

# Molecular Systems Biology

Mathematical models and theory for the study of the  
molecular systems underlying cell biology

F.J. Bruggeman, VU University, Amsterdam, The Netherlands, 2013



# Contents

|   |     |
|---|-----|
| <b>Preface</b>  | vii |
| <b>How to read this book</b>  | ix  |
| <b>1 Molecular networks and systems biology</b>   | 1   |
| 1.1 Living cells . . . . .  | 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 . . . . .                    | 8   |
| 1.3.1 Cells are dynamic! . . . . .  | 8   |
| 1.3.2 Cells and their networks are organized in space and time! . . . . .                                 | 10  |
| 1.3.3 Isogenic cells can display large cell-to-cell heterogeneity! . . . . .                              | 10  |
| 1.3.4 Evolutionary network adaptations can be tracked and predicted! . . . . .                            | 12  |
| 1.3.5 Network principles exist that apply to many species! . . . . .                                      | 12  |
| 1.3.6 Predictive mathematical modeling of molecular network dynamics is feasible! . . . . .               | 12  |
| 1.3.7 Foremost biology but also engineering, physics and mathematics in <i>systems</i> biology! . . . . . | 12  |
| <b>2 Kinetic description of the biochemical reactions between molecules</b>                               | 15  |
| 2.1 Reactions between molecules are the basic processes of life . . . . .                                 | 15  |
| 2.2 The quantitative description of molecular reactions . . . . .   | 16  |
| 2.2.1 Mass balances . . . . .   | 16  |
| 2.2.2 Mass-action-kinetics . . . . .  | 17  |
| 2.3 Linear growth . . . . .   | 20  |
| 2.3.1 Exercise . . . . .  | 20  |
| 2.4 Exponential growth . . . . .  | 21  |
| 2.5 Exponential decay . . . . .   | 22  |
| 2.6 Zeroth-order synthesis and first-order degradation . . . . .  | 22  |
| 2.6.1 Exercise . . . . .  | 24  |

|          |  |           |
|----------|--|-----------|
| 2.7      | How many products/targets can a catalyst produce/activate during its life time? . . . . .                                  | 24        |
| 2.8      | Rate characteristics, thermodynamic equilibrium and steady state . . . . .   | 26        |
| 2.9      | Rate characteristics, attracting states, and dynamics . . . . .  | 29        |
| 2.10     | Binding equilibria, association and dissociation constants . . . . .   | 31        |
| 2.11     | A number of biological examples . . . . .  | 32        |
| 2.11.1   | Protein complex formation (different perspective) . . . . .  | 32        |
| 2.11.2   | Cooperative transcription factor binding to DNA targets . . . . .  | 34        |
| 2.11.3   | Negative autoregulation of a gene . . . . .  | 37        |
| 2.12     | Example networks: G-protein coupled receptor signalling . . . . .  | 39        |
| <b>3</b> | <b>The remarkable and emergent functionality of small molecular circuits</b>   | <b>47</b> |
| 3.1      | Small, functional molecular systems are continuously exploited by cells . . . . .  | 47        |
| 3.2      | Example 1: Enzyme saturation as a mechanism for robustness . . . . .   | 49        |
| 3.3      | Example 2: Perfect adaptation of chemotaxis . . . . .  | 50        |
| 3.3.1    | Mathematical explanation: integral feedback control . . . . .  | 53        |
| 3.3.2    | Integral feedback control (the mathematical background) . . . . .  | 54        |
| 3.4      | Example 3: Fold change detection . . . . .   | 55        |
| 3.5      | Example 4: Kinetic proofreading . . . . .  | 57        |
| 3.5.1    | Different explanation for why kinetic proofreading improves accuracy . . . . .   | 61        |
| 3.6      | Example 5: Robustness of two component signalling networks . . . . .   | 62        |
| 3.7      | Concluding remarks . . . . .   | 64        |
| <b>4</b> | <b>Diffusion and reactions</b>   | <b>65</b> |
| 4.1      | Introduction . . . . .   | 65        |
| 4.2      | Characteristic dimensions for cell biology . . . . .   | 65        |
| 4.2.1    | Performing calculations with units . . . . .   | 66        |
| 4.3      | Diffusion of molecules . . . . .   | 70        |
| 4.4      | Example calculation of travelled distance by macromolecules . . . . .  | 74        |
| 4.5      | Expressing the diffusion coefficient in terms of the dimensions of the molecule and the properties of the medium . . . . . | 75        |
| 4.6      | Diffusion-limited association reactions . . . . .  | 77        |
| 4.7      | When an association reaction becomes diffusion limited . . . . .   | 78        |
| 4.8      | Mean single molecule search times in the cell . . . . .  | 78        |
| 4.9      | Search times by single molecules and association kinetics . . . . .  | 79        |
| 4.10     | Comparison of molecule search times . . . . .  | 80        |
| <b>5</b> | <b>Using diffusion times to get more insight into some basic cellular processes</b>  | <b>83</b> |
| 5.1      | Molecule search times and the response time of prokaryotic two-component signalling . . . . .                              | 83        |
| 5.2      | Leaky transcription rates due to spontaneous fluctuations in the promoter occupation by a repressor . . . . .              | 84        |

|          |  |            |
|----------|--|------------|
| 5.3      | A diffusion limited kinase . . . . .   | 85         |
| 5.3.1    | Diffusion-limited association . . . . .  | 85         |
| 5.3.2    | Diffusion-limited association and this is also the slowest process in the enzyme . . . . . | 86         |
| 5.3.3    | Diffusion-limited association and the catalytic rate is limiting enzyme rate . . . . .     | 86         |
| 5.4      | Facilitated diffusion through biological membranes . . . . .                               | 86         |
| 5.4.1    | Fick's law and membrane diffusion . . . . .  | 86         |
| 5.4.2    | Facilitated passive transport via a dedicated protein . . . . .                            | 87         |
| 5.4.3    | Active transport . . . . .   | 88         |
| 5.4.4    | Facilitated passive transport of a weak acid . . . . .                                     | 88         |
| 5.4.5    | Facilitated diffusion via a carrier protein . . . . .                                      | 89         |
| <b>6</b> | <b>Basic enzyme kinetics</b>   | <b>93</b>  |
| 6.1      | Enzymes . . . . .  | 93         |
| 6.2      | Irreversible Michaelis-Menten kinetics . . . . .   | 93         |
| 6.2.1    | Derivation of enzyme kinetics: quasi-steady state assumption . . . . .                     | 94         |
| 6.2.2    | Derivation of enzyme kinetics: equilibrium-binding assumption . . . . .                    | 96         |
| 6.2.3    | What's worse: assuming quasi-steady state or rapid equilibrium? . . . . .                  | 97         |
| 6.2.4    | Consideration of inhibitors and activators: equilibrium-binding assumption . . . . .       | 97         |
| 6.2.5    | Exercises . . . . .  | 99         |
| 6.2.6    | Sensitivity of the enzyme rate to reactants and effectors . . . . .                        | 99         |
| 6.3      | Reversible Michaelis-Menten kinetics: One enzyme-reactant state                            | 100        |
| 6.3.1    | Quasi-steady state approximation . . . . .   | 100        |
| 6.4      | Reversible Michaelis-Menten kinetics: Two enzyme-reactant states                           | 101        |
| 6.4.1    | Rapid-equilibrium approximation . . . . .  | 101        |
| 6.4.2    | Quasi-steady state approximation . . . . .   | 102        |
| 6.5      | Enzyme action and thermodynamics (advanced material) . . . . .                             | 105        |
| 6.6      | Enzyme inhibition: quasi-steady state approach . . . . .                                   | 108        |
| 6.7      | Equilibrium binding models and convenience kinetics . . . . .                              | 111        |
| <b>7</b> | <b>Cooperative enzymes</b>   | <b>115</b> |
| 7.1      | The regulatory potential of cooperative enzymes . . . . .                                  | 115        |
| 7.2      | The Monod Wyman Changeux model for cooperative enzyme kinetics . . . . .                   | 117        |
| 7.2.1    | Incorporation of allosteric modifiers . . . . .  | 120        |
| 7.2.2    | Generalisation to any number of subunits . . . . .   | 120        |
| 7.3      | The reversible Hill Equation . . . . .   | 123        |

|   |            |
|---|------------|
| <b>8 Examples of cooperative proteins</b>   | <b>129</b> |
| 8.1 Cooperative sensors: membrane receptors and transcription factors   | 129        |
| 8.1.1 Binding of the signal induces a conformational change when $K_R \neq K_T$                                 | 130        |
| 8.1.2 The fraction of bound subunits (the MWC perspective)  | 130        |
| 8.1.3 A dimer and state perspective   | 132        |
| 8.1.4 Consideration of additional regulators  | 133        |
| 8.1.5 Assessing the regulatory potential of $R_2S_2$  | 134        |
| 8.1.6 Exercise  | 135        |
| <b>9 Stoichiometric network analysis</b>  | <b>137</b> |
| 9.1 Introduction  | 137        |
| 9.2 The stoichiometric matrix, $\mathbf{N}$   | 138        |
| 9.3 The principles of this chapter explained with a simple example  | 140        |
| 9.4 Determining $\mathbf{L}$ by inspection by tracing the cutting and pasting of molecule parts by enzymes      | 143        |
| 9.5 Determining the moiety conservation relations and steady-state flux relationships with Gaussian Elimination | 146        |
| 9.6 Independent fluxes and the kernel matrix $\mathbf{K}$   | 150        |
| 9.7 Determination of $\mathbf{K}$ with linear algebra   | 153        |
| 9.8 Determining moiety conservation and independent fluxes in Mathematica                                       | 154        |
| 9.9 Advanced material   | 156        |
| 9.9.1 Relations between $\mathbf{N}$ and $\mathbf{L}$   | 156        |
| 9.9.2 Derivation of the $\mathbf{K}$ and $\mathbf{L}$ matrix using linear algebra for any $\mathbf{N}$ matrix   | 157        |
| <b>10 Flux balance analysis</b>   | <b>161</b> |
| 10.1 Steady-state flux space and constraint-based modeling  | 161        |
| 10.2 Flux Balance Analysis  | 167        |
| 10.3 Applications and pitfalls  | 171        |
| 10.4 Sensitivity analysis   | 172        |
| <b>11 Dynamics of molecular systems</b>   | <b>177</b> |
| 11.1 Stability of steady states   | 177        |
| 11.2 Bistable dynamics of single-variable systems   | 180        |
| 11.2.1 Emergence and disappearance of bistability as function of a parameter                                    | 183        |
| 11.2.2 Bistability as a mechanism for memory  | 183        |
| 11.2.3 Intermezzo: the Goldbeter-Koshland function  | 187        |
| 11.2.4 A one-way switch   | 190        |
| 11.2.5 A two-way switch   | 193        |
| 11.2.6 Homeostasis  | 194        |
| 11.3 Stability of two variable systems  | 197        |
| 11.4 Phase plane analysis for two variable dynamic systems  | 197        |
| 11.5 Stability of steady states of dynamic systems with two variables   | 202        |

|   |            |
|---|------------|
| 11.5.1 Analysis of the 2x2 jacobian matrix . . . . .  | 204        |
| 11.5.2 Exercises . . . . .  | 206        |
| <b>12 Introduction to metabolic control analysis</b>  | <b>209</b> |
| 12.1 Introduction . . . . .   | 209        |
| 12.2 Cellular self-regulation . . . . .   | 211        |
| 12.3 System properties . . . . .  | 214        |
| 12.4 MCA the basics . . . . .   | 216        |
| 12.4.1 Response of a metabolic pathway to a change in an enzyme concentration . . . . .   | 216        |
| 12.4.2 Elasticity coefficients: local properties defined at the level of single reactions . . . . .   | 217        |
| 12.4.3 Flux control coefficients defined at the level of the entire network . . . . .   | 219        |
| 12.4.4 The rate of an enzyme catalysed reaction, dependency on enzyme concentration and general elasticity equations (skip on a first read) . . . . . | 222        |
| 12.5 The machinery of MCA: response, control and elasticity coefficients  | 223        |
| 12.6 Control coefficients for a linear pathway . . . . .  | 226        |
| 12.6.1 Summation and connectivity theorems for linear pathways  | 229        |
| 12.7 Control shifts due to changes in enzyme concentrations . . . . .   | 230        |
| 12.7.1 Irreversible enzymes generally loose flux control when increased in concentration . . . . .  | 230        |
| 12.7.2 Reversible enzymes generally loose flux control when increased in concentration and then operate closer to thermodynamic equilibrium . . . . . | 230        |
| 12.8 A branched pathway: one robust branch and the other highly sensitive . . . . .   | 231        |
| 12.8.1 Control coefficients . . . . .   | 231        |
| 12.8.2 Low-flux branches and branch point control . . . . .   | 232        |
| 12.9 A metabolic pathway with negative feedback: homeostasis . . . . .  | 233        |
| 12.10 The central dogma of molecular biology; a transcription and translation cascade . . . . .   | 235        |
| 12.11 Ultrasensitivity of signaling networks . . . . .  | 236        |
| 12.11.1 A single covalent modification cycle . . . . .  | 236        |
| 12.11.2 A signaling cascade: sensitivity amplification . . . . .  | 238        |
| 12.11.3 Signaling cascades with feedback . . . . .  | 239        |
| 12.12 Using the summation and connectivity theorems to make deductions . . . . .  | 241        |
| 12.12.1 Metabolic pathway with one enzyme without product inhibition . . . . .  | 241        |
| 12.12.2 Metabolic pathway with one enzyme without product inhibition and one negative feedback . . . . .  | 242        |
| 12.12.3 Metabolic pathway without product inhibition of the first reaction and one negative feedback . . . . .  | 243        |

|   |            |
|---|------------|
| <b>13 Flux maximization by optimal metabolic enzyme expression</b>                            | <b>245</b> |
| 13.1 Metabolic pathways, enzyme levels, and fitness . . . . .                                 | 245        |
| 13.2 Numerical illustration of pathway flux optimisation . . . . .                            | 246        |
| 13.3 Growth rate reduction as a response to nonfunctional protein expression . . . . .        | 247        |
| 13.4 Theoretical illustration of pathway flux optimisation (advanced material) . . . . .      | 249        |
| <b>14 Optimal gene control systems</b>  | <b>251</b> |
| 14.1 What's so amazing about molecular control systems? . . . . .                             | 251        |
| 14.2 Sensing and signal integration by enzymes, genes and networks .                          | 252        |
| 14.2.1 Sensing and integration . . . . .  | 252        |
| 14.2.2 Regulation of single enzymes, proteins, and genes . . . . .                            | 252        |
| 14.2.3 Multiple transcription factors regulating the transcription rate of one gene . . . . . | 254        |
| 14.3 Regulation and networks . . . . .  | 254        |
| 14.4 Control of metabolism of by gene expression . . . . .                                    | 254        |
| 14.4.1 A simple example . . . . .   | 254        |
| 14.4.2 Optimal gene expression . . . . .  | 256        |
| <b>15 Answers to exercises</b>  | <b>259</b> |
| 15.1 Answers to Chapter 2 exercises . . . . .   | 259        |
| 15.2 Answers to Chapter 3 exercises . . . . .   | 268        |
| 15.3 Answers to Chapter 4 exercises . . . . .   | 270        |
| 15.4 Answers to Chapter 6 exercises . . . . .   | 276        |
| 15.5 Answers to Chapter 7 exercises . . . . .   | 282        |
| 15.6 Answers to Chapter 8 exercises . . . . .   | 284        |
| 15.6.1 Exercise . . . . .   | 284        |
| 15.7 Answers to Chapter 9 exercises . . . . .   | 285        |
| 15.7.1 Exercise . . . . .   | 285        |
| 15.8 Answers to Chapter 10 exercises . . . . .  | 293        |
| 15.9 Answers to Chapter 12 exercises . . . . .  | 297        |

# Preface

Living cells are the fundamental unit of all life on this planet. They occur as free-living unicellular organisms or as components of multicellular organisms, like us. Cells are amazing molecular systems. They assimilate nutrients to maintain themselves, they reproduce, communicate with each other, and compete. Free-living unicellulars continuously adapt to changes in external conditions by environmental sensing, information integration, and the execution of an altered physiological strategy. They can decide to remain dormant when conditions are poor. Some cells communicate with others, either in a multicellular context or a microbial community. Cells in a multicellular context actively coordinate their activities to give rise to astonishing emergent properties at the levels of tissues and whole organisms. On top of that, cells evolve. They change their capacities blindly through random mutations, while being subjected to a game of natural selection where the most fit – best adapted – genotype wins. This can lead to the appearance of new microbial species or renegade cancerous cells.

All aspects of these resourceful creatures – cells – emerge from the concerted action of molecules, working actively together in networks. Some of those molecules are small compounds, such as glucose or lactate. Others are complicated molecular machines, encoded on the organism's DNA, taking care of, for instance, protein synthesis, protein transport or DNA replication. All those molecules are in a constant flux: they are being synthesised and degraded, and move through the cell. Understanding the dynamics of one molecule therefore requires considering many others as well, all participating in a molecular network.

The emergent properties of molecular networks is what makes cells ‘alive’. Thus, the underlying molecular mechanisms are the natural terms for understanding life. This book explains how some of this understanding can be achieved using mathematical models and theory. It augments the experimental approaches to studying cellular systems as part of the research field *systems biology*. We will require theoretical approaches and models originally developed in physics, biochemistry, control theory, and dynamical systems theory. All of these will be explained at a level of undergraduate mathematicians, physicists, engineers and biologists. I hope you will enjoy reading and studying this book. Biology has never been so vibrant and challenging as in current times!



# How to read this book

This book is an introduction to the mathematical modelling of the molecular systems that we find inside living cells. It is used in the master course *Basic Models of Biological Networks* at the *VU University* in Amsterdam, The Netherlands. The book has evolved over the last few years to what it is at the moment. Every year it is updated and extended. Unfortunately, it still contains typographical errors and omissions. I apologise for them in advance and hope that you will inform me about them – I will then communicate them to the other students following the course and update the book.

The target audience of this book are master students in the master program *Systems Biology and Bioinformatics* at the *VU University*, which either have a biology, bioinformatics or some other background; generally they have not been exposed to modelling of biological systems in their bachelor teaching. For those with a biology background the modelling will be new and for those with a physics or engineering background the kind of thinking about biological systems will be useful.

During the writing stages of this book, many colleagues have helped improving it and in some cases wrote small contributions. I would particularly like to thank Bas Teusink (FBA), Timo Maarleveld (FBA), Jacky Snoep (proofreading of chapters), Klaas Krab, Evert Bosdriesz (integral control), Anne Schwabe (kinetic proofreading), Susanne Roth (fold change detection), Joost Hulshof (bifurcations and stability) and Jan Berkhou (FBA).

Amsterdam, 2015,  
Frank Bruggeman,  
Systems Bioinformatics,  
VU University,  
Amsterdam,  
The Netherlands.  
email: f.j.bruggeman@vu.nl



# Chapter 1

## Molecular networks and systems biology

### 1.1 Living cells

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 and its kinetic properties.

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. In these networks, all the molecules are indirectly coupled to all others.

Since 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 occur 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

## 2 CHAPTER 1. MOLECULAR NETWORKS AND SYSTEMS BIOLOGY

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 – even seemingly simple ones. 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, described 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 using models.

## 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. In other words, cells are organized systems kept continuously 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 see 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.

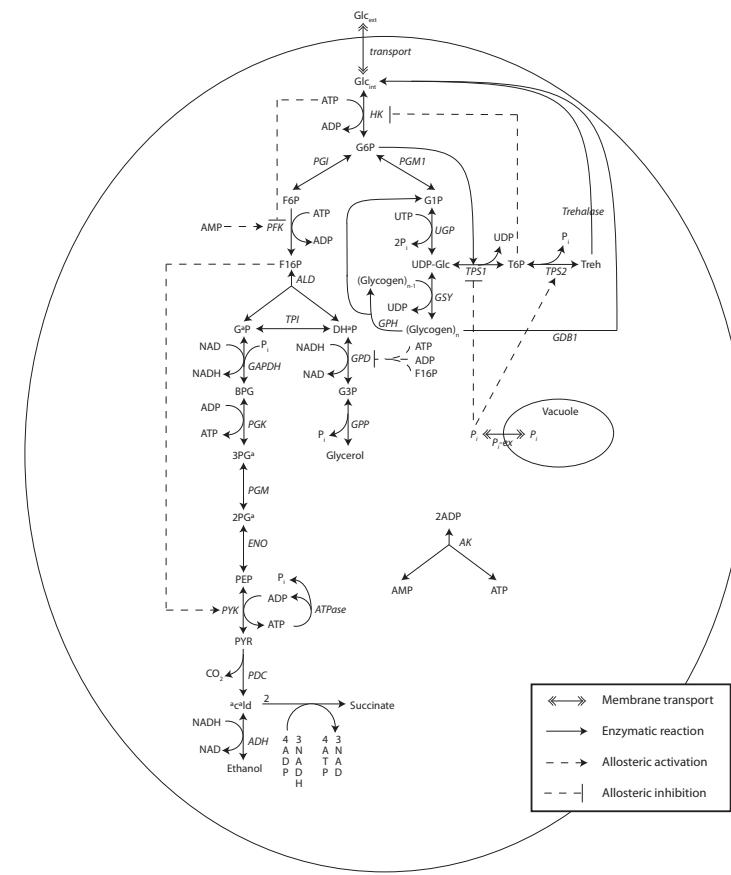


Figure 1.1: Network diagram of glycolysis as it occurs in *Saccharomyces cerevisiae*, a yeast species of interest to biotechnology. It is also often used as a eukaryotic model organism in cell biology and systems biology. The arrows denote reactions, every reaction has a dedicated enzyme as a catalyst. 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.

Glycolysis is involved in the conversion of sugars into building blocks and energy metabolites. Many organisms rely on glycolysis; including yeast, most

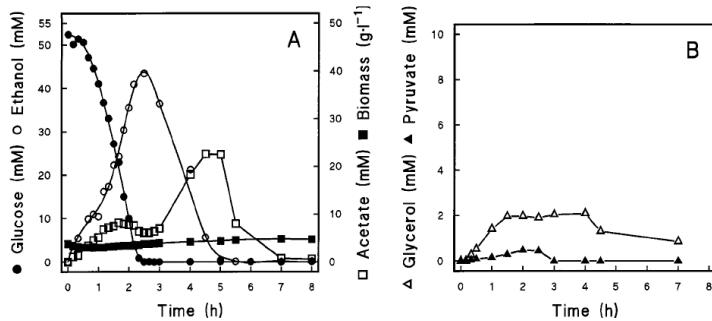


Figure 1.2: The dynamics of extracellular products of glycolysis upon a glucose pulse to *Saccharomyces cerevisiae* [12].

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. Up to 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. Up to 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  $\text{CO}_2$  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 cerevisiae* to a glucose pulse is shown in figure 1.2. Five gram of this yeast per liter medium consumed  $\approx 52 \text{ mM}$  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 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 [38]. 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 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 perceive 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  $\beta$  (TGF $\beta$ ), to alter cell growth, adhesion, differentiation, and controlled cell death [32].

A number of proteins are involved in this process of TGF $\beta$ -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 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 TFG $\beta$ . All of those are

## 6 CHAPTER 1. MOLECULAR NETWORKS AND SYSTEMS BIOLOGY

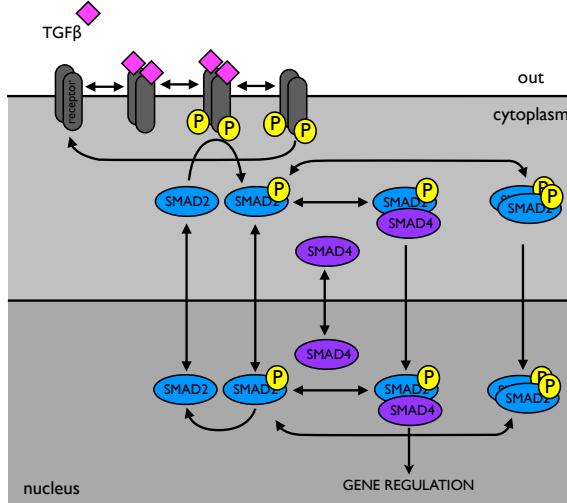


Figure 1.3: Network diagram of the Smad signaling network that responds to external TGF $\beta$  (TGF=transforming growth factor) levels [32]. This signaling pathway, as many others, is involved in regulation of cell growth, adhesion, migration, cell-fate determination and apoptosis.

not shown in the diagram. The action of the phosphatase in the nucleus leads to the dephosphorylation 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, 33].

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 its output, ERKPP, to the signal EGF, [22] and the system could display oscillations [24]. Even though, ultrasensitivity remains slightly controversial and is perhaps condition dependent, oscillations have now been observed experimentally at the level of single cells [35] (see figure 1.6).

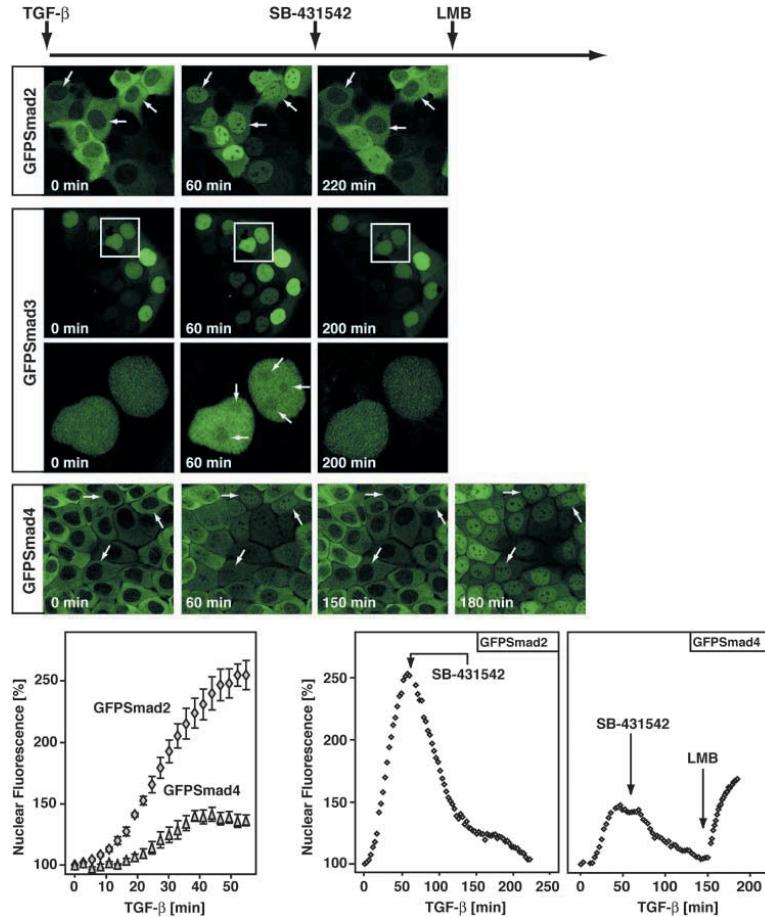


Figure 1.4: Single-cell dynamics of fluorescently labelled Smad2, GFPSmad2, and Smad4, GFPSmad4. Within 60 minutes both Smads translocate to the nucleus upon addition of TGF $\beta$ . The addition of the receptor inhibitor, SB-431542, causes Smad2 to return to the nucleus. Smad2 is then no longer phosphorylated 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 [30].

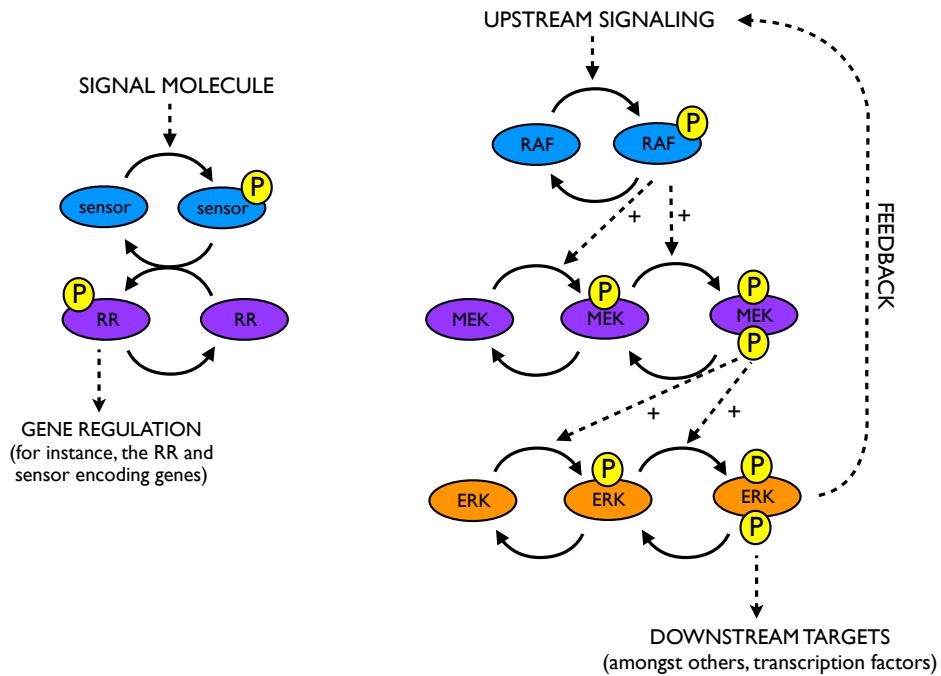


Figure 1.5: Protein phosphorylation is a recurrent mechanism for signal transduction (see also figure 1.3). Upon phosphorylation proteins change their affinity for signaling partners. Two component signaling as often observed in prokaryotes (and plants) and a mitogen activation protein kinase (MAPK) cascade. MAPK signaling has been studied in great depth in the last decade using mathematical models and quantitative experimentation [25, 35, 22].

## 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 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

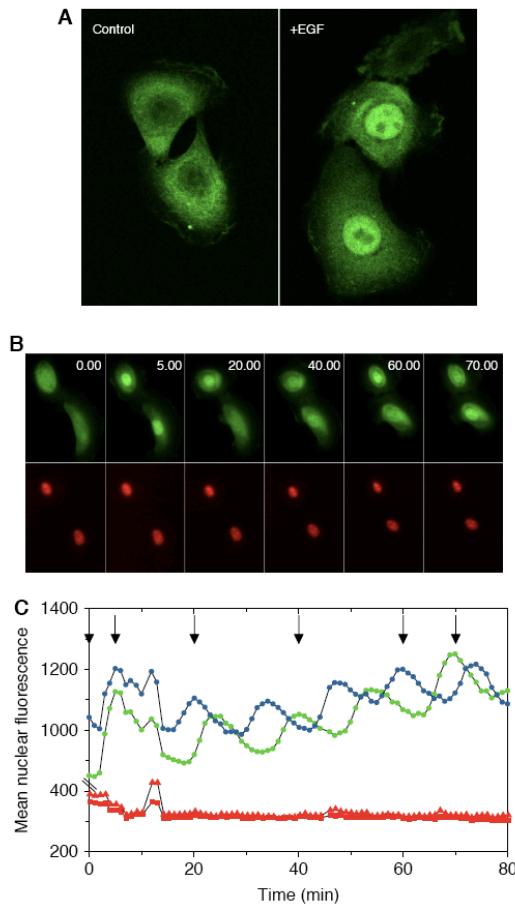


Figure 1.6: Single-cell monitoring of oscillatory dynamics of doubly phosphorylated ERK in both the cytoplasm and the nucleus [35]. Panel A shows that the total fluorescence increases upon addition of epidermal growth factor (EGF) - the signal of the signaling network. This indicates activation. In addition, *ERKPP* accumulates in the nucleus. On top this accumulation, 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.

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.

### 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.

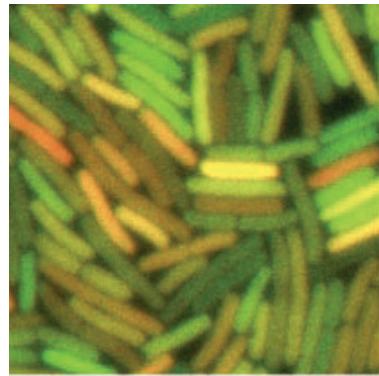


Figure 1.7: Heterogeneity of the expression of a fluorescent protein in a population of isogenic *Escherichia coli* cells [11].

### 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 is 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.

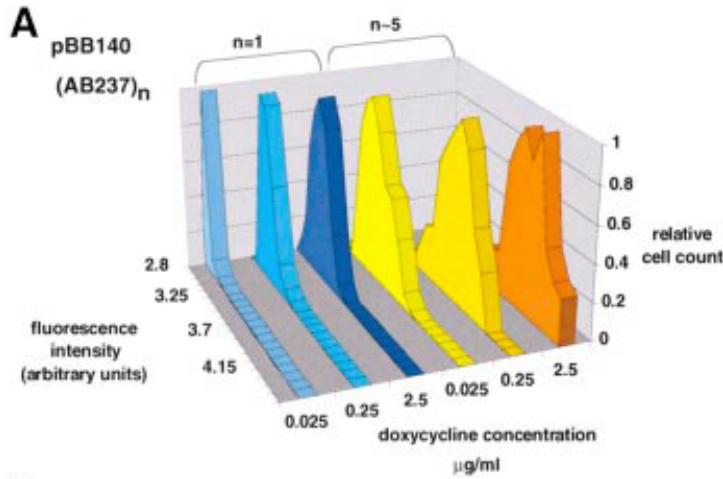


Figure 1.8: Distribution of fluorescence intensity per cell as function of a gene regulator [1].

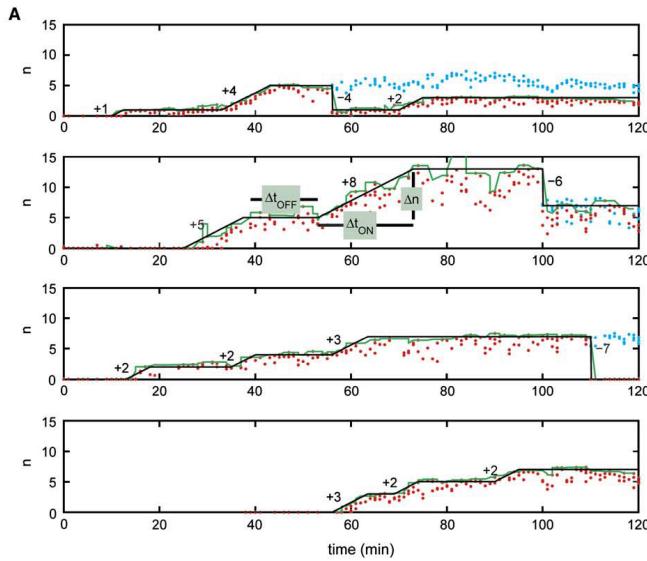


Figure 1.9: Variability of mRNA levels per cell as function of time for four cells [16].

### 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 out 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 [23]. 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.

### 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

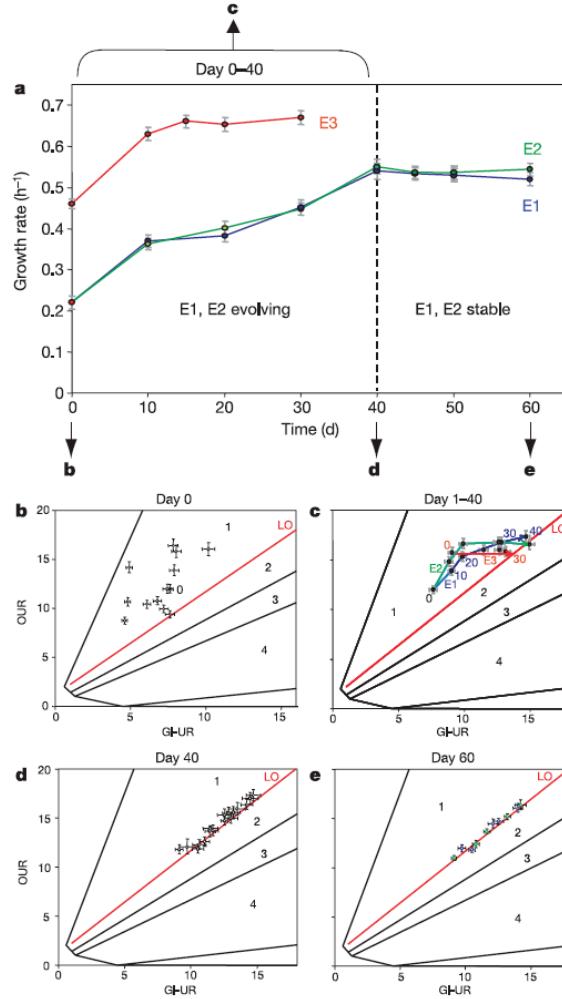


Figure 1.10: Adaptive evolution of *E. coli* growing on glycerol [23]. As function of time the population reaches the red line of optimality (figs b-e) as predicted by flux analysis of *E. coli*'s metabolic network.

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

# Kinetic description of the biochemical reactions between molecules

### 2.1 Reactions between molecules are the basic processes of life

To understand how the molecules inside cells bring about cellular behaviour requires understanding of cellular activities, such as 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, catalysing reactions, or may have constructive roles, for instance, actin plays a role as a monomer in the cytoskeleton, nucleosomes wrap DNA, or specific proteins make up a microorganism's flagellum (the propellor that microorganisms use to move through fluids). Proteins carry out their roles by interacting with other molecules, by forming complexes, or by catalysing conversions. These fundamental activities of proteins can be quantitatively described in terms of their kinetic properties. This means that the behaviours of a cell are ultimately the consequence of the kinetic properties of all its molecules! Those kinetics properties relate in a very complicated manner to the DNA sequence of the gene coding for this protein; even though this relationship is understood in principle, we are still not able to calculate the consequences of individual gene mutations for the kinetic properties of the associated protein.

In this chapter, we will study how we can quantitatively describe reaction rates in terms of kinetic equations, and how those rates bring about changes in the concentrations of these molecules, giving rise to dynamic cellular activities. We will limit ourselves to uncatalyzed reactions and postpone the discussion of enzyme kinetics to a later chapter. The kinetics treated in this chapter is the basic toolkit for modelling of complex molecular systems in living cells, such as

enzymes, protein complex formation, metabolic pathways, gene expression and signalling networks.

With the toolkit explained in this chapter, you will already be able to understand the principles of a whole range of unexpected and sophisticated behaviours of molecular systems, such as fold change detection, robustness, perfect adaptation, and error minimisation through kinetic proofreading. These emergent properties of small molecular systems will be discussed in the next chapter. We will first focus on the basic theory of mass-action kinetics in this chapter.

## 2.2 The quantitative description of molecular reactions

### 2.2.1 Mass balances

We will assume throughout this chapter that we can describe the reactions between molecules without having to consider the diffusion of molecules, spatial organisation of the cell and the inherent stochastic aspects of reactions. Some of those aspects will be considered in later chapters. These assumptions turn out to be warranted in most of the cases, so not much generality is lost by making them.

What kind of reactions exist between molecules? The basic interaction between two molecules is that they can form a complex. This molecular complex can then fall apart after some ‘life’ time. The concentrations of the two molecules and the complex change after the occurrence complex formation and the complex dissociation reaction. How fast those concentrations changes will depend on the rate of the reaction, i.e. how quickly those molecules find each other and form a complex and how stable the resulting complex is.

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 molecule plays a role in, as a substrate and product, we can determine the net change in the concentration of this molecule. This calculation resembles ‘molecular accounting’. A natural approach to accounting is to make use of balances – as you do when managing your bank account. Here we do not deal with money but with numbers of molecules, e.g. expressed as a concentration. (A concentration is just the number of molecules in a volume, and if the volume remains fixed, changes in the concentration are only due to changes in the number of molecules.) Since, you can count molecules in the same way as euros or dollars, the same principles apply to molecular and money accounting.

Thus, setting up a mass balance is a natural approach to 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 millimolar,  $mM$ . The concentration is now defined as the number of molecules of  $X$ ,  $n_X$ , divided by the volume,  $V$ , of the system it is in, e.g. the cell:  $x = \frac{n_X}{V}$ . Because we assume the volume to be fixed, the changes in the concentration are only due to the changes in the number of molecules, due to the activity of reactions.

The rate of change in the concentration,  $x$ , 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 at this time,  $v_{synth}(t)$  and  $v_{deg}(t)$ , of this molecule  $X$  with concentration  $x$ ,

$$\frac{d}{dt}x(t) = v_{synth}(t) - v_{deg}(t) = \sum_i v_{i,synth}(t) - \sum_j v_{j,deg}(t) \quad (2.1)$$

The net rates of synthesis and degradation equal the sum of the synthesis and degradation rates. 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  $t$ .

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 besides  $X$  and this dependency is given by a rate equation, which can either derive from mass action or enzyme kinetics. Mass action kinetics will be studied in this section and enzyme kinetics in a next chapter.

### 2.2.2 Mass-action-kinetics

Mass action kinetics applies to uncatalyzed reactions, so-called ‘spontaneous reactions’, reactions that do not require an enzyme as a catalyst. Intra-enzyme reactions, e.g. in the catalytic site, are also described by mass-action kinetics.

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  $S$ , of  $X$ , the intrinsic rate constant for isomerization,  $k^+$  and a similar rate constant,  $k^-$ , for the isomerization of  $X$  into  $S$ , i.e. the backward reaction. The reaction rate,  $v$ , is then given by:

$$v = k^+ s - k^- x \quad (2.2)$$

If the unit of the reaction rate is expressed in terms of  $\text{mM}/\text{min}$ , the unit of the concentration needs to be  $\text{mM}$  and the unit of the rate constants are then necessarily  $\text{min}^{-1}$ . The reversibility of the reaction dictates that the rate can also be negative, i.e. such that  $S$  is 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  $x^2$ . The reaction is said to be in thermodynamic equilibrium when  $v = 0$ , then  $\frac{x}{s} = \frac{k^-}{k^+}$ .

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

$$v = k^+ \cdot x \cdot y - k^- xy \quad (2.3)$$

Confirm that: the unit of the rate constant  $k^+$  should now be  $\text{min}^{-1}\text{mM}^{-1}$ . This rate constant is an example of a second-order rate constant, as its associated 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  $X + Y + Z \rightleftharpoons XYZ$  and would have unit  $\text{min}^{-1}\text{mM}^{-2}$ . A zeroth order rate constant is then associated with the reaction  $\rightarrow X$ , this may look weird, because  $X$  appears out of nothing, but this is often used as a shorthand notation when we do not want to be bothered with the substrate(-s) of the reaction. For convenience we subsume this information into the zeroth-order rate constant, which now has  $\text{mM}/\text{min}$  as unit. Obviously, a first-order reaction means  $X \rightarrow Y$  and has a rate constant with  $\text{min}^{-1}$  as unit.

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

$$v = k^+ x^2 - k^- x_2 \quad (2.4)$$

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,



the following rate equation for the reaction,

$$v = k^+ \prod_{i=1}^s x_i^{n_i} - k^- \prod_{j=1}^p y_j^{m_j} \quad (2.6)$$

The symbol  $\prod$  means product,

$$\begin{aligned} 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 &= \prod_{i=1}^5 i \\ Z_1 \cdot Z_2 \cdot Z_3 \cdot Z_4 \cdot Z_5 &= \prod_{i=1}^5 Z_i \end{aligned} \tag{2.7}$$

Returning to equation 2.6, this indicates that per unit time  $m_1 v$  molecules of  $Y_1$  are made, and  $m_i v$  molecules of  $Y_i$ . Note that the  $m_i$  can be positive or negative depending on whether molecules are produced or consumed, respectively.

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: per unit rate 2 molecules of  $X$  is consumed. Therefore the rate of the degradation of  $X$  is twice the rate of the production of  $X_2$ , which occurs at a rate  $v$ . We would obtain in this case for the mass balances of  $X$  and  $X_2$ ,

$$\begin{aligned} \frac{dx}{dt} &= -2(k^+ x^2 - k^- x_2) = -2v \\ \frac{dx_2}{dt} &= k^+ x^2 - k^- x_2 = v \end{aligned} \tag{2.8}$$

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 dimerization rate.

One more important aspect of the reaction  $2X \rightleftharpoons X_2$  is that the total amount of molecules of  $X$  remains fixed in this case: no molecules are lost; they are only interconverted. Thus we have the following relationship for the total concentration of  $X$ :  $x_T = x + 2x_2$ . The concentration  $x_T$  just equals the amount of molecules that the system started with at time zero and remains fixed over time: thus we have,  $x_T = x(0) + 2x_2(0) = x(t) + 2x_2(t)$ . 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  $dx/dt + 2dx_2/dt = -2v + 2v$  (see equation 2.8)! These tricks we will apply very often to reaction systems.

## Exercises

- Determine the mass balances and mass action kinetics for the following molecules and reactions. An underlined molecule indicates that it has a fixed concentration.
  - $S \rightleftharpoons X \rightleftharpoons P$
  - $\underline{S} \rightleftharpoons X \rightleftharpoons P$

- (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-phosphate and  $Z$  as ATP. What should be the substrate of reaction 1 and the product of reaction 3?

## 2.3 Linear growth

We consider the following reaction,



where  $X$  can be mRNA molecules produced during transcription, with  $x$  as the mRNA concentration. The associated mass balance equals,

$$\frac{dx}{dt} = k \quad (2.10)$$

So, per unit time we gain the same number of mRNA molecules, assuming that the volume remains fixed. In other words, the slope in the plot of  $x$  as function of  $t$ , i.e  $dx/dt$ , remains fixed: so, we  $x$  depends linearly on  $t$ ! Therefore,  $x$  grows linearly with time. We could also have concluded this by solving this differential equation,

$$\begin{aligned} \int_{x(0)}^{x(t)} dx &= \int_0^t k dt \\ \Rightarrow x(t) - x(0) &= kt \\ \Rightarrow x(t) &= x(0) + kt \end{aligned} \quad (2.11)$$

(See Figure 2.1.) This equation allows you to calculate the mRNA concentration as function of time when you know how many you start with and what the transcription activity,  $k$ , is.

### 2.3.1 Exercise

1. Solve the mass balances associated with



sketch the dependency of  $x$  on time, and show that this model simplifies to equation 2.11 when  $k_2$  equals zero. (Hint: the answer in the ‘Introduction to Systems Biology’ syllabus, on pages 43-45, that you can download from [http://bruggemanlab.nl/?page\\_id=229](http://bruggemanlab.nl/?page_id=229).)

2. Consider

$$\xrightarrow{k_1} X, \quad (2.13)$$

when the volume also changes as function of time as  $\frac{d}{dt}V = \mu V$  with  $\mu$  as the cellular growth rate. Given this information derive the equation  $\frac{dx}{dt}$ . (Hint: the answer in the ‘Introduction to Systems Biology’ syllabus, on pages 43-45, that you can download from [http://bruggemanlab.nl/?page\\_id=229](http://bruggemanlab.nl/?page_id=229).)

## 2.4 Exponential growth

Now, we consider autocatalysis. For instance, a cell makes itself – it grows and divides – and per cell we have a certain rate of synthesis of new cells. If the rate of cell synthesis per cell equals  $k$  then with  $n_x$  cells in a constant volume,  $V$ , the number of new cells formed at a given moment in time equals  $kn_x$ , and in terms of concentration  $kx$ . At a next moment in time we shall have more cells and therefore the synthesis rate of new cells is also higher. This is reflected in the associated mass balance for  $x$ ,

$$\frac{d}{dt}x = kx, \quad (2.14)$$

indicating that the slope of  $x$  as function of  $t$  increases with  $x$ . We can solve this equation by hand,

$$\begin{aligned} \int_{x(0)}^{x(t)} \frac{1}{x} dx &= \int_0^t k dt \\ \Rightarrow \ln x(t) - \ln x(0) &= \frac{\ln x(t)}{\ln x(0)} = kt \\ \Rightarrow x(t) &= x(0)e^{kt} \end{aligned} \quad (2.15)$$

(See Figure 2.1.)

How much time does it take to double the number of organisms?

$$\frac{x(t)}{x(0)} = 2 = e^{kt_d} \Rightarrow t_d = \frac{\ln 2}{k} \quad (2.16)$$

with  $t_d$  as the doubling time. In other words, we have,

$$\begin{aligned} n_x(t_d) &= 2^1 n_x(0) \\ n_x(2t_d) &= 2^1 n_x(t_d) = 2^2 n_x(0) \\ n_x(3t_d) &= 2^1 n_x(2t_d) = 2^2 n_x(t_d) = 2^3 n_x(0) \\ &\dots \\ n_x(gt_d) &= 2^g n_x(0) \end{aligned} \quad (2.17)$$

with  $g$  as the number of doublings or ‘generations’.

## 2.5 Exponential decay

Protein complexes play a fundamental role in cells, for instance in signalling and gene expression. The functionality of a protein complex is limited by its life time, i.e. the time when it falls apart. So, a natural question to ask is: what determines the life time of a protein complex? The life time of this protein complex can easily be determined, because we only have to consider the following mass balance,

$$\frac{dx}{dt} = -kx \quad (2.18)$$

Solving this equation gives,

$$\begin{aligned} \int_{x(0)}^{x(t)} \frac{dx}{x} &= \int_0^t -k dt \\ \ln x(t) - \ln x(0) &= -kt \\ \ln \frac{x(t)}{x(0)} &= -k \cdot t \\ \frac{x(t)}{x(0)} &= e^{-kt} \end{aligned} \quad (2.19)$$

(See Figure 2.1.) Thus, 50% of the complex has been degraded when  $\frac{1}{2} = e^{-kt\frac{1}{2}}$ , which gives rise to a half life time of  $\frac{\ln 2}{k} = t_{\frac{1}{2}}$ . And after  $t = 1/k$  time the ratio equals  $\frac{x(t)}{x(0)} = 1/e \approx 0.37$ ; thus, about 2/3 of the complex has been degraded; this time is often called the characteristic time. The characteristic time is the most relevant time as it corresponds to the average life time of an system of identical molecules that decay by a first-order process.<sup>1</sup>

## 2.6 Zeroth-order synthesis and first-order degradation

Clearly, the following models has broad applicability,




---

<sup>1</sup>This will perhaps be explained in a later chapter in more depth but the short version is this: The reaction:  $X \rightarrow$  with rate constant  $k$  dictates that the life time,  $t$ , for a *single* molecule of  $X$  is a random variable that is distributed according to an exponential distribution,  $t \sim k \cdot e^{-k \cdot t}$ . The “~”-symbol means “is distributed as”. The mean life time of the molecule,  $\tau$ , can be calculated and equals  $\tau = \int_0^\infty t \cdot k \cdot e^{-k \cdot t} dt = \frac{1}{k}$ . If you cannot follow this completely then do not worry; the important thing to remember is that the characteristic time is a proper measure for the life time of a molecule or protein complex. After this time, for a set of molecules 2/3 will be degraded on average.

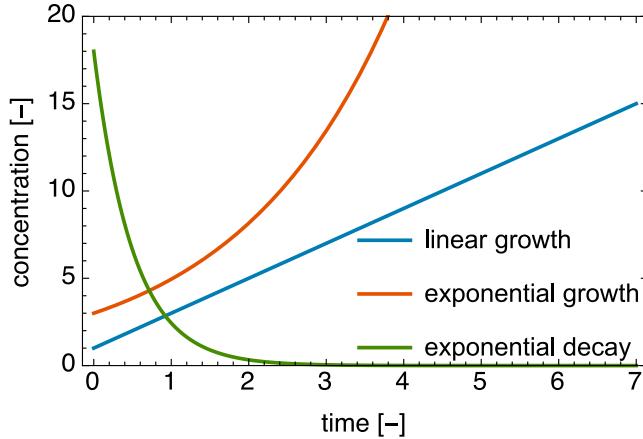


Figure 2.1: **Linear growth, exponential growth, and exponential decay.** Here we plot the following functions: linear growth  $x(t) = 1 + 2t$ , exponential growth  $x(t) = 3e^{0.5t}$ , and exponential decay  $x(t) = 18e^{-2t}$ .

using your basic math skills you can solve this equation and obtain,

$$x(t) = x(0)e^{-k_2 t} + (1 - e^{-k_2 t}) \frac{k_1}{k_2} \quad (2.21)$$

We also note that when  $dx/dt = 0$  that  $x = \frac{x_1}{x_2} \equiv x_s$  with  $x_s$  as the steady state concentration of  $x$ , it will become clear in a moment what this means, so now have,

$$x(t) = x(0)e^{-k_2 t} + (1 - e^{-k_2 t})x_s. \quad (2.22)$$

Analysing this equation we conclude that,

1. When you set the time to 0 then you obtain  $x(t = 0) = x(0)$ , like you should,
2. For very large times ( $t \rightarrow \infty$ , or  $t \gg 1/k_2$ ) then  $x(\infty) = x_s$ . So, the system eventually attains a steady state where: i.  $dx/dt = 0$ , ii.  $x = x_s$ , and iii.  $k_1 = k_2 x_s$ ,
3. when  $k_1 = 0$ , you get  $x(t) = x(0)e^{-k_2 t}$  indicating exponential decay,
4. when  $x(0) = 0$  then the system simplifies to  $x(t) = (1 - e^{-k_2 t})x_s$ , which indeed gives  $x(0) = 0$  and  $\frac{x(t)}{x_s} = \frac{1}{2} = 1 - e^{-k_2 t_{1/2}}$  and  $t_{1/2} = \frac{\ln 2}{k_2}$ . So, the half time for a system starting in a zero state is determined by the life time of the molecule  $X$ , so by  $1/k_2$ ! And, not by the synthesis time! This is important to remember and many of you do not realise this. It is very simple to understand, when we fix  $x_s$  and make  $k_2$  higher than indeed the time to reach  $x_s$  shortens, because we have to increase  $k_1$  as well to reach the same steady state,  $x_s$ , because  $x_s = k_1/k_2$ !

So, we can give the main equation an interpretation,

$$x(t) = \underbrace{x(0)e^{-k_2 t}}_{\substack{\text{Exponential decay} \\ \text{of initial concentration}}} + \underbrace{(1 - e^{-k_2 t})x_s}_{\substack{\text{Combined synthesis and degradation} \\ \text{and eventually } x_s \text{ is reached}}} . \quad (2.23)$$

This equation also tells you something else: regardless of the initial condition, the steady state  $x_s$  is always reached. So if  $x(0) > x_s$  or  $x(0) < x_s$ ,  $x_s$  is always the final state after some time of dynamics. When  $x(t) = x_s$  then the system remains in this state forever. Note that in this state, mass is continuously flowing at rate  $k_1 = k_2 x_s$ : the synthesis rate equals the degradation rate!

### 2.6.1 Exercise

Consider the following system,



with the concentration of  $S$  and  $P$  fixed (hence, the underline). Write down the mass balance for  $Y$  and show that  $y(t)$  can be found by analogy with the previous section, by rewriting the mass balance for  $y$ . Show that generally  $y_s$  is such that  $v_1 = v_2 \neq 0$ , this is called a steady state, a state of the system when mass flows continuously through the system. Does mass always flow in the same direction? Only when  $S$  and  $P$  are chosen in a particular manner do we get the so-called equilibrium state when  $v_1 = v_2 = 0$ . What is the expression that relates the equilibrium concentration of  $y$ ,  $y_e$ , to the parameters of the system? Set the parameters to the following values:  $k_1^+ = 10$ ,  $k_1^- = 1$ ,  $k_2^+ = 8$ , and  $k_2^- = 2$ , determine a concentration combination of  $S$  and  $P$  when equilibrium is reached. What happens to the mass flow when you decrease this  $P/S$  ratio and when you increase  $P/S$ ? Try to write the steady-state flux in terms of  $P/S$  and the remaining parameters of the system. Figure 2.2 should be helpful while doing this exercise.

## 2.7 How many products/targets can a catalyst produce/activate during its life time?

In many cellular mechanisms, a protein complex or enzyme in an active state activates several downstream processes, or from a single molecular template (gene or mRNA) several products are formed (mRNA or protein) catalysed by either a RNA polymerase or a ribosome (Figure 2.3). In all these cases, the number of downstream activated signalling molecules or the number of mRNA (or proteins) formed depends on: i. the life time of the catalyst or the template and ii. the time to make one product by conversion of substrate or to read off one product from a template. In this section, we will analyse a model that answers the following questions (Figure 2.3),

## 2.7. HOW MANY PRODUCTS/TARGETS CAN A CATALYST PRODUCE/ACTIVATE DURING ITS LIFE TIME

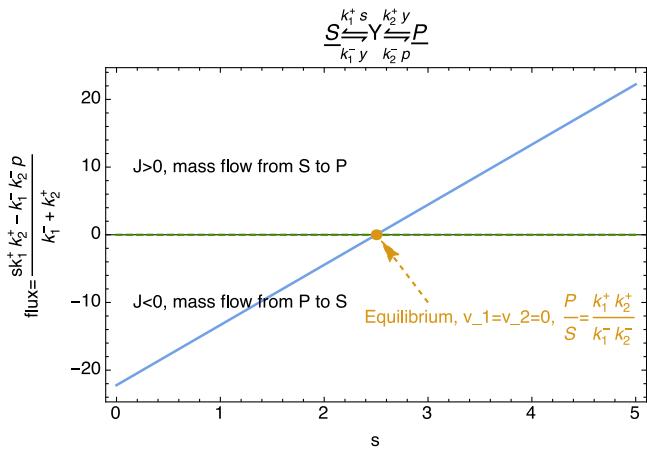


Figure 2.2: Two reversible reactions in series: steady state, equilibrium state, and flux reversal.

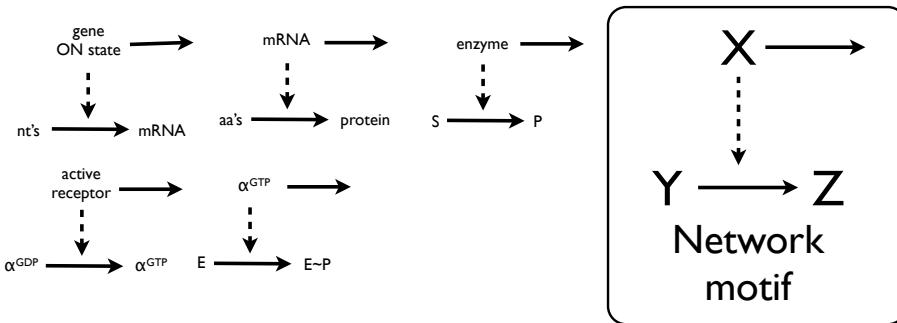


Figure 2.3: A recurrent network motif in molecular systems. In cells the following situation often occurs, a single protein or template lives for a short time before it is degraded and during this period it carries out some function, e.g. it catalyses a reaction or is used as a template and "read" to yield another product. With a network motif, a network is meant that occurs often in molecular circuits in cell biology. Legend: nt's = nucleotides, aa's = amino acids, s = substrate, p = product, E = enzyme, and E P = phosphorylated enzyme.

- How many proteins are made from a single mRNA molecule during its life time by ribosomes?
- How many downstream signaling proteins can be activated by an upstream signalling protein during its life time? E.g. by  $\alpha^{GTP}$  in Figure 2.8?
- How many products can a single enzyme produce during its life time?
- How many mRNAs can be produced during the life time of the active

state of a single gene by RNA polymerases?

This section therefore illustrates how little models can be exploited to answer a relevant biological question.

In a previous section, we determined the decay of the concentration of a molecule  $X$  in time due to a degradation reaction with a rate constant  $k$ , we will now call this rate constant  $k_d$ . Let's now consider the following situation, molecule  $X$  catalyses the conversion of  $Y$  into  $Z$  while it is not yet degraded. In the simplest, when the concentration of  $y$  is fixed, we have as the mass balance for  $Z$ ,

$$\frac{dz}{dt} = k_s \cdot y \cdot x(t) = k_s \cdot y \cdot x(0) \cdot e^{-k_d \cdot t} \quad (2.25)$$

Solving this differential equation with initial condition  $z(0) = 0$  gives,

$$z(t) = \frac{k_s \cdot y \cdot x(0)}{k_d} (1 - e^{-k_d \cdot t}) \quad (2.26)$$

When time goes to infinity, i.e. when all  $X$  has been degraded, the amount of  $Z$  produced equals,

$$z(t \rightarrow \infty) = z_{max} = \frac{k_s y}{k_d} \cdot x(0) \quad (2.27)$$

Where the ratio  $k_s y / k_d$  equals the number of molecules  $Z$ , given the substrate amount  $y$ , that a single unit of  $X$  can maximally produce and can be written as  $\tau_x^{max} / \tau_z$ : the maximal life time of  $X$  divided by the time to make one molecule of  $Z$ . This makes sense. Again a more realistic time measure is the characteristic time of  $X$ ,

$$z(t \rightarrow 1/k_d) = \frac{e - 1}{e} \cdot \frac{k_s y}{k_d} \cdot x(0) \approx 0.63 \cdot \frac{k_s y}{k_d} \cdot x(0) \quad (2.28)$$

An important aspect of signalling is amplitude amplification. Amplitude amplification is the phenomenon that one signal molecule leads to the activation of one or more downstream signalling molecules. A similar measure for transcription and translation is burst size. Transcription and translation bursts have to do with the number of mRNA produced per single ON state of a gene and the number of proteins produced per single mRNA. A quantification of both these phenomena is given by,

$$\frac{z(t \rightarrow 1/k_d)}{x(0)} = 0.63 \cdot \frac{k_s y}{k_d}. \quad (2.29)$$

If this quantity is larger than 1 then amplification or bursts occur.

## 2.8 Rate characteristics, thermodynamic equilibrium and steady state

In the previous sections, 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. Next we study the various so-called stationary states; states where the concentrations of molecules are fixed even though they do occur in reactions that synthesise and consume them.

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



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 for molecule  $X$ . If we assume the rates to follow mass action kinetics, we arrive at,

$$\frac{dx}{dt} = v_1 - v_2 = \underbrace{k_1^+ s - k_1^- x}_{v_1} - \underbrace{(k_2^+ x - k_2^- p)}_{v_2} \quad (2.31)$$

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 the difference between these rates  $x$  may rise or fall, steeply or only slightly. Alternatively, the rates balance and  $x$  remains fixed. This is shown in Figure 2.4D where the rate characteristics of this system is displayed. A rate characteristic is a plot of reaction rates as function of the concentration of its molecular reactants.

The two lines in Figure 2.4 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  $k_2^- p/k_2^+$ . 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  $x$  as function of time: it changes from its initial concentration to its value at the intersection between the two rate characteristics. The values of  $x$  where the two rates are equal is denoted by,  $x_S$ , and equals,

$$x_S = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} \quad (2.32)$$

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 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 all of the concentrations of the molecules are constant (and at least one of the reactions is unequal to zero). A stationary state requires 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 2.4 can become a state of thermodynamic 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,

$$\frac{p}{s} = \frac{k_1^+ k_2^+}{k_1^- k_2^-} \quad (2.33)$$

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, which is the requirement for a thermodynamic equilibrium state. The steady state and the thermodynamic equilibrium state are therefore both stationary states, as all the concentration are constant, but differ in the values of the reaction rates: in equilibrium *all* rates are equal to zero and in a steady state this is not the case.

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  $p$ .

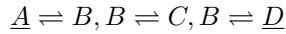
### Exercise

- 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 2.33 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

## 2.9. RATE CHARACTERISTICS, ATTRACTING STATES, AND DYNAMICS 29

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  $v_1$  equal  $v_2$ . 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  $v_2 = V_2x/(1+x)$  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:



## 2.9 Rate characteristics, attracting states, and dynamics

Using Figure 2.4 we will explain the relation between chemical kinetics, the mass balances, the rate equations, the stability of the stationary states, and the dynamics of concentrations of reactants. In figure 2.4, the left figures refer to the system  $\underline{S} \rightleftharpoons X$ ,

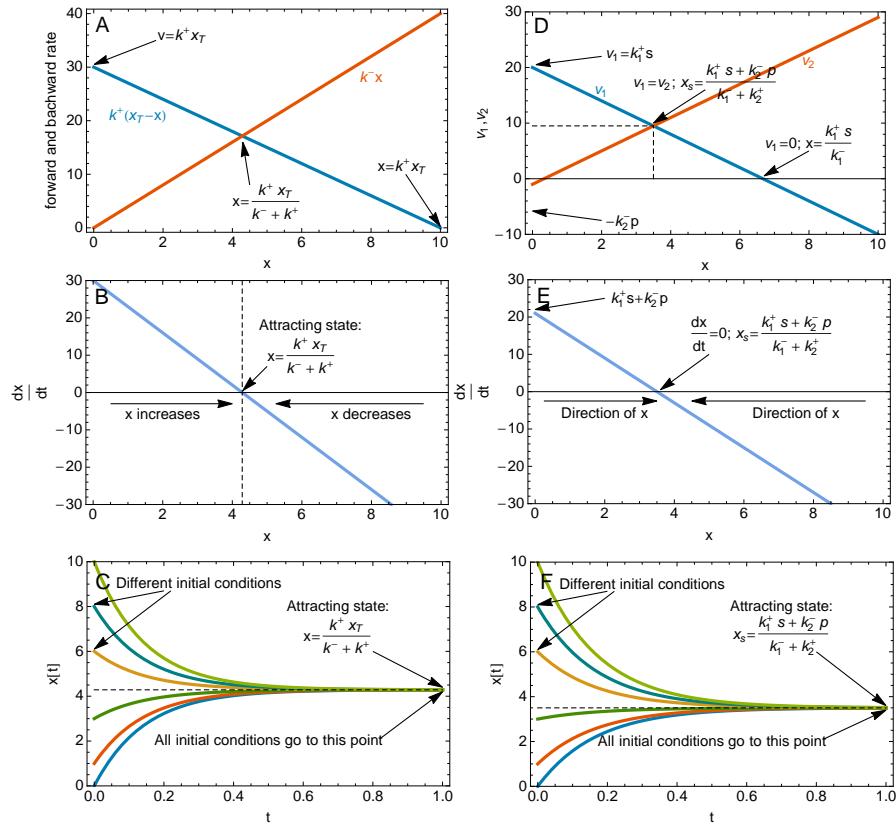
$$\begin{aligned} \frac{dx}{dt} &= v = k^+s - k^-x \\ x_T &= s(t) + x(t) = s(0) + x(0) \Rightarrow 0 \leq x \leq x_T \end{aligned} \quad (2.34)$$

Here we define  $x_T$  as the total concentration of  $x$ , which is set by the initial concentration of  $s$  ( $s(0)$ ) and  $x$  ( $x(0)$ ), such that at all times we have  $x_T = s(t) + x(t)$ . The rate characteristics of this system is shown in Figure 2.4A and we see that the two lines intersect when  $x = \frac{k^+x_T}{k^-+k^+}$ . Since, we are considering only one reaction, i.e.  $\underline{S} \rightleftharpoons X$ , the stationary state where  $dx/dt = v = 0$  is a thermodynamic equilibrium state because a reaction rate is zero. Note that this is not the case for the system considered in the second panel of plots in Figure 2.4. There  $v_1 = v_2 \neq 0$  at the stationary state and, hence, a steady state occurs.

Figure 2.4B & D indicate that the stationary state of the two systems is an attracting state: for all initial concentrations of  $x$  the system spontaneously evolves to the stationary state concentration. For  $x$  below the stationary state  $dx/dt > 0$  and for  $x$  above the stationary state  $dx/dt < 0$ . This attraction is

also illustrated in Figure 2.4C & F where the evolution of the concentration of  $X$  is shown as function of time for different initial conditions.

For a system with one variable the column figure shown in Figure 2.4 can always be generated and give a detailed insight into how the dynamics of the system follows from the dependency of the process rates on the concentration of the variable intermediate.



**Figure 2.4: Rate characteristics, stability of the final state, and dynamics.** On the left the system  $S \rightleftharpoons X$  is considered and on the right  $S \rightleftharpoons X \rightleftharpoons P$ . The upper figures show the rate characteristics; the dependency of the reaction rates on the concentration of the single variable concentration,  $x$ . The figures in the middle show the dependency of the rate of change  $\frac{dx}{dt}$  on the concentration of  $X$ ; both figures indicate that the stationary state – where  $\frac{dx}{dt} = 0$  – is an attracting state, as all concentrations of  $X$  are attracted to it. The lowest two figures illustrates the evolution of the concentration of  $X$  towards the stationary state as function of time.

## 2.10 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 proteins 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. Such constants will be introduced in this section and they will be used to study molecular complex formation.

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



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:  $a_T = a + ab$  (because  $ab \ll b_T$  and  $b \approx b_T$ ). 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  $A$ ,

$$\frac{da}{dt} = k_1^-(a_T - a) - k_1^+ a \cdot b \quad (2.36)$$

And this you can solve for the equilibrium concentrations using the information of the last section. In the equilibrium state, the association rate and dissociate rate are equal, such that the reaction rate is zero, and the total amount of  $A$  is fixed; thus we have the following relationships,

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

We can eliminate  $ab$  to obtain,

$$a_T = a + \frac{k_1^+ a \cdot b}{k_1^-} = a \left( 1 + \frac{b}{K_D} \right) \quad (2.37)$$

Here we have defined the *dissociation constant*  $K_D$ , which equals  $k_1^- / k_1^+$ . Note that it has concentration as unit! This means that unbound concentration of  $A$  equals,

$$a = \frac{a_T}{1 + \frac{b}{K_D}} \quad (2.38)$$

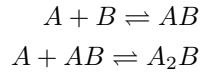
With the definition of the dissociation constant we can rewrite the equilibrium condition  $k_1^- ab = k_1^+ a \cdot b$  as  $ab = a \cdot b / K_D$  and we obtain for the bound concentration of  $A$ ,

$$ab = \frac{a_T \frac{b}{K_D}}{1 + \frac{b}{K_D}} = a_T \frac{b}{K_D + b} \quad (2.39)$$

Note that  $b = b_T$  as we assume  $B$  in excess. 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  $1\text{ nM}$ . 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:



Define a  $K_D$  for the first reaction and the second reaction. Do you understand that those can indeed be different? Assume 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  $a$ ,  $b_T$ ,  $K_{D1}$  and  $K_{D2}$ .

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



- (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?

## 2.11 A number of biological examples

### 2.11.1 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,



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  $ab \ll b$ . 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,



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,

$$\frac{dab}{dt} = k^+(a_T - ab)b - k^-ab \quad (2.43)$$

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,

$$ab(t) = \left(1 - e^{-(k^+b+k^-)t}\right) ab(\infty) = \left(1 - \frac{1}{e^{(k^+b+k^-)t}}\right) ab(\infty) \quad (2.44)$$

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 or the concentration  $b$  in  $k^+b+k^-$  is increased. At time  $t_{1/2} = \ln(2)/(k_b + b k_f)$  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,

$$\begin{aligned} \frac{ab(\infty)}{a(\infty) \cdot b} &= \frac{k^+}{k^-} \\ a_T &= a(\infty) + ab(\infty) \end{aligned} \quad (2.45)$$

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

$$ab(\infty) = ab_{EQ} = \frac{a_T \cdot b}{K_D + b} \quad (2.46)$$

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  $1 \text{ nM}$ . 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  $1 \text{ nM}^{-1}\text{s}^{-1}$ . Assume 10 transcription factors and 1 DNA site per cell and take  $E. coli$ 's cell volume ( $1 \text{ fl}$ ).

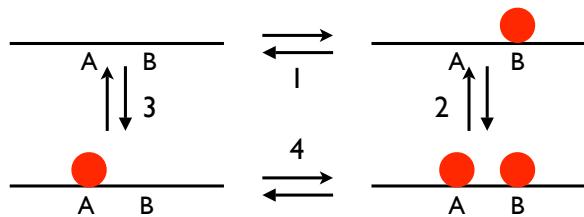
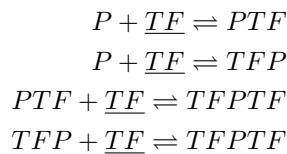


Figure 2.5: A state diagram of a gene promotor regulated by a transcription factor, the red ball. Two sites are present on the DNA,  $A$  and  $B$ .

### 2.11.2 Cooperative transcription factor binding to DNA targets

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 promoters in these states. We assume the transcription factor,  $TF$ , to be in excess. We have to deal with four reactions (PTF means binding site B is occupied and TFP means binding site A is occupied),



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

$$p_T = p + tfp + ptf + tfptf \quad (2.47)$$

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,

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

(You should know now how the  $K$ 's are defined in terms of the  $k$ 's.) Here the  $\alpha$  factor is an interaction coefficient that captures the effect of the presence of one transcription factor on the DNA when the next one binds. Positive cooperativity occurs when  $\alpha < 1$  and then the presence of a transcription factor on the promoter increases the affinity of the promoter for the second transcription factor. When  $\alpha > 1$  negative cooperativity occurs, and the second transcription factor binds with a lower affinity than the first. We allow for the possibility that the affinity of the DNA for the second transcription factor is altered when one transcription factor is already present. This cooperativity could derive 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 (by substitution of the last three equations in the promoter conservation relationship and solving for  $p$ ),

$$p = \frac{p_T}{1 + \frac{tf}{K_1} + \frac{tf}{K_2} + \frac{tf^2}{\alpha K_1 K_2}} \quad (2.48)$$

and for the fully occupied promotor concentration,

$$tfptf = \frac{p_T \frac{tf^2}{\alpha K_1 K_2}}{1 + \frac{tf}{K_1} + \frac{tf}{K_2} + \frac{tf^2}{\alpha K_1 K_2}} \quad (2.49)$$

If the binding sites are identical, such that  $K_1 = K_2$ ,

$$tfptf = \frac{p_T \frac{tf^2}{\alpha K^2}}{1 + 2 \frac{tf}{K} + \frac{tf^2}{\alpha K^2}} \quad (2.50)$$

If  $\alpha$  is very small then  $\frac{tf^2}{\alpha K^2}$  can become much larger than  $2 \frac{tf}{K}$ , which means that  $tfptf \gg tfp$  and  $tfptf \gg ptf$  and therefore

$$tfptf \approx \frac{p_T \frac{tf^2}{\alpha K^2}}{1 + \frac{tf^2}{\alpha K^2}} \quad (2.51)$$

This equation is often written in terms of new  $K$ 's without the mentioning of the  $\alpha$  constant,

$$tfptf \approx \frac{p_T \frac{tf^2}{K^2}}{1 + \frac{tf^2}{K^2}}, \quad (2.52)$$

and is often called the Hill equation. This equation we will study in more depth in the chapter on cooperative enzyme kinetics.

Alternatively, if the binding sites are equal, such that  $K_1 = K_2$ , and no cooperativity occurs, such that  $\alpha = 1$ , we obtain the Hill equation via another route,<sup>2</sup>

$$tfptf = \frac{p_T \frac{tf^2}{K^2}}{1 + 2\frac{tf}{K} + \frac{tf^2}{K^2}} = \frac{p_T \frac{tf^2}{K^2}}{\left(1 + \frac{tf}{K}\right)^2} = p_T \frac{\frac{tf}{K}}{\left(1 + \frac{tf}{K}\right)} \frac{\frac{tf}{K}}{\left(1 + \frac{tf}{K}\right)} = p_T \cdot \frac{tfp}{p_T} \cdot \frac{ptf}{p_T}. \quad (2.53)$$

Thus the saturation of the promoter with both TF's equals the product of the saturation of the individual binding sites! Because, the saturation of the single sites equals (this we derived above),

$$ptf = tfp = p_T \frac{\frac{tf}{K}}{1 + \frac{tf}{K}} \quad (2.54)$$

This illustrates an interesting problem of promoter saturation with transcription factors and how this can be overcome by positive cooperativity. In the absence of cooperativity if both sites are saturated by 50% - thus  $tfp/p_T$  and  $ptf/p_T$  are both equal to 0.5 - then the total saturation of promoter is  $tfptf/p_T = 0.5^2 = 0.25$ ! However, with positive cooperativity,  $\alpha < 1$ , this saturation reduction can be overcome and a much higher sensitivity to the transcription factor concentration can be achieved. Again two forms of sensitivity can be defined here: one related to the slope of the curve and the other related to the transcription factors concentration that gives rise to 50% promoter saturation.

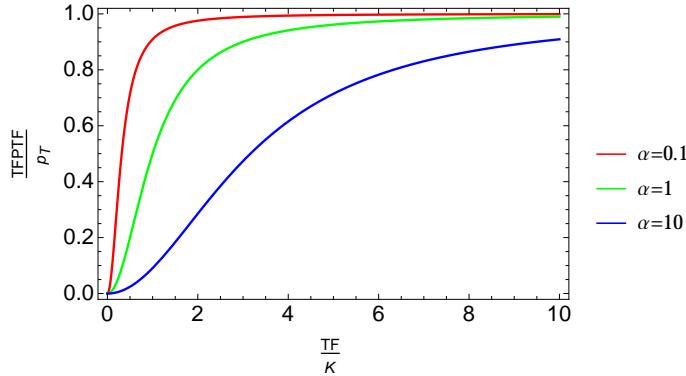
### Exercise

Consider the previous section.

1. How are the dissociation constants defined in terms of rate constants?
2. Why do we need to introduce the  $\alpha$  when the regulatory sites are different or when the transcription factor can interact on the DNA?
3. Make the derivation for  $tfptf$  (equation 2.49) yourself.
4. Plot the concentration of  $tfptf$  as function of the transcription factor concentration. Investigate the influence of  $K_1$ ,  $K_2$ , and  $\alpha$ .
5. What does  $K_1 < K_2$  indicate?
6. What does  $\alpha < 1$  indicate?

---

<sup>2</sup>Note that the "K" in the previous equation will be smaller.



**Figure 2.6: Influence of positive, negative and no cooperativity of transcription factor binding to DNA on saturation of DNA with transcription factors.** Positive cooperativity corresponds to the case where  $\alpha = 0.1$ , negative cooperativity when  $\alpha = 10$  and when  $\alpha = 1$  no cooperativity occurs.

Here  $\frac{tfptf}{p_T} = \frac{\frac{t_f^2}{\alpha K^2}}{1 + \frac{t_f^2}{\alpha K^2}}$  is plotted for different values of  $\alpha$ .

### 2.11.3 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 2.7. 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,

$$\begin{aligned}\frac{dmRNA}{dt} &= \frac{k_{sm}}{1 + TF^n} - k_{dm}mRNA \\ \frac{dTf}{dt} &= k_{stf}mRNA - k_{dtf}TF\end{aligned}\quad (2.55)$$

The first mass balance considers 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. When  $n > 0$  inhibition occurs; we will consider  $n$  as an integer.

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

$$0 = \frac{k_{sm}}{1 + \left(\frac{k_{stf}mRNAs}{k_{dtf}}\right)^n} - k_{dm}mRNAs \quad (2.56)$$

(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,

$$k_{sm} = k_{dm}mRNA_S + \left( \frac{k_{stf}}{k_{dtf}} \right)^n k_{dm}mRNA_S^{n+1} \quad (2.57)$$

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

$$mRNA_S = \left( \frac{k_{sm}}{k_{dm}} \right)^{\frac{1}{n+1}} \left( \frac{k_{dtf}}{k_{stf}} \right)^{\frac{n}{n+1}} \quad (2.58)$$

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

$$\frac{\partial mRNA}{\partial k_{sm}} \frac{k_{sm}}{mRNA} = \frac{\partial \ln mRNA}{\partial \ln k_{sm}} = \frac{1}{1+n} \quad (2.59)$$

This equation indicates that a 1% change in the transcription rate, due to some other factor not modeled here, will give rise to  $\frac{1}{n+1}\%$  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  $k_{dm}$ . Higher values of  $n$  makes the system gradually more sensitive to change in  $k_{stf}$  and  $k_{dtf}$ ,

$$\frac{\partial \ln mRNA}{\partial \ln k_{stf}} = \frac{\partial \ln mRNA}{\partial \ln k_{dtf}} = \frac{n}{1+n} \quad (2.60)$$

This sensitivity ranges between 0.5 and 1.

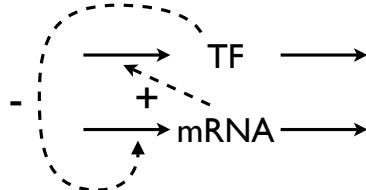


Figure 2.7: A transcription factor encoding gene that is inhibited by its own protein product.

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,

$$\frac{dmRNA}{dt} = \frac{k_{sm}}{1 + \left( \frac{k_{stf}mRNA}{k_{dtf}} \right)^n} - k_{dm}mRNA \quad (2.61)$$

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,

$$\frac{k_{sm}}{1 + \left(\frac{k_{stf}mRNA}{k_{dtf}}\right)^n} = k_{dm}mRNA = J$$

$$k_{sm} = k_{dm}mRNA = J \quad (2.62)$$

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 respond faster to a sudden increase in transcription activity!

## 2.12 Example networks: G-protein coupled receptor signalling

To illustrate what you can do with the principles you have learned in the last sections, we will construct the mass balances and mass-action kinetics for a model of G-protein coupled receptor (GPCR) activation and the subsequent activation of the associated G-protein.<sup>3</sup> The human genome encodes about 800 GPCRs and is therefore an important signalling mechanism for cells. GPCRs are found across the eukaryotic domain of life. GPCRs are typically active as dimers, as shown in Figure 2.8.

In Figure 2.8, the main biochemical reactions involved in receptor dimerisation, receptor activation, and receptor activation of a trimeric G-protein are shown. Each of these reactions has been assigned a number and we will determine the mass-action kinetics for all of them. In addition, the concentrations of all the different proteins and protein complexes will change over time determined by the rates of those reactions. We will first determine the mass action kinetics rate equation for each of the reactions. Then we will determine the mass balances. And finally we will identify the total concentrations of signalling proteins that remain fixed over time. We will consider *GTP*, *GDP*, *s* (the signal) and *Pi* at fixed concentrations; effectively we assume that these are continuously kept constant by the metabolism of the cell.

---

<sup>3</sup>If you are not familiar with G-proteins and receptors then check Wikipedia or a textbook; in principle the text here is sufficient, so only consult Wikipedia or a textbook when you think this is necessary.

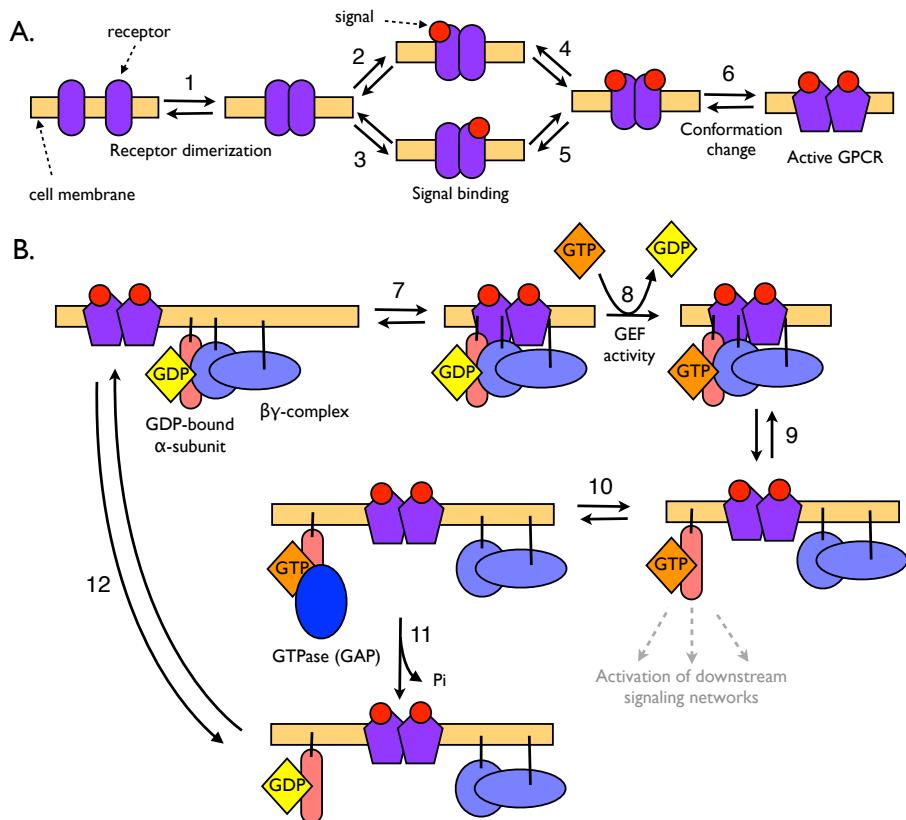


Figure 2.8: Activation by G-protein coupled receptor (GPCR) by signal binding induced conformation change of a receptor dimer and the subsequent activation of a trimeric G-protein by the receptor, via activation of a guanine exchange factor (GEF) activity, and the inactivation of the trimeric G-protein by a GTPase (GAP). A. Receptor dimerisation, signal binding, and receptor dimer conformation change. B. Activation of a trimeric G-protein, composed out of three subunits,  $\alpha$ ,  $\beta$ , and  $\gamma$ . These proteins membrane bound do not pass the membrane. We consider the  $\beta\gamma$  complex as a stable complex that does not dissociate.

Following the principles you have learned you can verify that the following

mass-action kinetic equation apply to Figure 2.8A,

$$\begin{aligned}
 v_1 &= k_1^+ \cdot r^2 - k_1^- \cdot r_2 \\
 v_2 &= k_2^+ \cdot r_2 \cdot s - k_2^- \cdot sr_2 \\
 v_3 &= k_3^+ \cdot r_2 \cdot s - k_3^- \cdot r_2 s \\
 v_4 &= k_4^+ \cdot sr_2 \cdot s - k_4^- \cdot r_2 s_2 \\
 v_5 &= k_5^+ \cdot r_2 s \cdot s - k_5^- \cdot r_2 s_2 \\
 v_6 &= k_6^+ \cdot r_2 s_2 - k_6^- \cdot \bar{r}_2 s_2
 \end{aligned} \tag{2.63}$$

Here we denote the receptor concentration by  $r$ , the dimer concentration by  $r_2$ , the signal, e.g. a growth factor, chemical or cytokine, by  $s$ , and the dimer forms by  $r_2s$ ,  $sr_2$ ,  $r_2s_2$  and  $\bar{r}_2s_2$ . The last concentration denotes the concentration of the active conformation of the receptor that activates the  $\alpha$  subunit of the G-protein.

The mass balances have to be derived for all the variable concentrations in the process. Following the principles you have learned in the last section, we obtain:

$$\begin{aligned}
 \frac{dr}{dt} &= -2v_1 \\
 \frac{dr_2}{dt} &= v_1 - v_2 - v_3 \\
 \frac{dsr_2}{dt} &= v_2 - v_4 \\
 \frac{dr_2s}{dt} &= v_3 - v_5 \\
 \frac{dr_2s_2}{dt} &= v_4 + v_5 - v_6 \\
 \frac{d\bar{r}_2s_2}{dt} &= v_6
 \end{aligned} \tag{2.64}$$

The total amount of receptor remains fixed over time, as no new receptors are added to the system (e.g. by membrane insertion) nor are any receptors removed (e.g. by endocytosis). Thus the following relationship must be correct,

$$r_T = r + 2r_2 + 2r_2s + 2sr_2 + 2r_2s_2 + 2\bar{r}_2s_2 \tag{2.65}$$

Here we summed the concentrations of the receptor containing species and multiplied them with the number of receptor molecules per species. This is the same as saying that you have a total amount of euros made out of euros, dimes, and quarters: *total euros = euro coins + 0.1 dime coins + 0.25 quarter coins*.

We can derive verify the previous relationship in the following manner. Because the total receptor remains constant over time the following must be correct as well,

$$\frac{dr_T}{dt} = 0 = \frac{dr}{dt} + 2\frac{dr_2}{dt} + 2\frac{r_2s}{dt} + 2\frac{dsr_2}{dt} + 2\frac{dr_2s_2}{dt} + 2\frac{d\bar{r}_2s_2}{dt} \tag{2.66}$$

This we can verify by substituting the mass balances (equation 2.64) into the last equation,

$$\begin{aligned} 0 &= \frac{dr}{dt} + 2\frac{dr_2}{dt} + 2\frac{r_2 s}{dt} + 2\frac{ds r_2}{dt} + 2\frac{dr_2 s_2}{dt} + 2\frac{d\bar{r}_2 \bar{s}_2}{dt} \\ &= -2v_1 + 2(v_1 - v_2 - v_3 + v_2 - v_4 + v_3 - v_5 + v_4 + v_5 - v_6 + v_6) \\ &= 0 \end{aligned} \quad (2.67)$$

This shows that equation 2.66 is correct!

If we would know the values for the concentrations of the protein concentration at some time point and the values for all the elementary rates constants, the  $k$ 's, we would be able to compute the change of the concentrations of the signalling intermediates over time, using a mathematical algorithm for solving the differential equations (equations 2.64). This we shall do later.

These is one more interesting aspect left regarding this model. This has to do with the equilibrium state it settles in after having had enough time at a constant concentration of the signal. It can be mathematically proven that this system will evolve over time to a state where all the concentrations remain fixed.<sup>4</sup> When all concentrations have become constant then all the rates of change, the  $\frac{d}{dt}$ 's, have become zero. As a consequence, we have:

$$\begin{aligned} v_1 &= 0 \\ v_2 &= v_3 = v_4 = v_5 = 0 \\ v_6 &= 0 \end{aligned} \quad (2.68)$$

Thermodynamics dictates that  $v_2$ ,  $v_3$ ,  $v_4$ , and  $v_5$  also need to be zero.<sup>5</sup> When all these rates are zero then the dependency of the concentration of the active receptor state,  $\bar{r}_2 \bar{s}_2$ , on the signal concentration can be easily calculated by

---

<sup>4</sup>You will learn how to prove this yourself in a later chapter; this proof will not be carried out here.

<sup>5</sup>You can skip this: This requires knowledge of microscopic reversibility which you do not have yet. This principle states that the product of equilibrium constant of reaction 2 and 4 equals the product of the equilibrium constants of reaction 3 and 5. Here the rates of reaction 2-5 have to be defined as positive when they run from left to right. The thermodynamic driving force of the reaction route via 2 and 4 then equals  $RT \ln \frac{\bar{r}_2 \bar{s}_2}{r_2 K_2 K_4}$  and for route via 3 and 5:  $RT \ln \frac{r_2 s_2}{r_2 K_3 K_5}$ ; this can only be the same if the system carries no net reaction rates, so:  $v_2 = v_3 = v_4 = v_5 = 0$ .

inspection of equations 2.64 and 2.63,

$$\begin{aligned}
 r_2 &= \frac{k_1^+ r^2}{k_1^-} \\
 sr_2 &= \frac{k_2^+ \cdot r_2 \cdot s}{k_2^-} = \frac{k_1^+ k_2^+}{k_1^- k_2^-} \cdot s \cdot r^2 \\
 r_2 s &= \frac{k_3^+ \cdot r_2 \cdot s}{k_3^-} = \frac{k_1^+ k_3^+}{k_1^- k_3^-} \cdot s \cdot r^2 \\
 r_2 s_2 &= \frac{k_4^+ \cdot sr_2 \cdot s}{k_4^-} = \frac{k_1^+ k_2^+ k_4^+}{k_1^- k_2^- k_4^-} \cdot s^2 \cdot r^2 \\
 \overline{r_2 s_2} &= \frac{k_6^+ r_2 s_2}{k_6^-} = \frac{k_1^+ k_2^+ k_4^+ k_6^+}{k_1^- k_2^- k_4^- k_6^-} \cdot s^2 \cdot r^2
 \end{aligned} \tag{2.69}$$

Substitution of all these relations into equation 2.65 gives an equation in terms of the constants,  $r_T$  and all the  $k$ 's, and the concentration of the monomeric form of the receptor,  $r$ .<sup>6</sup> This is quite a horrible equation to look at but can be easily obtained using the software package Mathematica or else by hand:

$$r = \frac{-1 + \sqrt{1 + 8K_1 r_T + 8K_1 K_2 r_{TS} + 8K_1 K_3 r_{TS} + 8K_1 K_2 K_4 r_{TS}^2 + 8K_1 K_2 K_4 K_6 r_T s^2}}{4(K_1 + K_1 K_2 s + K_1 K_3 s + K_1 K_2 K_4 s^2 + K_1 K_2 K_4 K_6 s^2)} \tag{2.70}$$

Here  $K_i$  is defined as  $k_i^+ / k_i^-$ . With this equation we can also express the active concentration of the receptor as function of the signal concentration by substituting the previous equation into this one – which we derived above;

$$\overline{r_2 s_2} = \frac{k_1^+ k_2^+ k_4^+ k_6^+}{k_1^- k_2^- k_4^- k_6^-} \cdot s^2 \cdot r^2 \tag{2.71}$$

This equation now only depends on parameters: the  $K$ 's,  $s$ , and  $r_T$ . When we take the limit of the resulting equation with respect to  $s$  going to infinity – so we determine the maximal level of the active receptor state – we obtain for the maximal concentration of  $\overline{r_2 s_2}$ ,

$$(\overline{r_2 s_2})_{MAX} = \frac{r_T}{2} \frac{K_6}{1 + K_6} \tag{2.72}$$

With  $K_6 = k_6^+ / k_6^-$ . This equation indicates that only  $K_6$  determines the maximal active receptor fraction, i.e.  $\frac{(\overline{r_2 s_2})_{MAX}}{r_T}$ .

In Figure 2.9, we plot the dependency of the concentration of the active receptor state on the concentration of the signal. This figure indicates that the saturation curve is slightly sigmoidal and that, for these parameter settings maximally 50% receptor pool can be activated, which fits with equation 2.72. This relationship is an important characterisation of the signalling process because it describes how the signal is "transduced" into its intracellular form at the level of the membrane. It characterises how:

---

<sup>6</sup>If you wonder about this, then try this yourself. For instance in Mathematica or Maple.

- sensitive the receptor is to the signal, which is quantified by the slope of this curve,
- the region where the cell is sensitive to the signal, i.e. where the receptor concentration changes as function of the signal level, and
- much the signal is amplified or attenuated in amount; receptor concentration units versus signal concentration units. For instance, if the receptor level is in  $\mu M$  and the signal level is in  $mM$  then the cell tunes down the signal to more "usable" concentrations - for instance because the cell membrane can only accommodate a finite amount of receptors.

The receptor model we have derived here is a bit simplified; in the chapter on enzyme kinetics we will discuss the Monod-Wyman Changeux equation, which is also frequently used to model GPCRs.

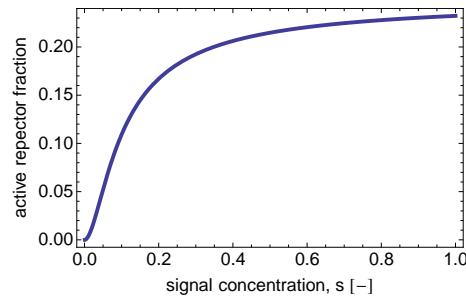


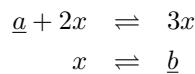
Figure 2.9: **Dependency of active fraction of the receptor,  $\frac{(r_2 s_2)_{MAX}}{r_T}$ , on the signal concentration.** This curve was calculated with equation 2.71 (with equation 2.70 substituted) and kinetic parameters:  $K_1 = K_2 = K_4 = 10$  and  $K_3 = K_6 = 1$  (all K's are in unit  $mM^{-1}$ ).

### Exercises

1. Determine the mass-action kinetics rate equations for all the reaction in Figure 2.8B,
2. Determine the mass balances for all the variable concentrations in Figure 2.8B,
3. Determine the total concentrations of signalling proteins that remain fixed over time in Figure 2.8B

### Exercises

1. Consider the following reactions,



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 3

# The remarkable and emergent functionality of small molecular circuits

### 3.1 Small, functional molecular systems are continuously exploited by cells

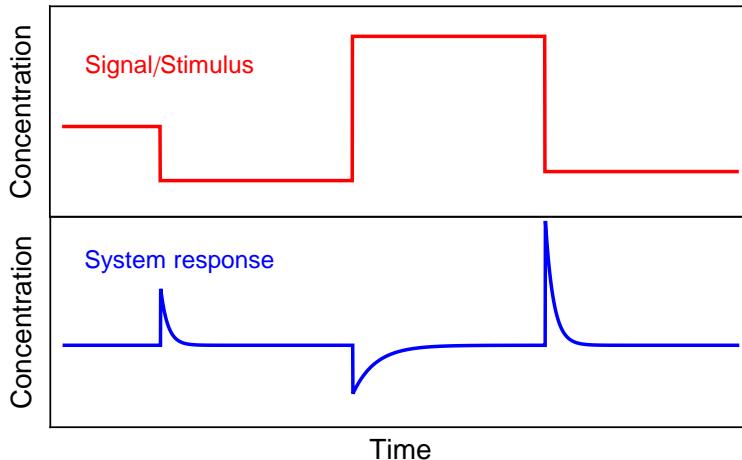
Cells are continuously confronted with changes in their environment. To some of them they have to be robust and maintain the cellular state and to others, cells have to respond adaptively and change their molecular make up - rewire signalling and molecular networks - to maintain functionality and fitness.

Robustness of physiology is important to be able to cope with sudden stresses, such as changes in temperature – which reduces protein stability – and changes in osmotic pressure – which can cause cells to explode or implode. Robustness requires cells to sense the stress and induce a response that reduces the negative influence on cell behaviour. For instance, chaperones are expressed when temperature increases to facilitate protein folding and protein stability or specific metabolites, osmolites, are accumulated intracellularly to compensate for enhanced extracellular osmotic pressure and reduce cellular water loss. When those responses are carried out, cells can continue functioning. Typically, these stress adaptations consume a lot of resources which prevent the cell from returning to its previous state of activity and, hence, the response should be shut off when the stress disappears.

Adaptation responses are properties of molecular networks - and not of single molecules; specific molecular control circuitry is required to achieve cellular adaptations. Molecular systems composed out of a small number of molecular constituents and their interactions can already give rise to sophisticated and, often, unexpected behaviours. Those behaviours are truly network behaviours; they cannot be explained in terms of a single molecular property but instead

depend in a nonlinear and nonintuitive manner on molecular properties, and typically on all molecular properties of the network. Such network properties are called emergent properties. Cells therefore function due to the emergent properties of molecular systems.

In this chapter, a number of mathematical models of small molecular systems, involved in signalling, metabolism, and gene expression control, will be introduced and explained in detail. In principle, basic knowledge of mass-action kinetics and mass balances will be sufficient to understand how the basic network functionalities for cells, studied in this chapter, emerge from molecular interactions. We will limit ourselves to basic mathematical models of the underlying biochemistry of the systems with functional properties. Our interest will be mostly in highlighting the principles of a class of molecular systems rather than in specific instances of such systems; in their full blown biochemical detail. Those details often confuse us and prevent us from appreciating the basic molecular control logic that is minimally required to give rise to a specific cellular functionality. These models offer a sensible start to study specific systems in more detail.



**Figure 3.1: A canonical perfect adaptation response of a molecular system.** A signal (or stimulus) of a molecular system is changed in a stepwise manner and the system shows a transient response and returns to a basal value that is independent of the signal level.

### 3.2 Example 1: Enzyme saturation as a mechanism for robustness

Robustness is defined as the independence of a steady-state concentration, flux or any function thereof with respect to one or more parameters. The temperature in your room is robust with respect to outside temperature if you have an air-conditioner linked to a thermostat. So, robustness is system property that you are actually quite familiar with. It results from the action of a sensor and controller. Molecular systems can also be robust and it will be shown in this section that we can view such systems in exactly the same way as you would view a thermostat. We consider an example discussed in Drengstig et al [10],

$$\begin{aligned}\frac{da}{dt} &= k_1 - k_2 \cdot a + k_3 \cdot e \\ \frac{de}{dt} &= k_4 - k_5 \cdot a \cdot \frac{e}{K_5 + e}\end{aligned}\quad (3.1)$$

(Draw the corresponding network yourself to obtain a network perspective on the system.) We would like to make the steady state concentration of  $A$ , i.e.  $a_s$ , robust with respect to parameters  $k_1$ ,  $k_2$ , and  $k_3$ . The question is how to do this. So, let's first write down the steady state equations,

$$0 = k_1 - k_2 \cdot a_s + k_3 \cdot e_s \quad (3.2)$$

$$0 = k_4 - k_5 \cdot a_s \cdot \frac{e_s}{K_5 + e_s} \quad (3.3)$$

If you would solve these two equations for  $a_s$  and  $e_s$  then you will find that the steady state concentrations depend on all parameters. (Use the Mathematica Solve command to verify this yourself.) To make  $a_s$  robust with respect to  $k_1$ ,  $k_2$ , and  $k_3$  then we should focus on equation 3.3 as it contains  $a_s$  and  $k_1$ ,  $k_2$ , and  $k_3$  do not appear. The problem is that it also contains  $e_s$  and therefore also equation 3.2 is required to solve for the steady state concentrations. As a result  $a_s$  will depend on all parameters. However, if we would be able to remove the dependency on  $e_s$  in equation 3.3 then we would have solved the problem: as we would have then obtained a single equation that solves  $a_s$  in terms of at most the kinetic parameters  $k_4$ ,  $k_5$  and  $K_5$  and this relation would not refer to  $k_1$ ,  $k_2$ , and  $k_3$ . If we choose  $K_5$  very small such that in most states of the system  $K_5 \ll e_5$  then  $\frac{e_5}{K_5 + e_5} \approx 1$  and we would have for equation 3.3,

$$0 \approx k_4 - k_5 \cdot a_s \Rightarrow a_s = \frac{k_4}{k_5} = a_{setpoint} \quad (3.4)$$

Hence, we can rewrite the dynamic system given the assumption  $K_5 \ll e_5$  to give rise to,

$$\begin{aligned}\frac{da}{dt} &= k_1 - k_2 \cdot a + k_3 \cdot e \\ \frac{de}{dt} &= k_5 (a_{setpoint} - a)\end{aligned}\quad (3.5)$$

So, this means that the system always goes to value  $a_{setpoint}$  for  $a$  at steady state. Therefore we can view  $a_{setpoint}$  as the setpoint value of  $a$ ;  $a_{setpoint}$  plays the same role as the set temperature on your thermostat device. The other remarkable thing is that the rate of change of  $e$ , i.e.  $\frac{de}{dt}$ , solely depends on  $a$  and only implicitly on  $e$ .<sup>1</sup> Figure 3.2 shows some example dynamics of the system and indicates that a reduction of  $K_5$  indeed improves robustness.

How does it now work? So, upon a change in either  $k_1$ ,  $k_2$ , or  $k_3$  the system returns always to  $a_s$  this must mean that  $e$  attains some value in equation 3.2 such that  $a = a_{setpoint}$ . Therefore,  $e$  always compensates for the deviation of  $a$  from  $a_{setpoint}$  induced by a perturbation of  $k_1$ ,  $k_2$ , or  $k_3$  such that eventually  $a \rightarrow a_{setpoint}$ . This is the function of  $E$ ; to make the concentration of  $A$  at steady state robust with respect to changes in  $k_1$ ,  $k_2$ , and  $k_3$ .

### 3.3 Example 2: Perfect adaptation of chemotaxis

The first realistic mathematical models we will study are minimal representations of a class of molecular systems that display perfect adaptation. Perfect adaptation is defined as the ability of a molecular system to display a transient response to a stepwise increase in a stimulus and subsequently return to the state prior to the stimulus. The main consequence of this behaviour is that the system responds to changes in its environment and not to the absolute value of the extracellular stimulus. This is a remarkable property and very useful for cells. This behaviour is shown graphically in Figure 3.1. This section is partially based on the paper by Yu & Doyle [42].

Figure 3.1 immediately indicates the mathematical condition for perfect adaptation to take place. The *steady state* value of the variable, e.g. a reaction rate or molecule concentration, that displays the perfect adaptation behaviour should not depend on the signal level. Only then will this variable always have the same value at steady states. This may sound obvious but how can molecular systems accomplish this?

A daily life example of perfect adaptation is the thermostat of the heating system at home. Upon a decrease in outside temperature (external condition) or opening of a window, the thermostat will switch on the heaters until the room is at the desired temperature again. In this way, the room is maintained at a desired temperature and the temperature displays perfect adaptation despite changes from the outside.

The best known example of the perfect adaptation strategy in biology functions in bacterial chemotaxis, the ability of bacteria to move to regions of higher chemoattractant or lower chemorepellant concentration. Bacteria do this by a so called biased random walk; They move for some time in a direction, then “tumble”, and then move in a different, random, direction. An increase in

---

<sup>1</sup>It is interesting to figure out what characterises the response time of the system to reach the setpoint value of  $a$ . Is it  $1/k_5$  or not? You could use Mathematica and figure this out.

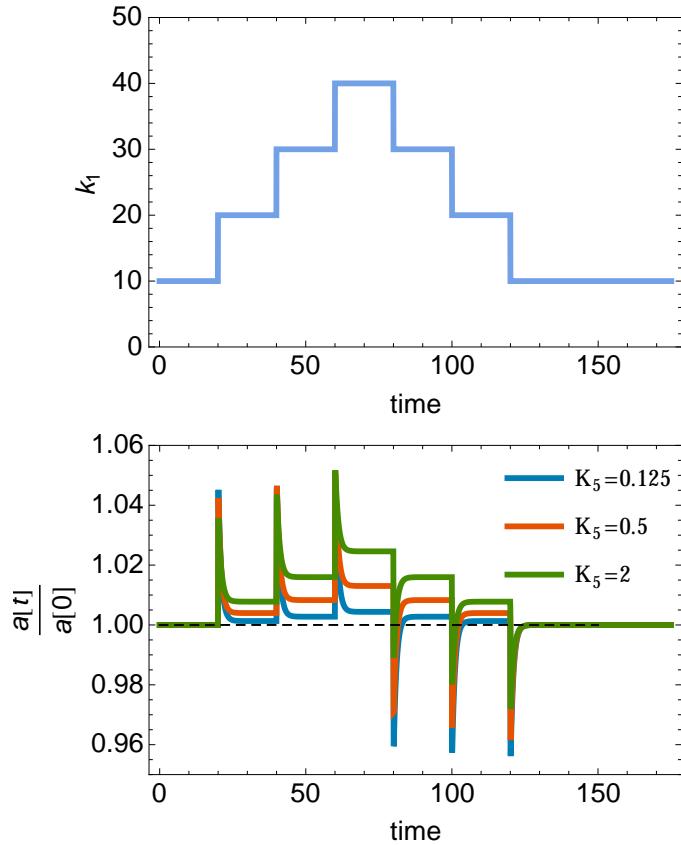


Figure 3.2: **Robustness improves when  $K_5$  is decreased.** The parameters were set to  $k_2 = 40$ ,  $k_3 = 40$ ,  $k_4 = 5$ ,  $k_5 = 1$ .

chemoattractant concentration reduces the chance to tumble. In this way, runs in a direction of increasing chemoattractant are, on average, longer and thus farther, causing the bacterium to move up the gradient. However, because nutrient levels can vary strongly in different environments, it is important that the duration of a run only depends on *changes* in concentration, not the absolute concentration. We will return to this feature in a next section.

The length of a run is, indirectly, regulated by a receptor that can bind the chemoattractant. This receptor,  $E$ , can be either in an active or in an inactive state. If this receptor is active, it phosphorylates some regulator molecules, which then can bind to the flagellar motors to induce tumbling. The amount of active receptor,  $A$ , can thus be interpreted as the output of the system. The receptor can be reversely methylated to form  $E_m$ . Only methylated receptors

can be active and their level depends on the level of the chemoattractant. The fraction of methylated receptors that is active is given by the factor  $\alpha = (1 + l/K_L)^{-1}$ , which decreases when ligand,  $L$  is bound to the receptor<sup>2</sup>.  $\alpha E_M$  is then the equilibrium amount of the methylated receptor that is not bound to the ligand,  $L$  with concentration  $l$ . We assume  $K_L$  to be equal to 1.

The network we shall be considering is shown in Figure 3.3. Methylation of the receptor,  $E$ , is done by an enzyme, CheR. Typically, CheR is always saturated with  $E$ . This means that the methylation rate is constant and denoted by  $V_{max}^R$ . Demethylation is done by the enzyme CheB, which can only demethylate active receptors; those receptors are methylated but not bound to  $L$ . The concentration of unbound receptors equals  $\frac{1}{1+l}E_M = \alpha(l)E_M$ ; here we have assumed that the binding of  $L$  to  $R$  occurs at thermodynamic equilibrium and that the  $K_D$  equals 1. We assume that CheB is operating far from saturation.<sup>3</sup> With those assumptions we obtain the following equations,

$$A = \alpha(l)E_M \quad (3.6)$$

and

$$\frac{dE_M}{dt} = \underbrace{V_{max}^R}_{\text{methylation rate}} - \underbrace{\frac{V_{max}^B}{K_b} A}_{\text{demethylation rate}} = V_{max}^R - \frac{V_{max}^B}{K_b} \alpha(l)E_M. \quad (3.7)$$

Here,  $V_{max}^R$  denotes the maximal methylation rate of  $E$  by CheR,  $V_{max}^B$  is the maximal rate at which (active)  $E_m$  can be demethylated by CheB and  $K_b$  is the Michealis-Menten constant of this reaction.

Figure 3.1 shows what happens to  $A$  when the ligand concentration is changed in a stepwise manner. We see that upon an decrease (increase) in  $l$ , there is a transient increase (decrease) in  $A$ . However, eventually  $A$  returns back to precisely its original level! To understand why this is so, let's have a look at the steady state of the system. Clearly, at steady state  $\frac{dE_M}{dt} = 0$ . This allows us to solve equation 3.9 for the steady state level of  $A$ ,  $A_S$ ;

$$A_S = \frac{V_{max}^R K_B}{V_{max}^B} = \alpha(l)E_{M,S}. \quad (3.8)$$

With  $E_{M,S}$  as the steady-state level of the methylated receptor. The steady state level of the active receptors, equals to  $\alpha(l)E_{M,S}$  is independent of the ligand concentration as it equals always  $\frac{V_{max}^R K_B}{V_{max}^B}$  in concentration! This means the system displays perfect adaptation: upon any change in the ligand concentration it will always return to the same basal activity state. This is because if  $A$  changes

---

<sup>2</sup>In reality, this system is much more complex. The receptor actually is a (stable) complex of multiple proteins, it has multiple ligand binding sites, and so on. However, the main conclusions also hold for the more complex, realistic, model.

<sup>3</sup>Generally, far from saturation the Michaelis-Menten equation simplifies:  $V_{max} \frac{x}{K_M + x} \approx \frac{V_{max}}{K_M} x$  because  $x \ll K_M$  in this concentration regime.

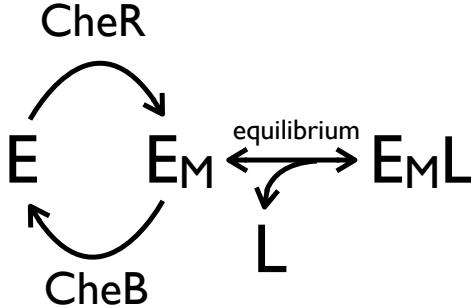


Figure 3.3: **Network scheme of chemotaxis system showing perfect adaptation.** CheR acts as the methylase of the receptor,  $E$ , yielding  $E_M$ . CheB demethylates  $E_M$ .  $E_M$  is the active form of the receptor. The receptor is active when it is not bound to the ligand  $L$ . The reaction  $E_M + L \rightleftharpoons E_M L$  is assumed to operate at thermodynamic equilibrium.

due to a change in  $l$ , this changes the rate of  $E_M$ -demethylation. Hence, the change in ligand concentration is compensated for by a change in methylation state of the receptor. Note that perfect adaptation does not depend on the details of how  $l$  affects  $A$  or how  $A$  affects the demethylation rate. Therefore, the perfect adaption of the system is called robust.

### 3.3.1 Mathematical explanation: integral feedback control

To better understand the basic requirement for perfect adaptation, we will rewrite the mass balance for  $E_M$  (equation 3.9),

$$\begin{aligned} \frac{dE_M}{dt} &= V_{max}^R - \frac{V_{max}^B}{K_B} A = \frac{V_{max}^B}{K_B} \left( \frac{K_B V_{max}^R}{V_{max}^B} - A \right) = \frac{V_{max}^B}{K_B} (A_S - A) \\ &= \frac{V_{max}^B}{K_B} (A_S - \alpha(l)E_M) \end{aligned} \quad (3.9)$$

The factor can be viewed as  $(A_S - A)$  is the "error" or the "deviation from the basal state", which needs to reduces to zero by a feedback action by the system. The feedback is the methylation of the receptor. When  $l$  increases  $\alpha(l)E_M$  decreases and  $E_M$  has to increase to make sure that  $A_S = \alpha(l)E_{M,S} = \frac{V_{max}^R K_B}{V_{max}^B}$ . We can rewrite the previous ordinary differential equation in integral form,

$$\underbrace{\Delta E_M = E_M(t_2) - E_M(t_1)}_{\text{Feedback control action}} = \underbrace{\int_{E_M(t_1)}^{E_M(t_2)} dE_M}_{\text{Time integral of deviation}} = \frac{V_{max}^B}{K_B} \int_{t_1}^{t_2} (A_S - \alpha(l)E_M) dt \quad (3.10)$$

This shows that the compensatory feedback on the system by  $E_M$  (total feedback is  $\Delta E_M$  in amount) due to a change in the ligand,  $l$ , equals the time integral of the deviation. So, the total deviation in the system induced by a change in  $l$ ,  $\Delta l = l_2 - l_1$ , is compensated for by the feedback. This is called integral feedback control and a mechanism for perfect adaptation.

### 3.3.2 Integral feedback control (the mathematical background)

The above system is an example of integral feedback control. Integral feedback is a general strategy to attain robust perfect adaptation. Generally, feedback systems are systems where the system-output is in some way influencing the state of the system itself. In this case, the system input is the ligand concentration, the state of the system depends on the methylation state of the receptor, the system output is the receptor activity, and this output is fed back to the system through the dependence of the demethylation rate on receptor activity.

In control theory, the error,  $\delta(t) = y(t) - y_S$ , is defined as the difference between a current system output,  $y(t)$ , and a desired system output,  $y_S$ . A system output could be a concentration of some molecule or a flux. If this error is kept zero at steady state by the action of the system upon environmental changes then the output is said to display perfect adaptation.

Generally, the system output,  $y(t)$ , depends on the input, typically called  $u$ , and the state of the system, called  $x(t)$ . Thus, mathematically we have  $y(t) = f(x(t), u)$ . In an integral feedback system, the time integral of the  $\delta(t)$  equals the feedback action, i.e.  $\Delta x = c \int_0^\infty \delta(t) dt$ , or, equivalently,  $\frac{dx}{dt} = c \cdot \delta(t)$ , where  $c$  is a constant. In the steady state,  $\delta = 1/c \frac{dx}{dt} = 0$ ; which indicates that the system can only be in steady state if the error is zero!

In our example, above we saw that the error  $\delta = A - A_S = A - \frac{V_{max}^R K_B}{V_{max}^B}$ . If we put this in equation (3.9), we see that  $\frac{dE_M}{dt} = -\frac{V_{max}^B}{K_B} \delta = -\frac{V_{max}^B}{K_B} (A - A_S) = -\frac{V_{max}^B}{K_B} (\alpha(l)E_M - A_S)$ . This is exactly the integral feedback condition and the system will always return to  $A_S$  at steady state! This is because the methylation state of the receptor compensates the ligand influence on the activity level of the receptor. Hence, we find at steady state that the steady state level of the methylation of the receptor equals  $E_{M,S} = \frac{A_S}{\alpha(l)} = A_S(1 + l)$  and increases linearly with the ligand level,  $l$ . Integral feedback control is a very widely used strategy in engineering. Here, we have seen that bacteria also use this strategy. We could say that bacteria "are very clever engineers"!

## Exercises

1. Show for the models in figure 3.4 that they each show perfect adaptation; identify the variables that are robust and the parameters, or reactions, to which they are robust.

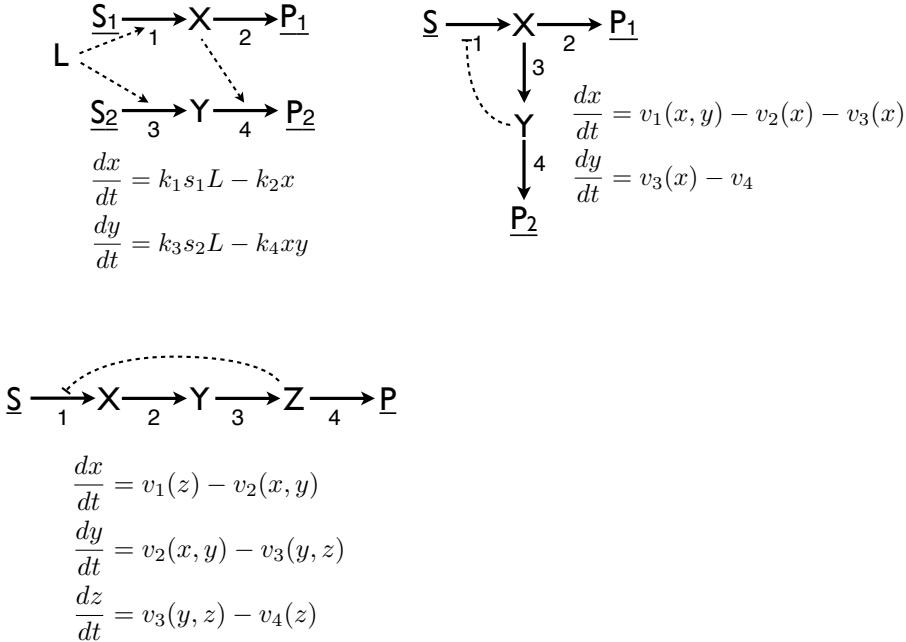


Figure 3.4: Three systems displaying perfect adaptation

### 3.4 Example 3: Fold change detection

This section is based on the paper by Goentoro, Shoval, Kirschner, and Alon [13]. Another dynamic response property that can be of functional importance is fold change detection. This is a mechanism that senses the *fold* change in an input and not the absolute level (change). Most systems would display different responses when their input changes from 5 to 10 or from 20 to 40 because the absolute levels of the signal input is different. A fold change detection system gives the same dynamic response because in both case the fold change is 2, i.e. 10/5 and 40/20. The question is what is minimally required to observe fold change detection? And, what are the examples in cell biology of fold change detection systems?

In Figure 3.5 a small regulatory network is shown composed out of the three regulatory proteins,  $X$ ,  $Y$ , and  $Z$ .  $X$  activates  $Y$ , which in turn inactivates  $Z$  and  $X$  also directly activates  $Z$ .  $X$ 's influence via  $Y$  introduces a delay such that  $Z$  first increases in time, due to the direct stimulatory influence of  $X$  on  $Z$ , and then, when  $Y$ 's activation by  $X$  kicks in,  $Z$  is inhibited by  $Y$ . This delayed inactivation of  $Z$  causes it to display a peaked response. The dynamics of  $Z$  shows two features: i. perfect adaptation (discussed in the previous section) and ii. fold change detection because for the same fold change in the signal,  $X$ , the dynamics of  $Z$  is exactly the same even though the basal level of  $X$  varies.

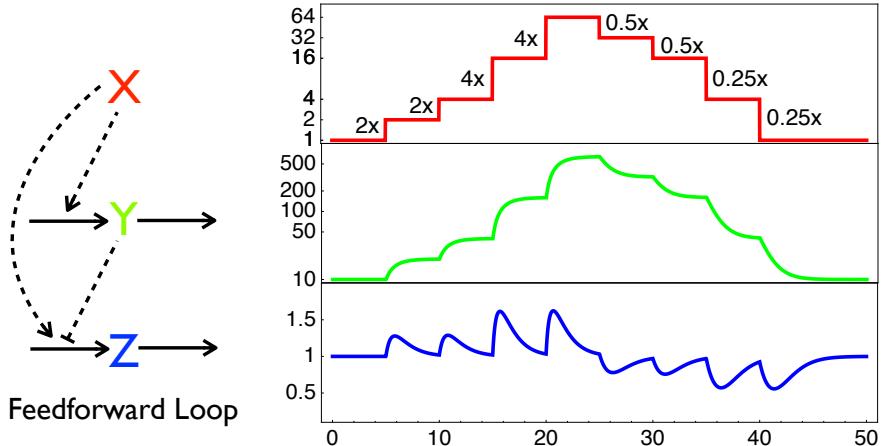


Figure 3.5: **A feedforward loop network that displays fold change detection.** The signal,  $X$ , activates  $Z$  directly and inhibits it indirectly via its activation of  $Y$  and the subsequent inhibition of  $Z$  by  $Y$ . The model is explained in the main text (equation 3.11). Kinetic parameters:  $\alpha_1 = \alpha_2 = 1$  and  $\beta_1 = \beta_2 = 10$ .

An inspection of the mass balances of this system indicates how fold change detection works. The mass balances are given by,

$$\begin{aligned} \frac{dy}{dt} &= \beta_1 x - \alpha_1 y \\ \frac{dz}{dt} &= \beta_2 \frac{x}{y} - \alpha_2 z \end{aligned} \quad (3.11)$$

Let's first confirm that this system displays perfect adaptation; at steady state we find that:

$$\begin{aligned} y_s &= \frac{\beta_1 x}{\alpha_1} \\ z_s &= \frac{\beta_2 x}{\alpha_2 y_s} = \frac{\beta_2 \alpha_1}{\beta_1 \alpha_2} \end{aligned} \quad (3.12)$$

These relations indicate that  $z_s$  does not depend on the level on  $X$  and therefore perfect adaptation occurs. In order to show that this system is also capable of fold change detection we have to show that the *dynamics* of  $Z$  only depends on the fold change in  $X$  and not on the basal level of  $X$ . This suggests that the differential equation of  $Z$  can be written in terms of the fold change in  $X$  as a

parameter. We define for all the variables the following fold changes,

$$\begin{aligned} F_x &= \frac{x}{x_0} \\ F_y &= \frac{y}{y_0} = \frac{y}{\frac{\beta_1 x_0}{\alpha_1}} \\ F_z &= \frac{z}{z_0} = \frac{z}{\frac{\beta_2 \alpha_1}{\beta_1 \alpha_2}} \end{aligned} \quad (3.13)$$

Thus,  $(y_0, z_0)$  is the steady state at signal level  $x_0$  and the reference state from where  $X$  is changed from  $x_0 \rightarrow x$ . Now we can rewrite the differential equation in terms of those fold change variables,

$$\begin{aligned} \frac{dF_y}{dt} &= \beta_1 \frac{x_0}{y_0} F_x - \alpha_1 F_y = \alpha_1 F_x - \alpha_1 F_y \\ \frac{dF_z}{dt} &= \beta_2 \frac{x_0}{z_0 y_0} \frac{F_x}{F_y} - \alpha_2 F_z = \alpha_2 \frac{F_x}{F_y} - \alpha_2 F_z \end{aligned} \quad (3.14)$$

The  $\alpha$  constants have unit 1/time and multiplication of  $t$  with  $\alpha_1$  therefore gives a dimensionless time  $\tau = \alpha_1 t$ ,

$$\begin{aligned} \frac{dF_y}{d\tau} &= F_x - F_y \\ \frac{\alpha_1}{\alpha_2} \frac{dF_z}{d\tau} &= \frac{F_x}{F_y} - F_z \end{aligned} \quad (3.15)$$

These equations indicate that all the dynamics of the system can be written in terms of the fold change of  $X$  and, hence, this is the "true" parameter of the system that determines the dynamics and not the reference level of  $X$ ,  $x_0$ . At steady state for this system we see that the fold change in  $Y$  equals the fold change in  $X$ :  $F_y = F_x$  and that the fold change in  $Z$ ,  $F_z$  equals 1 - indicating perfect adaptation.

### Exercises

1. Will two fold change detection systems in series give rise to fold change detection of the output of the second system with respect to the signal of the first?

## 3.5 Example 4: Kinetic proofreading

During many cellular processes mistakes can occur. For instance, the wrong nucleotide can be introduced during DNA repair or replication, the wrong amino acids can be incorporated in the polypeptide during translation, or "self" is mistaken for "alien" by the immune system. Cells exploit a specific error reduction

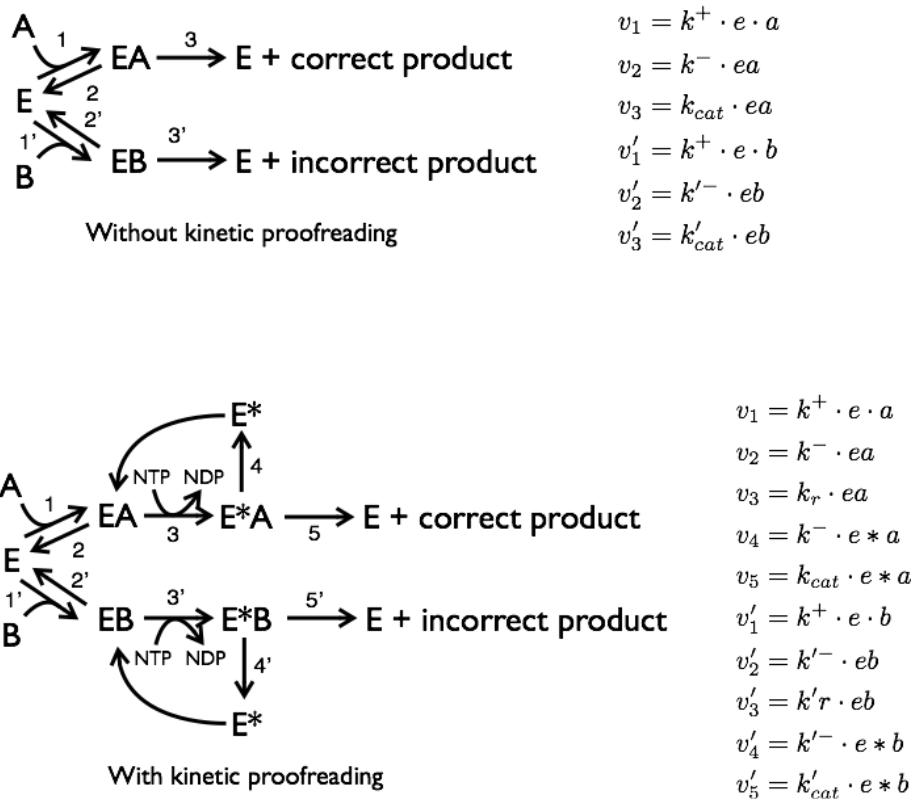


Figure 3.6: The kinetic proofreading mechanism for discrimination between correct and incorrect substrates. A correct substrate,  $A$ , and incorrect substrate,  $B$ , can both bind to an enzyme,  $E$ , which produces a product from the enzyme-substrate complexes,  $EA$  and  $EB$ . However, from  $EB$  the wrong product is made because  $A$  should have bound. Kinetic proofreading can reduce the ratio of the incorrect over correct product formation flux. It requires the introduction of an irreversible reaction. The irreversibility of this reaction is achieved by driving this reaction by a chemical free energy potential deriving from ATP/ADP, GTP/GDP, NADH/NAD, etc.  $E^*$  denotes the energised enzyme state, which can spontaneously or by the action of an enzyme return to the ground  $E$  state.

strategy in these discrimination problem, which is called kinetic proofreading. This section is based on the paper by Hopfield [21].

Before kinetic proofreading is explained we will look more carefully into the situation without this mechanism active. A relevant model for the discrimination problem is shown in Figure 3.6. An enzyme is shown that is supposed to make a product when  $A$  binds to it but  $B$ , the incorrect competitive substrate for the enzyme, can also bind to the enzyme and lead to the formation of an

erroneous product. The error probability (or rate) is defined by the following ratio,

$$\Phi = \frac{\text{Rate of incorrect product formation}}{\text{Rate of correct product formation}} = \frac{k'_{cat}eb}{k_{cat}ea} \quad (3.16)$$

If the error rate is 0.1 then this means that in 10% of the product formation events an error was made. Realistic values for error in transcription and translation are below  $10^{-8}$  and sometimes even a lot lower. What determines this error rate? This is what we will analyse next.

The steady-state concentrations of  $EA$  and  $EB$  are equal to,

$$\begin{aligned} ea &= \frac{k^+a}{k^- + k_{cat}} \\ eb &= \frac{k^+b}{k'^- + k'_{cat}} \end{aligned} \quad (3.17)$$

Here we assume that the association rate constant of  $A$  and  $B$  with the enzyme are the same. Note that the enzyme will have a slightly higher affinity for  $A$  and  $B$  and, therefore,  $k'^- > k^-$ .

Thus, the error rate becomes,

$$\Phi = \frac{k'_{cat}}{k_{cat}} \left( \frac{k^+b}{k'^- + k'_{cat}} \right) \left( \frac{k^+a}{k^- + k_{cat}} \right)^{-1} \quad (3.18)$$

We make another simplification, which is realistic in many cases (e.g. rate of amino acid incorporation by ribosome):  $k_{cat} \approx k'_{cat}$ .

$$\Phi = \left( \frac{k^+b}{k'^- + k_{cat}} \right) \left( \frac{k^+a}{k^- + k_{cat}} \right)^{-1} \quad (3.19)$$

This relationship indicates that the minimal error rate is attained when  $k_{cat} \ll k'^-$  and  $k_{cat} \ll k^-$ . In fact, this also makes sense because if  $B$  binds weaker to the enzyme than  $A$ , more  $B$  will have fallen off the enzyme if the catalytic rate is slower.<sup>4</sup> Then,

$$\Phi_{min} \approx \left( \frac{k^+b}{k'^-} \right) \left( \frac{k^+a}{k^-} \right)^{-1} \quad (3.20)$$

The ratio's  $\frac{k'^-}{k^+}$  and  $\frac{k^-}{k^+}$  are equilibrium constants (dissociation constants):  $\left( \frac{e \cdot a}{ea} \right)_{eq} = \frac{k^-}{k^+} = K_A$  and  $\left( \frac{e \cdot b}{eb} \right)_{eq} = \frac{k'^-}{k^+} = K_B$ . Thus,

$$\Phi_{min} \approx \left( \frac{b}{K_B} \right) \left( \frac{a}{K_A} \right)^{-1} = \frac{b}{a} \frac{K_A}{K_B} \quad (3.21)$$

---

<sup>4</sup>But the same applies to  $A$  or does more  $B$  fall off than  $A$  because the  $EB$  is more unstable than the  $EA$  complex? Here is the answer:  $\frac{d}{dt} \frac{eb}{ea} = \frac{ea \cdot deb/dt - eb \cdot dea/dt}{ea^2} = (k^- - k'^-) \frac{eb}{ea} < 0$ . Thus, the ratio decreases with time indicating that more of  $EB$  is lost than of  $EA$ ; an estimate for the duration of time that it decreases is  $1/(kr(ea + eb))$ . Even though the ratio decreases and relatively more  $EA$  will persist than  $EB$ , the concentration of  $EA$  of course also decreases; which reduces the formation rate of correct products, and slow down the system, but at the same time the error rate is reduced.

This indicates that the minimal error rate is achieved when the catalytic rate is much slower than the association and dissociation rates and, hence, the enzyme binding is close to thermodynamic equilibrium.<sup>5</sup> Error rates of less than  $10^{-8}$ , with substrates that are very similar in chemical properties and structures, e.g. nucleotides and amino acids, (such that  $K_A \approx K_B$ ) can essentially only be achieved by  $b/a$  concentration ratio's of  $< 10^{-8}$ . But we know that concentrations of nucleotides and amino acids do not vary this much in concentrations.<sup>6</sup> The cell must therefore use a different mechanism. This mechanism is kinetic proofreading and shown in Figure 3.6.

Again we will evaluate the error performance of this system by using the error rate,  $\Phi$ , as defined above,

$$\Phi = \frac{k'_{cat}e^*b}{k_{cat}e^*a} = \frac{k'_{cat}}{k_{cat}} \frac{\frac{k_r}{k'^- + k'_{cat}}}{\frac{k_r}{k^- + k_{cat}}} \frac{\frac{k^+ \cdot e \cdot b}{k'^- + k_r}}{\frac{k^+ \cdot e \cdot a}{k^- + k_r}} \quad (3.22)$$

Note that  $e^*a = \frac{k_r ea}{k^- + k_{cat}}$ ,  $ea = \frac{k^+ \cdot e \cdot a}{k^- + k_r}$ ,  $e^*b = \frac{k_r eb}{k'^- + k'_{cat}}$ , and  $eb = \frac{k^+ \cdot e \cdot b}{k'^- + k_r}$ . The rate constant  $k_r$  belongs to the irreversible reaction. Again, if the association and dissociation reactions are fast ( $k'^- \gg k'_{cat}$ ,  $k'^- \gg k_r$ ,  $k^- \gg k_{cat}$ ,  $k'^- \gg k_r$ ) and  $k'_{cat} = k_{cat}$  then the minimal error rate becomes,

$$\Phi_{min} \approx \frac{\frac{1}{k'^-} \frac{k^+ b}{k'^-}}{\frac{1}{k^-} \frac{k^+ a}{k^-}} = \frac{\frac{1}{k'^-} \frac{k^+ b}{k'^-}}{\frac{1}{k^-} \frac{k^+ a}{k^-}} \frac{k^+}{k^+} = \frac{b}{a} \left( \frac{K_A}{K_B} \right)^2 \quad (3.23)$$

This indicates that the irreversible reaction reduces the error rate greatly. It does this by introducing a delay of  $1/(k_r \cdot (ea + eb))$  such that more  $b$  will fall off the enzyme.

### Exercises

1. Show that with two irreversible reactions the error rate becomes:

$$\Phi = \frac{b}{a} \left( \frac{K_A}{K_B} \right)^3$$

2. And with  $n$  irreversible reactions,

$$\Phi = \frac{b}{a} \left( \frac{K_A}{K_B} \right)^{n+1}$$

---

<sup>5</sup>Under those conditions, the concentration of  $EA$  and  $EB$  is  $ea = e_T \frac{a/K_A}{1+a/K_A+b/K_B}$  and  $eb = E_T \frac{b/K_B}{1+a/K_A+b/K_B}$  and  $\Phi = eb/ea$ . This is in agreement with the results of the main text. We could not have started with this result because we did not know yet that the minimal error rate is achieved with fast equilibration of the enzyme-substrate complexes.

<sup>6</sup>This would also not work because if alanine, rather than proline, is required given the current position of the ribosome on the mRNA, it can be that at the next position proline is required rather than alanine; which would require a sudden concentration ratio change of alanine and proline. In addition, many mRNA's are translated in parallel and, hence, simultaneous demands for alanine and proline can occur. So, concentration ratios cannot solve the problem.

3. What is the negative consequence of kinetic proofreading? (Hint: i. a delay is introduced with every new irreversible reaction and ii.  $A$  also dissociates during this delay time. What is the net outcome of both of these phenomena?)

### 3.5.1 Different explanation for why kinetic proofreading improves accuracy

The delay introduced by hydrolysis of NTPs causes that the concentration of the enzyme state bound to the correct molecule rises relatively to the concentration of the enzyme bound to the incorrect molecule. Say, the correct state is called  $EA$  and the wrong state  $EB$  and that  $EA$  and  $EB$  can both dissociate into  $EA \xrightarrow{k_a} E + A$  and  $EB \xrightarrow{k_b} E + B$  prior to forming the correct or incorrect product, which occurs after a delay. The delay time introduced by the addition of one NTP hydrolysis equals  $\delta t$ , such that  $n$  of those reactions introduce a delay of  $\Delta t = n \cdot \delta t$ . Then we find for the rate of the correct enzyme state over the incorrect enzyme state (for times:  $t < \Delta t$ , so prior to the production of either the correct or incorrect product),

$$\frac{ea(\Delta t)}{eb(\Delta t)} = \frac{ea(0)e^{-k_a\Delta t}}{eb(0)e^{-k_b\Delta t}} = \frac{ea(0)}{eb(0)} e^{-k_a\Delta t} e^{k_b\Delta t} = \frac{ea(0)}{eb(0)} e^{(k_b - k_a)\Delta t} \quad (3.24)$$

Since  $k_a < k_b$ ,  $\frac{ea(t)}{eb(t)}$  increases in time; so, when the delay increases the correct enzyme preferentially remains. Let's now rewrite this a bit,

$$\begin{aligned} \frac{ea(\Delta t)}{eb(\Delta t)} &= \frac{ea(0)}{eb(0)} e^{(k_b - k_a)\Delta t} = \frac{ea(0)}{eb(0)} e^{(k_b - k_a)n\delta t} \\ &= \frac{ea(0)}{eb(0)} \left( e^{(k_b - k_a)\delta t} \right)^n = \frac{ea(0)}{eb(0)} \Omega^n \end{aligned} \quad (3.25)$$

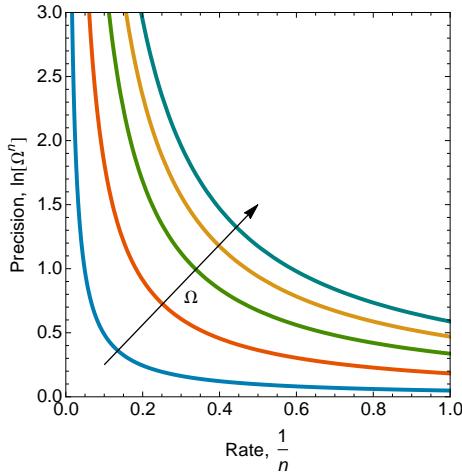
The improvement after 1 delay step of duration  $\delta t$  equals  $e^{(k_b - k_a)\delta t}$  which is a constant if  $\delta t$  is constant and we denote it by  $\Omega$ . Let's set  $\frac{ea(0)}{eb(0)}$  to 1 to simplify the analysis,

$$\frac{ea(\Delta t)}{eb(\Delta t)} = \Omega^n \quad , \text{ and } \Omega > 1 \quad (3.26)$$

The disadvantage of the delay is that rate of product formation decreases with  $n$  as,

$$\text{product formation rate} = v \propto \frac{1}{n\delta t} \quad (3.27)$$

These two equations suggest a trade off in kinetic proofreading: a higher precision is at the expense of the process rate. So, this means that  $\Omega$  can be under strong selective pressure. Which means that  $E$  evolves towards enhancement of the affinity difference for  $A$  and  $B$  such that  $k_b - k_a$  becomes larger and  $\Omega$  increases.



**Figure 3.7: Precision-rate trade off in kinetic proofreading.** The problem in kinetic proofreading is the following: the system can be made more precise by adding more delay but this occurs at the expense of the process rate. Along the lines  $n$  was changed from 1 to 100 and the lines differ in the  $\Omega$  value: 1.05, 1.2, 1.4, 1.6, and 1.8.

### 3.6 Example 5: Robustness of two component signalling networks

Two component signalling systems are the most abundant signalling systems for prokaryotes (Figure 3.8). Theory predicts that some of those systems may show a remarkable robustness property: small changes in the abundances of sensor and response regulator molecules, due to protein partitioning errors in cell division, transcription stochasticity, or other inevitable, dynamic fluctuations in protein abundance, do not influence the steady state phosphorylation degree of the response regulator. This was found by Shinar & Alon [36] and this section is based on their approach. This is a remarkable property and allows bacteria to respond more effectively to changes in the signal level and not show responses when no changes in the environment occurred but rather spontaneous changes in cellular levels of sensor and response regulator molecules. To show this robustness, we study the phosphate balance of the signaling network,

$$\frac{dp}{dt} = v_2 - v_6 = k_2 \cdot sl - k_6 rps = k_2 \frac{s \cdot l}{K_L} - k_6 rps \quad (3.28)$$

The rate of reaction 2 captures the influx of phosphate into the network and the rate of reaction 6 equals the efflux rate. We assume rapid binding of ligand to the sensor protein such that this reaction operates close to thermodynamic equilibrium and therefore  $sl \approx \frac{s \cdot l}{K_L}$ . The steady state requirement for  $RPS$

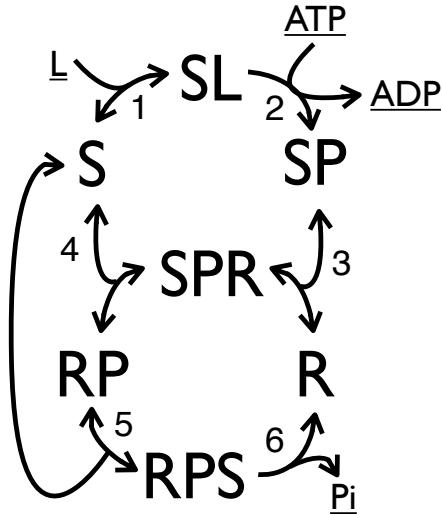


Figure 3.8: **Bacterial two component signal transduction.** Many bacteria use two component signal transduction systems to sense and initiate responses to environmental conditions. Two component signalling systems rely on just two proteins: a sensor ( $S$ ) and a response regulator ( $R$ ). The response regulation is often a transcription factor, which becomes activated by phosphorylation. Binding of a ligand (signal),  $L$ , to the sensor activates its autokinase activity leading to a phosphorylated sensor. This phosphoryl group can be transferred to a cognate response regulator. The phosphorylated response regulation,  $RP$ , can regulate transcription or become dephosphorylated by the unphosphorylated sensor. The sensor is often membrane embedded and the response regulator cytoplasmic. *Escherichia coli* exploits more than 20 of such systems.

forces  $v_5$  and  $v_6$  to be equal,

$$\begin{aligned} k_5^+ \cdot rp \cdot s - k_5^- rps &= k_6 rps \quad \Rightarrow \\ rps &= \frac{k_5^+}{k_5^- + k_6} \cdot rp \cdot s \end{aligned} \quad (3.29)$$

Substitution of this relationship in the previous steady state mass balance for phosphate gives,

$$\begin{aligned} 0 &= k_2 \frac{s \cdot l}{K_L} - k_6 \frac{k_5^+}{k_5^- + k_6} \cdot rp \cdot s \quad \Rightarrow \\ rp &= \frac{k_2(k_5^- + k_6)}{k_5^+ k_6} \frac{l}{K_L} \leq r_T \end{aligned} \quad (3.30)$$

Here  $r_T$  denotes the total response regulator concentration,  $r_T = r + rps + rp + spr$ . This is the input-output relationship of this signalling network. It is linear and does not depend on the total sensor and response regulator concentration!

Experimental confirmation of this theoretical prediction is only indirect at the moment. And, even though the robustness property is advantageous, the system has a linear input-output relationship, which does not make it very sensitive to changes in the ligand concentration. In fact, in reality the sensor and response regulators are dimeric molecules, which could allow for a quadratic dependency of  $rp$  on  $l$ . Under this condition, the robustness has been predicted to remain valid.

### Exercises

1. Is the equilibrium binding assumption for  $S$  and  $L$  required for the robustness of  $RP$  with respect to the total sensor and response regulator level?

## 3.7 Concluding remarks

The aim of this chapter was to show:

- how capable small molecular systems are in achieving quite sophisticated behaviours, which can be highly functional and fitness enhancing for cells – many more of those will follow in this book,
- how small mathematical models of molecular systems can help you understand their design and functional principles, and
- that a whole class of systems can be modelled by a minimal mathematical model that captures the essence of the principles underlying the system behaviour.

The latter two messages are very useful guidelines for modelling molecular systems, even when most of the kinetic parameters of a molecular system are known, and a detailed model can be developed, our advise to always also make a small model, or a few of them, that captures the essence(-s) of the system you are studying. Those models will always be very enlightening and often also models with minimal complexity that are easy to use and explain your findings to colleagues.

## Chapter 4

# Diffusion and reactions

### 4.1 Introduction

Diffusion of molecules inside cells lies at the basis of molecular reactions. The association of proteins into complexes, the binding of enzymes to their substrates, and protein association to DNA are basic processes in cell biology and all rely on diffusive encounters of molecules. In this chapter, we will predict some of the kinetic properties of molecular association reactions from fundamental physical principles about diffusion and introduce you to some basic methods to exploit knowledge of time and length scales of cellular processes in studying cell biology.

Collisions between molecules are probabilistic events – molecule movement is random – and the collision frequency depends on the diffusion properties of the two molecules – partially determined by the medium in which diffusion takes place – and their concentrations. After these collisions, reactions or binding associations can occur, provided that the molecules collided in the right orientation. This means that the association time can be viewed as the sum of a diffusion time and a reaction time. The latter time starts when the two molecules have collided and solely involves intramolecular processes. When the diffusion time is much slower than the reaction time, the association reaction rate is effectively diffusion limited; the association time is set by the diffusion time. In this regime, association rate constants can be estimated from the diffusional search time, which are known for many types of diffusion processes. These estimates represent the maximal value of the association rate constant, a limit dictated by physics.

### 4.2 Characteristic dimensions for cell biology

A cellular compartment, such as the cytoplasm of a cell or the interior of an organelle, is a volume packed with (macro-) molecules. (Figure 4.1 and 4.2). It is often said to be a crowded environment, because the average distance between macromolecules in a cell is roughly the dimensions of a single macromolecule

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

Table 4.1: Properties (estimates) of a bacterium, such as *E. coli*, compared to the properties of a mammalian cell (an eukaryote).

(average radius:  $\approx 5 \text{ nm}$ ). This makes the intracellular milieu very different from highly diluted solutions, such as those most often used in the lab. Macromolecular interactions can be strongly influenced by their environment.

Cells contain numerous molecules of different types having different physicochemical properties and functions; DNA, RNA, proteins, lipids, fatty acids, sugars, and low molecular weight weak acids, such as pyruvate, acetate, etc. All of this makes the intracellular milieu an environment difficult to mimic *in vitro*. This does not necessarily mean that the *in vivo* milieu will be critical for all molecular properties but certainly for some. Macromolecular crowding favours for instance the association of proteins.

It is instructive to envision cells as shown in Figure 4.2 but then with all the molecules moving erratically. This mental picture plus a number of other physicochemical properties of molecules and cells will turn out to be useful in understanding the fundamental basis of molecular reactions. Table 4.1 summarises some properties we will need throughout this chapter. It shows the extremely small volumes of cells. The number of proteins that approximately inside cells. The numbers of specific proteins/cell ranges from 1 - 10000's in bacteria and 1-1000000 in eukaryotes! So many orders of magnitude are spanned by protein concentrations. mRNA concentrations are typically 100 molecules per cell, often even below 20! Genes occur either at single copies (haploid organisms) or as two (diploids; mammals) or more copies. For instance, when *Escherichia coli* grows at maximal growth rate, with a cell division event roughly every hour, the number of genome copies per cell can reach a value of  $\sim 4$  whereas it has roughly 1 copy at low growth rates. Thus, gene number per cell does not need to be fixed.

### 4.2.1 Performing calculations with units

Throughout this book you will carry out calculations with numbers with dimensions attached to them, such as the membrane area, cell volume, or the number of molecules per cell. When those quantities are put into equations the basic units, i.e. time, number of molecules, and the length scale, need to match; thus, all quantities needs to be for instance in *minutes*, *mmol*, and *dm*. When interconverting units the following three basic mathematical relationships

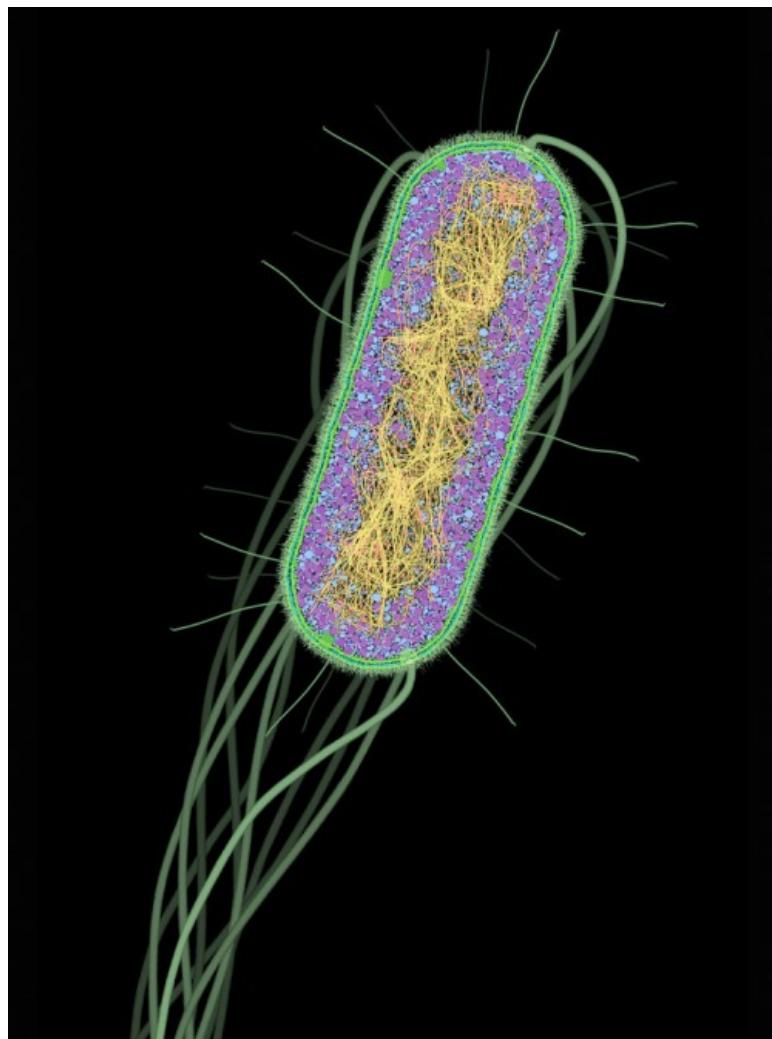
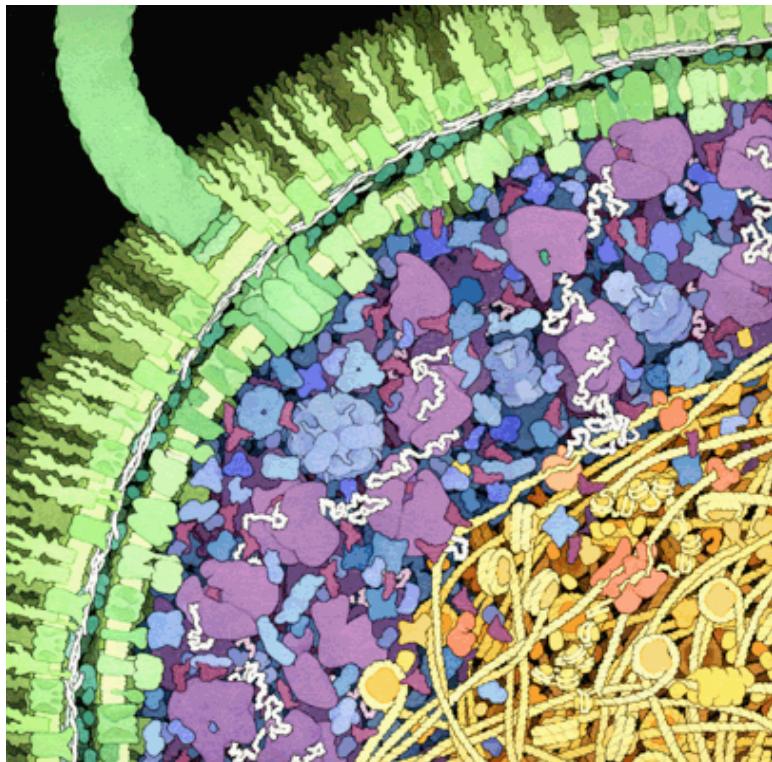


Figure 4.1: A drawing of *Escherichia coli* with realistic relative dimensions By Goodsell (see <http://mgl.scripps.edu/people/goodsell/>).



**Figure 4.2: 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  $1.4 \cdot 10^4 \text{ g/mol}$  moves at  $1.3 \cdot 10^3 \text{ cm/sec}$ ! 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!

as used,

1.  $(x \cdot y)^a = x^a \cdot y^a$ , e.g.  $m^3 = (10^1 \text{ dm})^3 = (10^1)^3 \text{ dm}^3$
2.  $(x^a)^b = x^{a \cdot b}$ , e.g.  $(10^1)^3 \text{ dm}^3 = 10^3 \text{ dm}^3$

$$3. x^a x^b = x^{a+b}, \text{ e.g. } 10^3 dm^3 = 10^3 \cdot (10^{-1} m)^3 = 10^3 \cdot 10^{-3} \cdot m^3 = 10^{3-3} m^3 = 10^0 m^3 = 1 m^3.$$

To make sure you are sufficiently familiar with interconverting units; here follow a few examples,

1. 1 liter = 1 dm<sup>3</sup> = 1 (10<sup>-1</sup> m)<sup>3</sup> = 1 (10<sup>-1</sup>)<sup>3</sup> m<sup>3</sup> = 1 10<sup>-3</sup> m<sup>3</sup>.
2. changing the units of a diffusion coefficient,  $D = 2 \mu\text{m}^2/\text{s} = 2 (10^{-6})^2 \text{ m}^2/\text{s} = 2 10^{-12} \text{ m}^2\text{s}^{-1} = 2 10^{-12} \text{ m}^2(1/60\text{min})^{-1} = 2 10^{-12} \text{ m}^2(1/60)^{-1} \text{ min}^{-1} = 120 10^{-12} \text{ m}^2/\text{min} = 0.12 10^{-9} \text{ m}^2/\text{min}$
3. working out the net unit of an equation for a diffusion limited second order rate constant for the association reaction  $A + B \rightarrow AB$ ;  $k = 4\pi(D_A + D_B)(r_a + r_b)$  with the diffusion coefficients  $D_A$  and  $D_B$  in  $\mu\text{m}^2/\text{s}$  and the molecular reaction radii,  $r_a$  and  $r_b$ , in  $\mu\text{m}/\text{molecule}$ . The unit of  $k$  becomes:  $\mu\text{m}^2/\text{s} \cdot \mu\text{m}/\text{molecule} = \mu\text{m}^3/(s \cdot \text{molecule}) = 1/(s \cdot \text{molecule}/\mu\text{m}^3)$ . This is correct because the unit of a second order rate constant should be  $1/(\text{time} \cdot \text{concentration})$ . Often we do not use for the concentration unit,  $\text{molecule}/\mu\text{m}^3$ , but, for instance,  $nM = 10^{-9} M = 10^{-9} \text{ mol/l}$ . Let's change the unit. To convert moles into molecule amount we will need to use the Avogadro's number,  $N_A = 6 \cdot 10^{23} \text{ molecules/mol}$ .<sup>1</sup> To convert the rate constant unit to  $nM^{-1}s^{-1}$  we start from the current unit and carry out the following unit conversion calculation,  $k = 1/(s \cdot \text{molecule}/\mu\text{m}^3) = \frac{1}{s \cdot \frac{\text{molecule}}{\mu\text{m}^3}} = \frac{1}{s \cdot \frac{1}{N_A} \cdot \frac{\text{mol}}{\text{molecule}} \cdot \frac{\text{molecule}}{\mu\text{m}^3}} = \frac{\frac{N_A}{\text{mol}}}{s \cdot \mu\text{m}^3} = \frac{\frac{N_A}{\text{mol}}}{s \cdot \frac{\text{mol}}{(10^{-5}\text{dm})^3}} = \frac{\frac{N_A}{\text{mol}}}{s \cdot \frac{\text{mol}}{10^{-15}(\text{dm})^3}} = \frac{\frac{N_A}{\text{mol}}}{s \cdot 10^{15} \cdot M} = \frac{\frac{N_A}{\text{mol}}}{s \cdot 10^{15} \cdot 10^9 nM} = \frac{\frac{N_A}{\text{mol}}}{s \cdot 10^{24} \cdot nM}$ . Thus if the rate constant is equal to  $k = 1 \text{ s}^{-1} (\text{molecule}/\mu\text{m}^3)^{-1} = 1 \frac{N_A}{s \cdot 10^{24} \cdot nM} = 1 \cdot 6 \cdot 10^{23} \frac{1}{10^{24}} \frac{1}{s \cdot nM} = 0.6 \cdot \frac{1}{s \cdot nM}$ .

## Exercise

1. Perform the following conversions:
  - (a) pM to nM (pl=picoliter=10<sup>-12</sup>).
  - (b) m<sup>3</sup> to μl.
  - (c) pM to molecules · l<sup>-1</sup>.
  - (d) dm<sup>2</sup>/min to μm<sup>2</sup>/s.
  - (e) 1 molecules/cell to nM if the cell has a volume of 1 fl (fl=femtoliter=10<sup>-15</sup> l; is approximately the volume of *E. coli*).
2. The volume of *E. coli* is approximately 1 μm<sup>3</sup>. Assume that *E. coli* is a sphere.
  - (a) Calculate the concentration of one molecule per cell in nM.

---

<sup>1</sup>This is an approximation, the real value is  $N_A = 6.02214129 \cdot 10^{23} \text{ molecules/mol}$ .

- (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?

### 4.3 Diffusion of molecules

Inside cells molecules move by diffusion. Diffusive motion (without drift) is unbiased in direction. A molecule moving by diffusion along one dimension, e.g. a protein diffusion along DNA, 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 4.3.

To get some insight into the statistics of diffusing molecules, we will consider 1D diffusion. This works exactly the same as flipping coins. If "head" is to move left and "tail" to move right then one can ask: "What is the probability to have 8 heads in 10 throws?". This would be the same as asking: "What is the probability that a molecule moved 8 diffusive steps to the left when it has taken 10 diffusive steps in total?". Many of properties of diffusion in 1D can easily transferred to 2D and 3D because diffusion in each dimensions occurs independently. So, 3D diffusion is like throwing with three coins; one for left/right, one for up/down, and one for towards you and away from you. And 2D with two coins. Clearly, those coin throws are independent events: the outcome of one coin throw does not influence the outcome of the other coin throw.

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 (the same distribution that describes coin flipping statistics),

$$p(N, n_L) = \binom{N}{n_L} p^{n_L} (1-p)^{N-n_L} \quad (4.1)$$

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,...,  $N$ . 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, unbiased diffusion those probabilities both 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 was expected. This does not describe the process completely



**Figure 4.3: Examples of four 2D random walks of 10000 diffusive steps.** The start position is indicated with a red dot. The net distance travelled is given by  $\sqrt{dN\delta^2}$  for a random walk of  $N$  diffusive steps of size  $\delta$  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. The equation tells us that he has travelled  $N = \frac{100^2}{d\cdot\delta^2} = \frac{100^2}{2\cdot1} = 5000$  steps; this is 5 km ( $= N \cdot \delta$ ).

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 measures the spread of the molecules and the width of

the binomial distribution. The variance increases with the number of steps  $N$ , indicating that over time, when  $N$  becomes larger, more spreading of molecules over the diffusive space occurs. Again, as expected.

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 diffuse 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  $N$ . However, when we are considering a great many molecules, like in the case of the blue colour dye, the number of particles at a given location approximates a continuous variable; then the binomial distribution can be approximated by a well-known probability distribution for continuous variables, the Gaussian or normal distribution (or bell shaped distribution).

When we consider movement to the right relative to the origin as a positive distance, the travelled distance  $\Delta$  equals  $n_R - n_L$ ; and then movement to the left becomes a negative distance. 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 be approximated by a continuous Gaussian distribution; which is sometimes called a normal distribution or a bell curve (see: figure 4.4). This is a standard limit of the binomial distribution and the derivation does not concern us here but can be found in elementary probability theory books.<sup>2</sup> The probability for a molecule to move distance  $\Delta$  after time  $t$  then obeys the following Gaussian distribution,<sup>3</sup>

$$p(\Delta, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{\Delta^2}{4Dt}} \quad (4.2)$$

Here we have defined time as  $t = N\tau$  as the number of diffusive steps times the waiting times for diffusive steps. The diffusion coefficient,  $D$ , is defined as  $D = \delta^2/(2\tau)$  with  $\tau$  and  $\delta$  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 in a moment. The spread of the probability at different times as prescribed by this equation is shown in figure 4.4. Because diffusion in all three dimensions is independent, these curves apply to the  $x$ ,  $y$ , and  $z$  dimension in case of 3D diffusion.

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

<sup>2</sup>Or the excellent physics book by Dill & Bromberg, Molecular Driving Forces: Statistical Thermodynamics in Chemistry, Physics, Biology, and Nanoscience, Garland Science, 2010. This book is a great introduction to the physics relevant for biology and biochemistry.

<sup>3</sup>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.  $x_1, \dots, x_i, \dots, x_n$  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  $x$ .

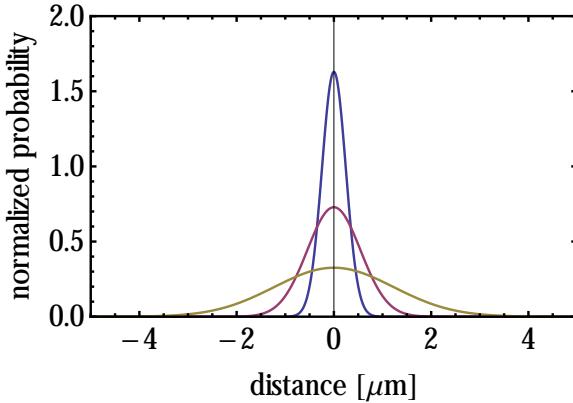


Figure 4.4: Examples of the probability density (equation 4.2) at three different times ( $3 \cdot 10^{-3}$ ,  $15 \cdot 10^{-3}$  and  $75 \cdot 10^{-3}$  seconds; the diffusion coefficient corresponds to a realistic value ( $10 \mu\text{m}^2/\text{s}$ ). The radius of an *E. coli* cell is about  $1\mu\text{m}$ .

calculate this we get,<sup>4</sup>

$$\langle \Delta \rangle = \int_{-\infty}^{\infty} \Delta \cdot p(\Delta, t) \cdot d\Delta = 0 \quad (4.3)$$

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. The spreading we can quantify again with the variance. The results that we will use most often relates the travelled distance to the diffusion coefficient through the variance of  $\Delta$ , denoted in physics by  $\langle \delta\Delta^2 \rangle$ ,

$$\langle \delta\Delta^2 \rangle = \int_{-\infty}^{\infty} \Delta^2 \cdot p(\Delta, t) \cdot d\Delta - \langle \Delta \rangle^2 = \langle \Delta^2 \rangle - \langle \Delta \rangle^2 = 2Dt \quad (4.4)$$

(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.  $\mu\text{m}^2$ . 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 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,

$$P(\Delta > \sqrt{2Dt}, t) = 1 - \int_{-\sqrt{2Dt}}^{\sqrt{2Dt}} P(\Delta, t) d\Delta = 0.32 \quad (\approx 1/3) \quad (4.5)$$

---

<sup>4</sup>How to perform these calculations you do not need to know at this stage of the course.

This last result shows the applicability of equation 4.4 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,

$$\langle \Delta^2 \rangle = \langle \Delta_x^2 \rangle + \langle \Delta_y^2 \rangle + \langle \Delta_z^2 \rangle = 6Dt \quad (4.6)$$

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, as we shall see below! Because the association of two proteins can of course not be faster than their diffusion speed.

#### 4.4 Example calculation of travelled distance by macromolecules

With the knowledge of the previous section, we can address problems related to the duration of diffusive phenomena. For instance, the diffusion coefficient of green fluorescent protein (GFP) is  $25 \mu\text{m}^2/\text{s}$  in mammalian cytoplasm (in water  $87 \mu\text{m}^2/\text{s}$  and in bacterial cytoplasm  $7.7 \mu\text{m}^2/\text{s}$ ). This means that GFP can travel in  $10 \text{ s}$  a distance of  $\sqrt{6 * 25 * 10} = 38 \mu\text{m}$ . The length of *E. coli* is about  $1 \mu\text{m}$  so this means a single molecule travels the length of *E. coli* in about  $1^2/(6 * 7.7) = 0.02 \text{ s}$  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.

#### Exercise

1. Equation 4.2 gives the probability distribution for the particles at some location along a line and the distribution broadens over time. Often, it is more convenient to picture the number of molecules than probability. Equation 4.2 can also be written as,

$$n(\Delta, t) = N \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{\Delta^2}{4Dt}} \quad (4.7)$$

With  $N$  as the total number of diffusing particles and  $n(\Delta, t)$  as the number of particles at distance  $\Delta$  and time  $t$ . This means that the probability  $p(\Delta, t)$  equals  $\frac{n(\Delta, t)}{N}$ .

#### 4.5. EXPRESSING THE DIFFUSION COEFFICIENT IN TERMS OF THE DIMENSIONS OF THE MOLECULE AND THE PROPERTIES OF THE MEDIUM

- (a) Explain why this probability definition indeed makes intuitive sense.
- (b) If  $D = 20 \mu\text{m}^2/\text{s}$  how much time does it take before  $2/3$  of the particles are more than  $5 \text{ cm}$  away from  $\Delta = 0$ ? Those would be realistic times for an ion diffusing through an axon by diffusion. Why doesn't your brain function like this?

### 4.5 Expressing the diffusion coefficient in terms of the dimensions of the molecule and the properties of the medium

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, and stop moving at  $T = 0 \text{ Kelvin}$ ). The physicochemical properties of the particle influence the diffusion coefficient according to the following (Einstein-Smoluchowski) relation,

$$D = \frac{kT}{f} \quad (4.8)$$

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

$$D = \frac{kT}{6\pi\eta a} \quad (4.9)$$

This equation indeed agrees with the intuition sketches above: the diffusion coefficient

- increases with temperature
- decreases with the size of the diffusing particle
- decreases with the viscosity of the medium

#### Exercise

1. Diffusion of GFP
  - (a) Plot the diffusion coefficient as function of the radius of a spherical particle. Take  $30^\circ\text{C}$  and express this temperature in units Kelvin.
  - (b) Double the temperature and make the same plot.
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  $\eta = 10^{-3} \text{ Pa s}$  and  $kT = 4 * 10^{-21} \text{ J}$ . Is this a realistic value?

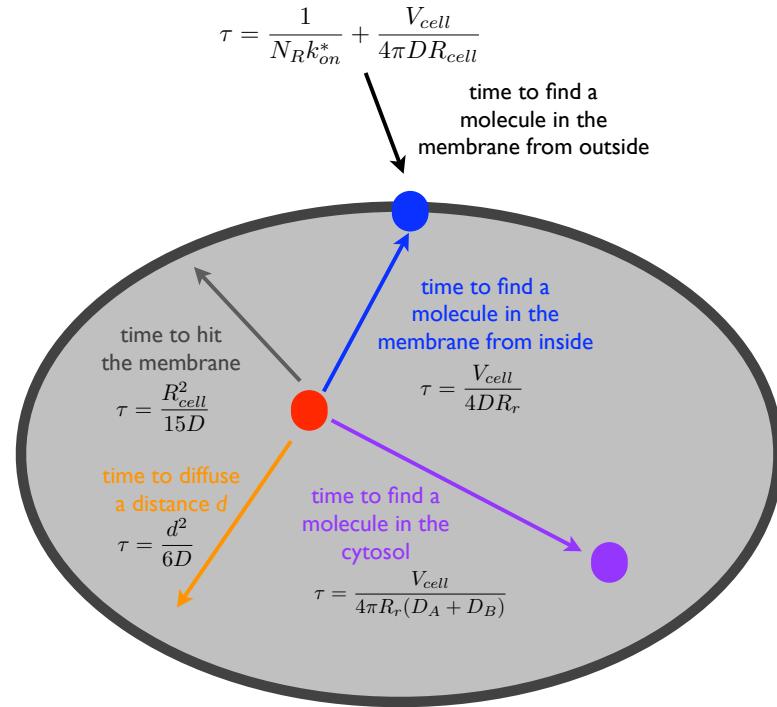


Figure 4.5: Diffusion times for a single molecule to find the membrane (Dobrzynski and Bruggeman, unpublished), a molecule in the membrane (from inside and outside) [37, 2], to travel a certain distance  $d$ , and to find a (diffusing) molecule in the cytosol [17]. The  $D$ ,  $D_A$ , and  $D_B$  denote a diffusion coefficient (in  $\mu\text{m}^2/\text{s}$ ),  $R_{cell}$  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  $\mu\text{m}$  for a prokaryote, a diffusion coefficient is typically 5  $\mu\text{m}^2/\text{s}$  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. The time to find a molecule in the membrane from the outside equals  $\frac{V_{cell}}{4\pi D R_{cell}}$  and the time to bind to a receptor is given by  $\frac{1}{N_R k_{on}^*}$  if  $N_R$  receptors are considered.

## 4.6 Diffusion-limited association reactions

In this section, we will equate association rates in terms of the diffusion coefficients of the two binding molecules. 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  $k_a$ . Hence, the reaction we consider is:  $A + B \rightarrow AB$ . The rate of the reaction is given by mass-action kinetics:  $v = k_a \cdot a \cdot b$ . So far, we have considered the reaction rate in units concentration per time. With  $a$  and  $b$  denoting the concentration of  $A$  and  $B$  defined as the number of molecules, e.g. in mol, per unit volume. Next we will derive how this rate constant can be expressed in terms of the diffusion time for the molecules of  $A$  and  $B$  to find each other.

The *mean* diffusion (collision) time,  $\tau_d$ , for two *single* molecules,  $A$  and  $B$ , is given by the Smoluchowski equation (Figure 4.5; we will not derive this equation here but this is quite straightforward),<sup>5</sup>

$$\tau_d = \frac{V}{4\pi(D_A + D_B)(r_A + r_B)} \quad (4.10)$$

Let's check the unit of this equation,

$$\tau_d = \frac{\mu m^3}{\frac{\mu m^2}{s} \cdot \mu m} = s \quad (4.11)$$

In fact we can make this a bit more explicit – this is normally not done, but very useful for our purposes,

$$\tau_d = \frac{\frac{\mu m^3}{cell}}{\frac{\mu m^2}{s} \cdot \frac{\mu m}{molecule}} = s \cdot \frac{molecule}{cell} \quad (4.12)$$

The diffusion-limited association rate constant,  $k_a$ , in unit  $\frac{1}{s \cdot \frac{molecule}{cell}}$  is then given by  $1/\tau_d$ . This is an odd unit and we would prefer a more familiar and useful unit  $\frac{1}{s \cdot \frac{molecule}{volume}}$  or  $\frac{1}{s \cdot \text{concentration}}$ ; let's do that:

$$\begin{aligned} \frac{1}{s \cdot \frac{molecule}{volume}} &= \frac{1}{s \cdot \frac{molecule}{cell}} \frac{1}{\frac{cell}{volume}} = \frac{1}{s \cdot \frac{molecule}{cell}} \frac{volume}{cell} \Rightarrow \\ \frac{1}{\tau_d} V &= 4\pi(D_A + D_B)(r_A + r_B) = k_a \text{ in } \frac{1}{s \cdot \frac{molecule}{volume}} \end{aligned} \quad (4.13)$$

---

<sup>5</sup>Of course, since this is an average time, this means that individual molecules will display a great spread in their search times. This makes the search time a random variable – we cannot say beforehand what the search time of a specific molecule will be. The search time will be often distributed according to an exponential distribution; because the search time for one molecule for another molecule in a large volume is independent of the initial position of the molecule; hence, it is memoryless.

To arrive at the unit  $\frac{1}{s \cdot \text{concentration}}$  we have to use Avogadro's constant,

$$\frac{1}{s \cdot \text{concentration}} = \frac{1}{s \cdot \frac{\text{molecule}}{\text{volume}} \cdot \frac{\text{mol}}{\text{molecule}}} = \frac{1}{s \cdot \frac{\text{molecule}}{\text{volume}}} \cdot \frac{\text{molecule}}{\text{mol}}$$

$$\frac{1}{\tau_d} V N_A = 4\pi(D_A + D_B)(r_A + r_B)N_A = k_a \text{ in } \frac{1}{s \cdot M} \text{ if } D \text{ and } r \text{ in } dm \quad (4.14)$$

Let's perform a double check,

$$\begin{aligned} k_a &= 4\pi(D_A + D_B)(r_A + r_B)N_A = dm^2/s \cdot dm/\text{molecule} \cdot \text{molecule/mol} \\ &= \frac{dm^3}{s \cdot mol} = \frac{1}{s \cdot M} \end{aligned} \quad (4.15)$$

So, it is correct. This is the main equation for a second-order diffusion limited rate constant for use in rate equation.

## 4.7 When an association reaction becomes diffusion limited

A simple view of an association reaction between molecule *A* and *B* decomposes this process into two sequential events: i. the two molecules *A* and *B* have to collide with each other and ii. after they have collided, *A* and *B* have to bind (react). The sum of those two times,  $\tau_d$  and  $\tau_r$ , gives the association time,  $\tau_a$ . When the diffusion time is much larger than the reaction time then the association reaction is diffusion limited. Then, the rate constant for the association reaction,  $k_a$ , equals the diffusion limited second order rate constant as derived in the previous section.

### 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 rate 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  $10 * 10^{-10} m$ .

## 4.8 Mean single molecule search times in the cell

Equation 4.10 expresses the mean time for two single molecules *A* and *B*, to find each other in a cellular compartment given their radius and diffusion coefficients. From this equation we can derive the diffusion-limited second order rate constant

as was done above (equation 4.15). In this section, we will discuss other single molecule search problems shown in Figure 4.5.

Many more search times are relevant in cell biology, besides the time for two molecules in the cytosol to encounter each other. All those times can be used to estimate rate constants that we can use in kinetic models. The relationships for search times are shown in figure 4.5. Their units are:

$$\tau = \frac{R_{cell}^2}{15D} = \frac{\frac{\mu m^2}{cell}}{\frac{\mu m^2}{s}} = \frac{s}{cell} \quad (4.16)$$

$$\tau = \frac{d^2}{6D} = \frac{\mu m^2}{\frac{\mu m^2}{s}} = s \quad (4.17)$$

$$\tau = \frac{V_{cell}}{4DR_r} = \frac{\frac{\mu m^3}{cell}}{\frac{\mu m^2}{s} \frac{\mu m}{molecule}} = s \frac{molecule}{cell} \quad (4.18)$$

$$\tau = \frac{V_{cell}}{4\pi DR_{cell}} = \frac{\frac{\mu m^3}{cell}}{\frac{\mu m^2}{s} \frac{\mu m}{molecule}} = s \frac{molecule}{cell} \quad (4.19)$$

$$\tau = \frac{V_{cell}}{4\pi R_r(D_A + D_B)} = \frac{\frac{\mu m^3}{cell}}{\frac{\mu m}{molecule} \frac{\mu m^2}{s}} = s \frac{molecule}{cell} \quad (4.20)$$

The last relationship ( $R_r = r_A + r_B$ ) was analysed above and has the same unit as the equation 4.18 and 4.19; they are all single-molecule search times for another molecule. They can be easily converted to equations for diffusion-limited rate constants with proper units for use in rate equations:

$$k_a = \frac{1}{\tau} V_{cell} N_A = \frac{cell}{molecule \cdot s} \frac{l}{cell} \frac{molecule}{mol} = \frac{l}{s \cdot mol} = \frac{1}{s \cdot M} \quad (4.21)$$

## 4.9 Search times by single molecules and association kinetics

In the previous sections, the search times of individual molecules for other molecules has been related to the diffusion-limited second order rate constant for the association reaction (equation 4.21). The association rate equation for the association of molecule  $A$  and  $B$  is given by,

$$\frac{dab}{dt} = v = k_a \cdot a \cdot b = \frac{V_{cell} N_A}{\tau} \cdot a \cdot b, \quad (4.22)$$

and gives the number of association events per unit time and per unit volume,  $v$ , e.g. in  $M/s$ , and  $\tau$  is either given by equation 4.19, 4.18 or 4.20. Here,  $\tau$  denotes the search time for 1 molecule and  $\tau/N_A$  denotes the search time for 1 mol of molecules. But how can 1 mol of molecules search quicker than 1 molecule? In fact, the association time corresponds to the first collision of  $a$  and  $b$  and the situation we are considering is that we either have 1 mol of  $A$  (or  $B$ ) and 1

molecule of  $B$  (or  $A$ ). Then having more molecules engage in the search implies that the single molecule is found quicker.<sup>6</sup>

## 4.10 Comparison of molecule search times

Let's study the equations in Figure 4.5 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,

$$\frac{\text{time to find a molecule in the membrane}}{\text{time to find the membrane}} = 5\pi \frac{R_{cell}}{R_r} \quad (4.23)$$

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  $8 \cdot 10^5$ . This means that if a molecule has found another molecule in the membrane that it has crossed the cell 8000 times! It is like looking for a needle in a haystack.

The time to find a molecule in the membrane divided by the time to find a molecule in the cytosol gives,

$$\frac{\text{time to find a molecule in the membrane}}{\text{time to find a molecule in the cytosol}} = 8\pi \quad (4.24)$$

The time to find a molecule in the membrane is always longer than the time for finding a molecule in the cytosol.<sup>7</sup>

### Exercises

1. Study Figure 4.5 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 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  $80 \mu m$ . 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?

---

<sup>6</sup>In fact, the rates we are considering are so-called mean first-passage times. The first-passage time probability distribution is probability density for the search times of a single molecule and the first-passage time distribution out of  $N$  searchers can also be derived and this has a mean first-passage time which  $N$  times smaller than the mean search time of a single molecule.

<sup>7</sup>I do not understand intuitively why.

3. Can an average reaction time be shorter than an average diffusion time?
4. A mammalian cell has typically a dimension of  $10000 \mu m^3$ . 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 5

# Using diffusion times to get more insight into some basic cellular processes

### 5.1 Molecule search times and the response time of prokaryotic two-component signalling

Two-component signal transduction is the prevalent signalling mechanism for prokaryotes – we studied its robustness properties already in the previous chapter. For most two-component signalling networks, the sensor protein typically sits in the membrane and the response regulator – the transcription factor – resides in the cytoplasm. Several diffusive searches are important for the signalling process. When the sensor is active, after having phosphorylated itself, the response regulator can become active when it collides with a phosphorylated sensor. The time for a single response regulator protein to find a single phosphorylated sensor in the membrane is given by,  $\tau_1 = \frac{V_{cell}}{4 \cdot D \cdot R_r}$ . With *E. coli* properties we find for this time,  $\tau_1 = \frac{\frac{4}{3} \pi (1 \mu m)^3}{4 \cdot 7.7 \frac{\mu m^2}{s} \cdot 10 \cdot 10^{-3} \mu m} = 13.6 \text{ s}$ . With  $n_r$  response regulators and  $n_m$  membrane proteins, this time reduces to  $\tau / (n_r \cdot n_m)$ . So, with 10 molecules each the search time is about one tenth of a second. This is fast: the normal response to two-component signalling is gene expression, it takes about a minute to make a protein after the gene has been activated in *E. coli*. The next diffusion process concerns the search of the activated response regulator for the promoter site. How much time will this take? We now require the following equation  $\tau_2 = \frac{V_{cell}}{4 \pi R_r D}$  (the DNA site does not diffuse), which is a factor  $1/\pi$  smaller than the previous time if you compare the equations, so  $13.6/\pi = 4.3 \text{ s}$ . But, the response regulator could also have found the unphosphorylated sensor that can dephosphorylate the response regulator – the sensor became unphosphorylated after having transferred its phosphate to the response

regulator. The probability,  $p$ , that the response regulator found the DNA site equals  $p = \frac{k_2}{k_1+k_2} = \frac{1/\tau_2 N_A V_{cell}}{1/\tau_1 N_A V_{cell} + 1/\tau_2 N_A V_{cell}} = \frac{1/\tau_2}{1/\tau_1 + 1/\tau_2} = \frac{\tau_1}{\tau_1 + \tau_2} = \frac{\pi}{\pi + 1} = 0.75$  (the  $\pi$  appeared because  $\tau_1 = \pi \cdot \tau_2$ ; you will see this when you look carefully at those  $\tau$  equations). So, only  $\sim 3/4$  times does the phosphorylated transcription factor actually encounter the DNA site to activate transcription, in all the other cases it is inactivated before gene expression. This means that the gene network response only occurs on average after  $4/3$  response regulator phosphorylations, which means that the extracellular activator for the sensor should at least be present for a few tens of seconds ( $\approx 4/3 \cdot (\tau_1 + \tau_2)$ ). In this way, the signaling network does not respond immediately to every extracellular fluctuation in the signal. Note, that we here assumed a single sensor and response regulator molecules per cell, which is a simplification. In reality, single cells may have tens to hundreds of those proteins per cell.

## 5.2 Leaky transcription rates due to spontaneous fluctuations in the promoter occupation by a repressor

Suppose that we have 10 repressor proteins in *E. coli*. Their concentration is,

$$r = 10 \frac{\text{molecules}}{\text{cell}} \cdot \frac{1}{N_A} \frac{\text{mol}}{\text{molecules}} \cdot \frac{1}{\frac{4}{3}\pi(10^{-5} \text{ dm})^3} \frac{\text{cell}}{\text{volume in liter}} = 4 \text{ nM}.$$

The association rate constant for repressor and DNA site binding is in the diffusion-limited regime equal to  $k_a = 4\pi DR_r N_A \cdot 10^{-9}$  in  $\frac{1}{\text{s} \cdot \text{nM}}$ , which equals

$$4\pi \cdot 7.7 \cdot 10^{-10} \text{ dm}^2/\text{s} \cdot 10 \cdot 10^{-8} \text{ dm} \cdot 6 \cdot 10^{23} \cdot 10^{-9} = 0.58 \text{ s}^{-1} \text{ nM}^{-1}.$$

We know that the repressor-promoter complex is equal to  $rp = p_T \frac{r}{K_D + r}$  in concentration. If all the repressor molecules are active and we would like to have  $\sim 90\%$  inhibition ( $rp/p_T = 0.9$ ) then the dissociation constant should equal,  $0.11 \cdot r = 0.44 \text{ nM}$ . The dissociation constant equals  $K_D = k_d/k_a$ . So, the dissociation rate constant equals  $k_d = k_a \cdot K_D = 0.58 \cdot 0.44 = 0.25 \text{ s}^{-1}$ . This means that the repressor-promoter complex has a characteristic life time of  $1/k_d = 4 \text{ s}$ . Thus, the repressor-promoter complex dissociates on average after 4 s. The number of associations between the repressor and the promoter per unit time equals  $k_a \cdot r = 0.58 \cdot 4 = 2.3 \text{ s}^{-1}$ . Thus, after the promoter-repressor complex has dissociated the promoter remains empty for  $1/(k_a \cdot r) = 0.4 \text{ s}$ . During this time window either a RNA polymerase can bind to the promoter – to induce leaky, faulty transcription – or a repressor can bind. What is the probability that the RNA polymerase binds? This is equal to  $p = \frac{k_a/N \cdot pol}{k_a/N \cdot pol + k_a \cdot r} = \frac{pol/N}{pol/N + r}$ . Here we assume a concentration  $pol$  for RNA polymerase, diffusion-limited association, and on average  $N$  free promoters in total in the *E. coli* genome. The free concentration of RNA polymerases per cell is about 250 on average. For a 5% probability that a polymerase binds the number of competing promoters,  $N$ ,

should be  $N = 19 \cdot \frac{pol}{r} = 19 \cdot 260/10 = 494$ , which is about 10% of the genome size of *E. coli*. In addition, if the RNA polymerase association reaction operates in a more reaction limited regime then this number would reduce further. This means that if 5% leakiness is achievable for *E. coli* then in 1/20 cases of unoccupied promoters the RNA polymerase will bind – even though the gene is supposed to be off. This means that every  $20 \cdot (4 + 0.3) = 86$  s a faulty mRNA is produced. Depending on its life time, this can lead to quite some production of protein. If the mRNA would live for 10 min then at the steady state leaky transcript level is 7 mRNAs/cell. This number is still too high, suggesting that the RNA polymerase binding should operate in a more reaction limited regime than the transcription factor because then  $p = \frac{k_p/N \cdot pol}{k_p/N \cdot pol + k_r \cdot r} = \frac{\alpha/N \cdot pol}{\alpha/N \cdot pol + r}$  with  $\alpha = k_p/k_r < 1$ . This can reduce  $p$  far below 5% and the number of mRNAs/cell below 1.

## 5.3 A diffusion limited kinase

### 5.3.1 Diffusion-limited association

When a kinase,  $K$ , phosphorylates a protein target,  $E$ , it first needs to bind to it to form an enzyme-substrate complex,  $EK$ , and then it can phosphorylate the target to form  $EP$ ,



Here, we simplified the mechanism and subsumed the ATP and ADP concentrations into the elementary rate constants. At steady state, the rate of phosphorylation equals,

$$\frac{dEP}{dt} = v_2 = k_2 ek. \quad (5.2)$$

Then,

$$k_1^+ \cdot e \cdot k - k_1^- ek = k_2 ek, \quad (5.3)$$

which gives for,

$$ek = \frac{k_1^+}{k_1^- + k_2} \cdot e \cdot k \quad (5.4)$$

In addition, the total kinase concentration is conserved, i.e.  $k_T = k + ek$ , which allows for the elimination of  $k$  from the previous equation and the following expression for  $ek$  is obtained,

$$ek = k_T \frac{e}{e + K_M} \quad (5.5)$$

Here  $K_M$  is defined as  $\frac{k_1^- + k_2}{k_1^+}$ . If  $k_1^+$  achieves its highest possible value, i.e. when it achieves the bound set by diffusion limit, then the minimal value for the  $K_M$  is attained. The rate of the phosphorylation equals,

$$v_2 = k_2 k_T \frac{e}{e + K_M}. \quad (5.6)$$

### 5.3.2 Diffusion-limited association and this is also the slowest process in the enzyme

Suppose that this system is limited by binding of  $E$  and  $K$ , i.e.  $k_1^+$  is the slowest reaction constant, this means that  $ek \approx 0$ ; and therefore

$$\frac{dep}{dt} = v_2 = v_1 \approx k_1^+ \cdot e \cdot k \approx k_1^+ \cdot e \cdot k_T. \quad (5.7)$$

If  $k_1^+$  is then given by the diffusion-limit of the second-order rate constant (equation 4.15), we found the maximal enzyme rate that biology can attain,

$$v = k_1^+ \cdot e \cdot k_T = 4\pi(D_E + D_K)(r_E + r_K)N_A \cdot k_T \cdot e. \quad (5.8)$$

### 5.3.3 Diffusion-limited association and the catalytic rate is limiting enzyme rate

Another limit of the enzyme functioning is when  $k_2$  is the slowest rate constant then reaction  $v_1$  operates so fast that it operates close to thermodynamic equilibrium,

$$\begin{aligned} k_1^+ \cdot e \cdot k &\approx k_1^- ek \quad \Rightarrow \\ k_1^+ \cdot e \cdot (k_T - ek) &\approx k_1^- ek \quad \Rightarrow \\ ek &\approx k_T \frac{e}{e + K_D} \end{aligned} \quad (5.9)$$

Here  $K_D = \frac{k_1^-}{k_1^+}$  is defined as a dissociation constant. If  $k_1^+$  attains its highest possible value, which is the diffusion limit, the minimal value for the  $K_D$  is achieved.

## 5.4 Facilitated diffusion through biological membranes

### 5.4.1 Fick's law and membrane diffusion

The diffusion flux over a biological membrane leads to the following rate of change in the concentration of intracellular molecules<sup>1</sup>,

$$\frac{d}{dt}x_i = K \cdot D \cdot \frac{A}{V} \cdot \frac{x_o - x_i}{d} = P \cdot (x_o - x_i) \quad (5.10)$$

---

<sup>1</sup>This equation can be derived from Fick's law,

$$J = D \frac{\partial c}{\partial x},$$

with  $c$  as concentration and  $x$  as a distance coordinate, via an analysis of the dimensions of the problem.

The parameter  $K$  is the partition coefficient<sup>2</sup>,  $d$  is the membrane thickness,  $D$  the diffusion coefficient of the compound in the membrane,  $A$  is the membrane area,  $V$  is the cell volume, and  $x_o$  and  $x_i$  are extra- and intracellular concentration of the diffusing molecule. The parameter  $P$  is the so-called permeability coefficient and hides all the information, which is typically unknown for the transport process. It is a very useful constant that can be obtained from experiment. Note that this mechanism cannot lead to the accumulation of a molecule inside the cell, i.e. to achieve  $x_i > x_o$ . In addition, since we are considering the same molecule intracellularly and extracellularly – we are not considering a conversion of a molecule, only transport – the equilibrium constant for this process is 1. Therefore, at thermodynamic equilibrium  $x_i = x_o$ . Small non-charged molecules, such as weak acids, can typically diffuse spontaneously over biological membranes, such as pyruvate, and acetate.

If we are considering a molecule that is consumed by the cell then one interesting limit is the one where the intracellular concentration equals 0,

$$\frac{d}{dt}x_i = K \cdot D \cdot \frac{A}{V} \cdot \frac{x_o}{d} = P \cdot x_o = v \quad (5.11)$$

If the demand of the cell for this molecule is higher than this rate then a transporter is required to enhance the rate of uptake. This could for instance occur because: i. the molecule is very hydrophilic and does not spontaneously dissolve in the membrane or ii. the diffusion of the molecule through the membrane is very slow. In both cases, a transporter could yield an uptake rate for the cell which is higher than  $v$ . Two kinds of transporters can be distinguished; transporters that act as facilitated diffusion mechanisms or transporters that invest energy in the transport process, so-called active transport.

#### 5.4.2 Facilitated passive transport via a dedicated protein

”Facilitated” means that a protein is used to transport a molecule over a membrane. ”Passive” means that it does not require cellular energy. The protein can either function as a channel or it can flip in the membrane such that its extracellular domain flips intracellularly when it binds a molecule extracellularly. The reverse process also occurs. In all cases of facilitated transport, the net reaction we are considering is,



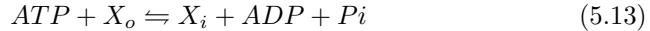
Since, we have the same molecule,  $X$ , on both side of the equation the equilibrium constant of this process is 1. As a consequence, this process has a strong tendency to also export the molecules out of the cell. In the next chapter on enzyme kinetics, we will look a bit more into membrane transporter kinetics.

---

<sup>2</sup>This is the equilibrium constant for solvation of the molecule into the membrane  $K = \frac{x_o^m}{x_o}$ , with  $x_o^m$  as the membrane concentration.

### 5.4.3 Active transport

The consequence of active transport is that the net transport process changes, for instance if ATP is hydrolysed, we obtain,



This equation suggests that if the cell manages to keep the ratio of  $\frac{ATP}{ADP \cdot Pi}$  high then  $x_i$  can become higher than  $x_o$  in thermodynamic equilibrium than in the case of facilitated transport.<sup>3</sup> However, the cell needs to continuously invest energy to maintain the ratio  $\frac{ATP}{ADP \cdot Pi}$  high, as in the transport process for every transported molecule of  $X$ , one ATP molecule is hydrolysed to yield ADP and Pi, such that the ratio  $\frac{ATP}{ADP \cdot Pi}$  reduces. In order to determine the  $\frac{ATP}{ADP \cdot Pi}$  ratio to keep the  $x_i/x_o$  ratio at some value would require a thermodynamic analysis, which we will not carry out here.

### 5.4.4 Facilitated passive transport of a weak acid

Because charged molecules do not move over biological membranes, a pH difference over a biological membrane – maintained by some cellular process – can lead to the accumulation of the protonated form of a weak acid intracellularly. At both sides of the membrane we have the following acid-base equilibrium,



We will consider the situation that  $h_o^+ \neq h_i^+$  and that the proton concentration remains fixed. Remember that  $pH = -\log h^+$  and that therefore  $h^+ = 10^{-pH}$ . Thus we have the following rate equation for the import process,

$$v = P \cdot (xh_o - xhi) = P \cdot \left( \frac{x_o^- \cdot h_o^+}{K_a} - \frac{x_i^- \cdot h_i^+}{K_a} \right), \quad (5.15)$$

with  $P$  as the permeability coefficient. Thus at thermodynamic equilibrium we obtain,

$$\frac{x_o^- \cdot h_o^+}{K_a} = \frac{x_i^- \cdot h_i^+}{K_a} \Rightarrow \frac{x_i^-}{x_o^-} = \frac{h_o^+}{h_i^+} = 10^{pH_i - pH_o} \quad (5.16)$$

Thus, if the environment is more acid than the cellular milieu, which it typically is, then  $x_i^- > x_o^-$ . The ratio of the total concentrations equals,

$$\frac{x_i + xh_i}{x_o + xh_o} = \frac{x_i^- \left( 1 + \frac{h_i^+}{K_a} \right)}{x_o^- \left( 1 + \frac{h_o^+}{K_a} \right)} = 10^{pH_i - pH_o} \frac{1 + \frac{10^{-pH_i}}{K_a}}{1 + \frac{10^{-pH_o}}{K_a}} \quad (5.17)$$

---

<sup>3</sup>This can easily be concluded from the simplest kinetics for this process:  $v = k^+ \cdot atp \cdot x_o - k^- \cdot adp \cdot pi$ .

Let's for instance consider acetate with a  $K_a = 1.8 \cdot 10^{-5}$ ,  $pH_o = 5$ , and  $pH_i = 7$ ; then,

$$\frac{acetate_{in}}{acetate_{out}} = 10^2 \frac{1 + \frac{10^{-7}}{1.5 \cdot 10^{-5}}}{1 + \frac{10^{-5}}{1.5 \cdot 10^{-5}}} = 60.4 \quad (5.18)$$

Note that,

$$K_a = \frac{x_i^- \cdot h_i^+}{xh_i} = \frac{x_o^- \cdot h_o^+}{xh_o} \quad (5.19)$$

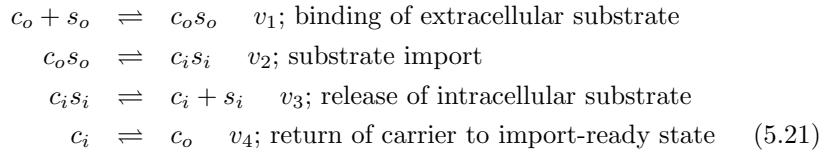
Hence,

$$\frac{x_i^- \cdot h_i^+}{x_o^- \cdot h_o^+} = \frac{xh_i}{xh_o} = 1, \quad (5.20)$$

which also fits with the relation  $\frac{x_i^-}{x_o^-} = \frac{h_o^+}{h_i^+}$  we derived above; moreover,  $\frac{xh_i}{xh_o} = 1$  because we assumed transport equilibrium. If the environment is more acidic then  $h_o^+ > h_i^+$  and, since  $xh_i = xh_o$ , we require  $x_o^- < x_i^-$  (because we have an acid-base equilibrium) and, hence, more weak acid occurs intracellularly.

#### 5.4.5 Facilitated diffusion via a carrier protein

In Figure 5.1, the state diagram is shown for a carrier protein that imports and exports the molecule  $S$ . The extra- and intracellular concentrations of the molecule are denoted as  $s_o$  and  $s_i$ , respectively. The carrier binding site for  $S$  on the outside is denoted by  $C_o$  and the inside site by  $C_i$ ; the concentration of carrier bound to extracellular substrate is then equal to  $c_o s_o$  and to intracellular substrate  $c_i s_i$ . Movement of the molecule from outside to inside is  $c_o s_o \rightarrow c_i s_i$  and to the outside is the reverse process. We assume that the molecule moves through the carrier by a diffusion process; e.g. the carrier can be viewed as a channel. The following four reactions occur:



We will assume that the binding of extra- and intracellular substrate occurs at thermodynamic equilibrium with the same dissociation constant,

$$K_s = \frac{c_o \cdot s_o}{c_o s_o} = \frac{c_i \cdot s_i}{c_i s_i} = \frac{k_1^-}{k_1^+} \quad (5.22)$$

This means that any given moment in time the total concentration of  $c_{o,T}$  and  $c_{i,T}$  equals,

$$\begin{aligned} c_{o,T} &= c_o + c_o s_o = c_o \left(1 + \frac{s_o}{K_s}\right) \\ c_{i,T} &= c_i + c_i s_i = c_i \left(1 + \frac{s_i}{K_s}\right) \end{aligned} \quad (5.23)$$

Hence,  $c_o = \frac{c_{o,T}}{1 + \frac{s_o}{K_s}} = c_{o,T} \frac{K_s}{K_s + s_o}$  and  $c_o s_o = \frac{c_{o,T}}{1 + \frac{s_o}{K_s}} \frac{s_o}{K_s} = c_{o,T} \frac{s_o}{K_s + s_o}$  and the similar equations hold for  $c_i$  and  $c_i s_i$ .

We have the following differential equations for the concentrations of the enzyme states,

$$\begin{aligned}\frac{dc_o}{dt} &= -v_1 + v_4 \\ \frac{dc_o s_o}{dt} &= v_1 - v_2 \\ \frac{dc_i s_i}{dt} &= v_2 - v_3 \\ \frac{dc_i s_i}{dt} &= v_3 - v_4\end{aligned}\tag{5.24}$$

Since we assume equilibrium conditions for reaction 1 and 3, their rates will always be zero and, hence, we can better look at the following differential equations,

$$\begin{aligned}\frac{dc_o + c_o s_o}{dt} &= \frac{dc_{o,T}}{dt} = v_4 - v_2 \\ \frac{dc_i + c_i s_i}{dt} &= \frac{dc_{i,T}}{dt} = v_2 - v_4\end{aligned}\tag{5.25}$$

At steady state of the enzyme we then have that,

$$\begin{aligned}v_2 = v_4 \Rightarrow k_{cs} (c_o s_o - c_i s_i) &= k_c (c_i - c_o) \Rightarrow \\ k_{cs} \left( c_{o,T} \frac{s_o}{K_s + s_o} - c_{i,T} \frac{s_i}{K_s + s_i} \right) &= k_c \left( c_{o,T} \frac{K_s}{K_s + s_o} - c_{i,T} \frac{K_s}{K_s + s_i} \right)\end{aligned}\tag{5.26}$$

And,

$$c = c_{o,T} + c_{i,T}\tag{5.27}$$

The latter two equation are sufficient to express the steady-state concentrations if  $c_{o,T}$  and  $c_{i,T}$  in terms of kinetic parameters,

$$c_{o,T} = \frac{c(k_c K_s + k_{cs} s_i)(K_s + s_o)}{2k_c K_s^2 + k_c K_s s_i + k_{cs} K_s s_i + k_c K_s s_o + k_{cs} K_s s_o + 2k_{cs} s_i s_o}\tag{5.28}$$

And  $c_{i,T} = c - c_{o,T}$ . The rate of the enzyme is,

$$\begin{aligned}
v &= v_2 = k_{cs} (c_o s_o - c_i s_i) \\
&= \frac{ck_c k_{cs} K_s (s_o - s_i)}{2k_c K_s^2 + (k_c K_s + k_{cs} K_s) s_i + (k_c K_s + k_{cs} K_s) s_o + 2k_{cs} s_i s_o} \\
&= \frac{\frac{ck_c k_{cs} K_s}{2k_c K_s^2} (s_o - s_i)}{1 + \frac{(k_c K_s + k_{cs} K_s)}{2k_c K_s^2} s_i + \frac{(k_c K_s + k_{cs} K_s)}{2k_c K_s^2} s_o + \frac{2k_{cs}}{2k_c K_s^2} s_i s_o} \\
&= \frac{\frac{ck_{cs}}{2K_s} (s_o - s_i)}{1 + \underbrace{\frac{(k_c + k_{cs})}{2k_c K_s} s_i}_{1/K_M} + \frac{(k_c + k_{cs})}{2k_c K_s} s_o + \frac{k_{cs}}{k_c K_s^2} s_i s_o} \\
&= \frac{\frac{ck_{cs}}{2K_s} (s_o - s_i)}{1 + \frac{s_i}{K_M} + \frac{s_o}{K_M} + \frac{k_{cs}}{k_c K_s^2} s_i s_o} \\
&= \frac{\overbrace{\frac{ck_{cs}}{k_c (k_c + k_{cs})} \left( \frac{s_o}{K_M} - \frac{s_i}{K_M} \right)}^{V_{MAX}}}{1 + \frac{s_i}{K_M} + \frac{s_o}{K_M} + \underbrace{\frac{k_{cs}}{k_c} \left( \frac{2k_c}{k_c + k_{cs}} \right)^2 \frac{s_i s_o}{K_M^2}}_{K_i}} \\
&= \frac{\frac{V_{MAX}}{K_M} (s_o - s_i)}{1 + \frac{s_i}{K_M} + \frac{s_o}{K_M} + K_i \frac{s_i s_o}{K_M^2}}
\end{aligned} \tag{5.29}$$

This rate equation for a carrier is used frequently for transporters in kinetic models of metabolic pathways.

Note that the equilibrium constant for the carrier, obtained by setting equation 5.29 to 0, equals 1,

$$\frac{V_{MAX}}{K_M} s_{o,eq} = s_{i,eq} \frac{V_{MAX}}{K_M} \Rightarrow K_{eq} = \frac{s_{i,eq}}{s_{o,eq}} = 1 \tag{5.30}$$

This should not be a surprise, since we are looking at the facilitated diffusion process.

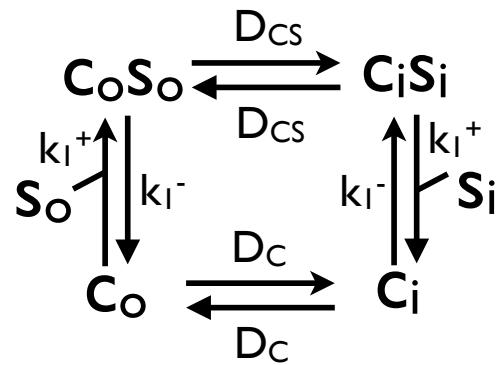


Figure 5.1: Mechanism for a membrane transporter acting as a facilitated diffusion carrier. The transport solute occurs at the external cell side with concentration,  $s_o$ , where it binds the carrier, occurring at concentration,  $c_o$ . As the carrier diffuses continuously through the membrane it sometimes occurs at the intracellular side, at concentration  $c_i s_i$ , where it can deposit the solute.

# Chapter 6

## Basic enzyme kinetics

### 6.1 Enzymes

Most reactions inside cells are catalysed 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 reaction rates; gene expression can alter the levels of enzymes and effectors can alter enzyme properties through allosteric regulation. Enzymes are true catalysts, they can only speed up reactions and are not consumed during the catalysis process. Also, they cannot alter the equilibrium constant of a reaction, which is determined by the thermodynamic properties of the reactants. In their catalytic site, enzymes offer a favourable physicochemical environment for the reaction chemistry to occur. In this way, enzymes speed up reactions that would otherwise occur in the cytoplasm at much lower rates. An enzyme may besides its catalytic site have regulatory, allosteric sites that affect the kinetic properties through the induction of conformation changes of the protein. Enzyme kinetics can be derived from a mass-action description of the elementary reactions involved in the enzyme mechanism. In this chapter, some of the basic concepts of enzymology will be explained.

### 6.2 Irreversible Michaelis-Menten kinetics

The study of enzyme kinetics, enzymology, is a large field with many details. Many of these details and tricks in the field can be found in this chapter, although we cannot cover the entire field. The book by Segel on enzyme kinetics does achieve this. 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.

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



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

$$\begin{aligned} v_1 &= k_1^+ e \cdot s - k_1^- es \\ v_2 &= k_2 es \end{aligned} \quad (6.2)$$

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

$$v = V_{MAX} \frac{s}{K_M + s} \quad (6.3)$$

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 constant of the enzyme for  $S$  denoted by  $K_M$ . A measure for the affinity of the enzyme for the substrate is  $1/K_M$ . This equation prescribes a hyperbolic relationship between  $v$  and  $s$  and is shown in Figure 6.1. The following relations become apparent when studying this figure:  $v = 1/2V_{MAX}$  at  $s = K_M$  and  $v \rightarrow V_{MAX}$  when  $s \gg K_M$ . 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. These relations can be obtained 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.

### 6.2.1 Derivation of enzyme kinetics: quasi-steady state assumption

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



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

$$\begin{aligned} \frac{de}{dt} &= -(k_1^+ e \cdot s - k_1^- es) + k_2 es \\ \frac{des}{dt} &= (k_1^+ e \cdot s - k_1^- es) - k_2 es \end{aligned} \quad (6.5)$$

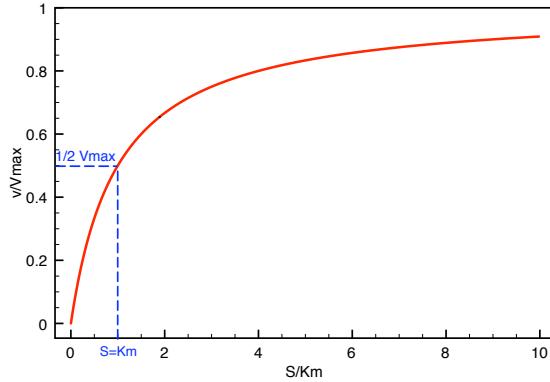


Figure 6.1: A sketch of the relative activity of an irreversible enzyme following Michaelis-Menten kinetics (equation 6.3) as function of its normalized substrate concentration.

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,

$$\begin{aligned} 0 &= -(k_1^+ e \cdot s - k_1^- es) + k_2 es \\ 0 &= k_1^+ e \cdot s - k_1^- es - k_2 es \end{aligned} \quad (6.6)$$

These equations allow 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  $es$ ,

$$\begin{aligned} 0 &= k_1^+ (e_T - es) \cdot s - k_1^- es - k_2 es \\ &= k_1^+ e_T \cdot s - k_1^+ es \cdot s - k_1^- es - k_2 es \\ &= k_1^+ e_T \cdot s - es(k_1^+ s + k_1^- + k_2) \Rightarrow \\ es &= \frac{k_1^+ e_T \cdot s}{k_1^+ s + k_1^- + k_2} \\ &= \frac{e_T \frac{k_1^+}{k_1^- + k_2} \cdot s}{\frac{k_1^+}{k_1^- + k_2} s + 1} \end{aligned} \quad (6.7)$$

The rate of the enzyme under quasi-steady state conditions equals  $v = v_1 = v_2 = k_2es$ ,

$$\begin{aligned} v &= k_2es \\ &= k_2e_T \frac{\frac{k_1^+}{k_1^- + k_2} \cdot s}{\frac{k_1^+}{k_1^- + k_2}s + 1} \\ &= V_{MAX} \frac{\frac{s}{K_M}}{\frac{s}{K_M} + 1} = V_{MAX} \frac{s}{s + K_M} \end{aligned} \quad (6.8)$$

Here the maximal rate of the enzyme,  $V_{MAX}$ , is defined as  $k_2e_T$  and the Michaelis-Menten constant as  $K_M = \frac{k_1^- + k_2}{k_1^+}$ . Note that the affinity of the enzyme can be defined as  $1/K_M$ .

### 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.

#### 6.2.2 Derivation of enzyme kinetics: equilibrium-binding assumption

Again we assume that the substrate level is fixed,



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,

$$k_1^+ e \cdot s = k_1^- es \Rightarrow e = \frac{k_1^- es}{k_1^+ s} = K_S \frac{es}{s} \quad (6.10)$$

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  $es$ ,

$$\begin{aligned} e_T &= e + es = es \left( \frac{K_S}{s} + 1 \right) \Rightarrow \\ es &= \frac{e_T}{\frac{K_S}{s} + 1} = \frac{e_T \frac{s}{K_S}}{\frac{s}{K_S} + 1} = \frac{e_T s}{s + K_S} \end{aligned} \quad (6.11)$$

Again the rate of the enzyme equals  $v = v_2 = k_2es$  and therefore,

$$v = k_2e_T \frac{s}{s + K_S} = V_{MAX} \frac{s}{s + K_S} \quad (6.12)$$

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?

#### 6.2.3 What's worse: assuming quasi-steady state or rapid equilibrium?

With the quasi-steady state assumption we found,

$$K_M = \frac{k_1^- + k_2}{k_1^+} \quad (6.13)$$

whereas the rapid-equilibrium assumption gave rise to the following result,

$$K_S = \frac{k_1^-}{k_1^+} \quad (6.14)$$

Those two constants play the same role because we found that,

$$\text{Rapid-equilibrium assumption: } v = V_{MAX} \frac{s}{s + K_S} \quad (6.15)$$

$$\text{Quasi-steady state assumption: } v = V_{MAX} \frac{s}{s + K_M} \quad (6.16)$$

$K_S$  equals  $K_M$  if  $k_1^- \gg k_2$ . Hence, this indicates that the rapid-equilibrium assumption is more stringent: it assumes also that the dissociation of  $es$  into  $s$  and  $e$  occurs much more often than the conversion  $es \rightarrow e + p$ .

#### 6.2.4 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 a mechanism for modulation of 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,

$$e_T = e + ex + es + esx \quad (6.17)$$

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),

$$e_T = e \left( 1 + \frac{x}{K_1} \right) + es \left( 1 + \frac{x}{K_2} \right) \quad (6.18)$$

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  $e + s \rightleftharpoons es$  is in equilibrium: then,

$$\begin{aligned} e_T &= K_S \frac{es}{s} \left( 1 + \frac{x}{K_1} \right) + es \left( 1 + \frac{x}{K_2} \right) \\ &= es \left( \frac{K_S}{s} \left( 1 + \frac{x}{K_1} \right) + \left( 1 + \frac{x}{K_2} \right) \right) \Rightarrow \\ es &= \frac{e_T}{\frac{K_S}{s} \left( 1 + \frac{x}{K_1} \right) + \left( 1 + \frac{x}{K_2} \right)} \\ &= \frac{e_T}{\left( 1 + \frac{x}{K_2} \right)} \frac{1}{\frac{K_S}{s} \frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}} + 1} \\ &= \frac{e_T}{\left( 1 + \frac{x}{K_2} \right)} \frac{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}}{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}} + 1} \\ &= \frac{e_T}{\left( 1 + \frac{x}{K_2} \right)} \frac{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}}{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}} + 1} \end{aligned} \quad (6.19)$$

And the rate of the enzyme now becomes,

$$\begin{aligned} v &= \frac{k_2 e_T}{\left( 1 + \frac{x}{K_2} \right)} \frac{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}}{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}} + 1} \\ &= V_{MAX}^{APP} \frac{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}}{\frac{s}{K_S} \frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}} + 1} \\ &= V_{MAX}^{APP} \frac{s}{s + K_S^{APP}} \end{aligned} \quad (6.20)$$

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.  $V_{MAX}^{APP}$  and  $K_S^{APP}$  are defined as,

$$\begin{aligned} V_{MAX}^{APP} &= \frac{k_2 e_T}{1 + \frac{x}{K_2}} = \frac{V_{MAX}}{1 + \frac{x}{K_2}} \\ K_M^{APP} &= K_S \frac{1 + \frac{x}{K_1}}{1 + \frac{x}{K_2}} \end{aligned} \quad (6.21)$$

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

### 6.2.6 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,

$$v = V_{MAX} \frac{s^n}{K_S^n + s^n} \quad (6.22)$$

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  $n$ ,

$$\frac{\partial \ln v}{\partial \ln s} = \frac{s}{v} \frac{\partial v}{\partial s} = n \frac{K_S^n}{K_S^n + S^n} \quad (6.23)$$

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.

### 6.3 Reversible Michaelis-Menten kinetics: One enzyme-reactant state

Monomeric enzymes have only one catalytic unit. We will consider enzymes with multiple subunits in Chapter 7. The simplest reversible enzyme mechanism considers an enzyme,  $E$ , that converts a single substrate  $S$  into a single product  $P$ ,



The two reactions are considered reversible and described by mass-action kinetics.  $ES$  is often referred to as the enzyme-substrate complex. Reactions with a single substrate and single product are called uni-uni reactions; two substrates and a single product, bi-uni reactions, etc.

#### 6.3.1 Quasi-steady state approximation

At quasi-steady state we have,

$$\begin{aligned} e_T &= e + s \\ \frac{des}{dt} &= k_1^+ \cdot e \cdot s - k_1^- \cdot es + k_2^- \cdot e \cdot p - k_2^+ \cdot es = 0 \\ \Rightarrow es &= e \cdot \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} = e \cdot \frac{\gamma}{\delta} \\ \text{With: } e_T &= e + es \Rightarrow e = \frac{e_T}{1 + \frac{\gamma}{\delta}} \text{ and } es = \frac{\frac{\gamma}{\delta} e_T}{1 + \frac{\gamma}{\delta}} \end{aligned} \quad (6.25)$$

The rate of the enzyme equals,

$$\begin{aligned} v = v_2 &= k_2^+ \cdot es - k_2^- \cdot e \cdot p = \frac{k_2^+ \gamma e_T - k_2^- \delta e_T p}{\delta + \gamma} \\ &= \frac{k_2^+ k_1^+ s e_T + k_2^+ e_T k_2^- p - k_2^- k_1^- e_T p - k_2^- k_2^+ e_T p}{k_1^- + k_2^+ + k_1^+ s + k_2^- p} \\ &= \frac{k_2^+ k_1^+ s e_T - k_2^- k_1^- e_T p}{k_1^- + k_2^+ + k_1^+ s + k_2^- p} = \frac{V_{MAX}^+ \frac{s}{K_s} - V_{MAX}^- \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \end{aligned} \quad (6.26)$$

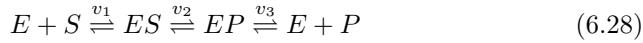
With  $K_s = \frac{k_1^- + k_2^+}{k_1^+}$ ,  $K_p = \frac{k_1^- + k_2^+}{k_2^-}$ ,  $V_{MAX}^+ = k_2^+ e_T$ , and  $V_{MAX}^- = k_1^- e_T$ . So,

$$v = \frac{V_{MAX}^+ \frac{s}{K_s} - V_{MAX}^- \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \quad (6.27)$$

is the reversible Michaelis-Menten equation that simplifies to the irreversible Michaelis-Menten equation when  $p = 0$ .

## 6.4 Reversible Michaelis-Menten kinetics: Two enzyme-reactant states

Next, we consider<sup>1</sup>



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.

### 6.4.1 Rapid-equilibrium approximation

Again the simplest derivation requires the rapid-equilibrium approximation. We assume that the rates of  $v_1$  and  $v_3$  are sufficiently close to zero to set them equal to zero. Then, we have,

$$\begin{aligned} es &= \frac{e \cdot s}{K_1} \\ ep &= \frac{e \cdot p}{K_3} \\ e_T &= e + es + ep = e \left( 1 + \frac{s}{K_1} + \frac{p}{K_3} \right) \\ \Rightarrow e &= \frac{e_T}{1 + \frac{s}{K_1} + \frac{p}{K_3}} \end{aligned} \quad (6.29)$$

So, the conversion of  $es \rightleftharpoons ep$  is the rate limiting step in the enzyme mechanism and therefore sets the reaction rate  $v$  for the conversion  $S \rightleftharpoons P$ ,

$$\begin{aligned} v &= k_2^+ \cdot es - k_2^- \cdot ep \\ &= \frac{k_2^+ \cdot e_T \cdot \frac{s}{K_1} - k_2^- \cdot e_T \cdot \frac{p}{K_3}}{1 + \frac{s}{K_1} + \frac{p}{K_3}} \end{aligned} \quad (6.30)$$

This last equation is often written as,

$$v = \frac{V_{max}^+ \cdot \frac{s}{K_S} - V_{max}^- \cdot \frac{p}{K_P}}{1 + \frac{s}{K_S} + \frac{p}{K_P}} \quad (6.31)$$

Where the forward maximal rate equals  $V_{max}^+ = k_2^+ e_T$ , the backward maximal rate  $V_{max}^- = k_2^- e_T$ ,  $K_S = K_1$  and  $K_P = K_3$ .

---

<sup>1</sup>A comment about this scheme. This scheme implies that  $S$  and  $P$  cannot bind simultaneously to the enzyme, as otherwise the state  $ESP$  would have existed. Therefore, having a higher concentration of  $P$  goes at the expense of  $es$  and a higher concentration of  $S$  goes at expense of  $ep$ . Hence,  $S$  and  $P$  compete for binding to  $E$  and the rate of  $P$  formation out of  $S$  is inhibited by higher concentrations of  $P$  and, likewise, the rate of  $S$  formation out of  $P$  is inhibited by higher concentrations of  $S$ . This is already evident from the scheme and the enzyme rate equation derived in this section agrees with this intuition. So, studying these schemes gives you some useful definition.

### 6.4.2 Quasi-steady state approximation

The mass balances for all the species in mechanism 6.28 are given by,

$$\begin{aligned}\frac{ds}{dt} &= -v_1 \\ \frac{de}{dt} &= -v_1 + v_3 \\ \frac{des}{dt} &= v_1 - v_2 \\ \frac{dep}{dt} &= v_2 - v_3 \\ \frac{dp}{dt} &= v_3\end{aligned}\tag{6.32}$$

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,

$$\begin{aligned}\frac{de}{dt} &= -v_1 + v_3 = 0 \\ \frac{des}{dt} &= v_1 - v_2 = 0 \\ \frac{dep}{dt} &= v_2 - v_3 = 0\end{aligned}\tag{6.33}$$

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

$$e_{tot} = e + es + ep\tag{6.34}$$

This equation can be checked to be true from equation 6.33 as,

$$\frac{de}{dt} + \frac{des}{dt} + \frac{dep}{dt} = 0 \quad (6.35)$$

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

$$v = dp/dt = v_3 = k_3^+ es - k_3^- \cdot e \cdot p \quad (6.36)$$

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 6.36. As the equations in equation 6.33 are linearly dependent - they obey equation 6.35 - we need to use the conservation of total enzyme, equation 6.34, when solving for the enzyme species. This can be done by hand (as was done in section 6.2) 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,<sup>2</sup>

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1^- - k_2^+ & k_2^- \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e \\ es \\ ep \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \quad (6.37)$$

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 6.2) for the three enzyme species,

$$\begin{pmatrix} e \\ es \\ ep \end{pmatrix} = - \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1^- - k_2^+ & k_2^- \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \quad (6.38)$$

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 6.36 gives,

$$\frac{v}{e_{tot}} = \frac{\overbrace{k_1^+ k_2^+ k_3^+}^{num_1} s - \overbrace{k_1^- k_2^- k_3^-}^{num_2} p}{\underbrace{k_1^- k_2^- + k_1^- k_3^+ + k_2^+ k_3^+}_{const} + \underbrace{k_1^+ (k_2^- + k_2^+ + k_3^+)}_{coef_s} s + \underbrace{k_3^- (k_1^- + k_2^- + k_2^+)}_{coef_p} p} \quad (6.39)$$

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 [34]. The maximal rate of the enzyme in the forward and the backward direction are defined as,

$$V_{MAX}^+ = \frac{num_1}{coef_s} \quad (6.40)$$

$$V_{MAX}^- = \frac{num_2}{coef_p}$$

---

<sup>2</sup>Do this yourself once to convince yourself that can to this.

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

$$\begin{aligned} K_{MS} &= \frac{\text{const}}{\text{coef}_s} \\ K_{MP} &= \frac{\text{const}}{\text{coef}_p} \end{aligned} \quad (6.41)$$

Substitution of these equations into 6.39 gives the reversible Michaelis-Menten rate equation,

$$v = \frac{V_{MAX}^+ \frac{s}{K_{MS}} - V_{MAX}^- \frac{p}{K_{MP}}}{1 + \frac{s}{K_{MS}} + \frac{p}{K_{MP}}} \quad (6.42)$$

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 6.42 simplifies into,

$$v = V_{MAX}^+ \frac{s}{s + K_{MS}} \quad (6.43)$$

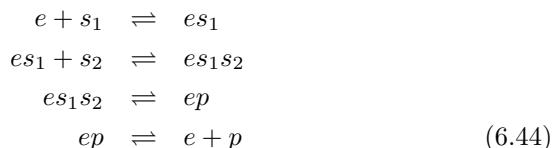
A sketch of this curve is plotted in figure 6.1. A number of conditions clarify the meaning of the terms in this equation and give rise to a number of frequently used concepts,

1. If  $s \gg K_{MS}$  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  $s \ll K_{MS}$  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 6.42 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 6.42 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 6.42.

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 its catalytic mechanism,



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.

## 6.5 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 6.2) 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.<sup>3</sup>

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

---

<sup>3</sup>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.

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,

$$\mu_A = \mu_A^{0'} + RT \ln a \quad (6.45)$$

The unit of molar Gibbs free energy is  $J/mol$ , the universal gas constant  $R$  has as its unit  $J/(mol \cdot K)$ , and temperature  $T$  is in Kelvin.<sup>4</sup> 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,  $\Delta G_R$ , is the difference in Gibbs free energy of the products and the substrates taking into account their stoichiometric coefficients. For the reaction  $2A + B \rightleftharpoons A_2B$  we obtain,

$$\Delta G_R = \mu_{A_2B} - 2\mu_A - \mu_B = \Delta G_R^{0'} + RT \ln \frac{a_2b}{a^2 \cdot b} \quad (6.46)$$

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

$$\frac{a_2b_{EQ}}{(a_{EQ})^2 \cdot b_{EQ}} = e^{-\frac{\Delta G_R^{0'}}{RT}} \equiv K_{EQ} \quad (6.47)$$

In this equation, the equilibrium concentrations appear (subscript EQ) and the equilibrium constant,  $K_{EQ}$ . The actual ratio  $\frac{a_2b}{a^2 \cdot b}$  is defined as the mass action ratio  $\Gamma$ . The deviation from thermodynamic equilibrium is captured by  $\Gamma/K_{EQ}$ . Indeed, using equation 6.46 and 6.47 we can write the Gibbs free energy potential of the reaction as,

$$\Delta G_R = RT \ln \frac{\Gamma}{K_{EQ}} \quad (6.48)$$

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:  $v = v^+ - v^- = k^+a^2b - k^-a_2b$ .

We can rewrite equation 6.42 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,

$$K_{EQ} = \frac{p_{EQ}}{s_{EQ}} = \frac{V_{MAX}^+ K_{MP}}{V_{MAX}^- K_{MS}} \quad (6.49)$$

This relationship is known as the Haldane relationship, which allows us to rewrite equation 6.42 as,

$$v = \frac{V_{MAX}^+ \frac{s}{K_{MS}} \left(1 - \frac{p}{sK_{EQ}}\right)}{1 + \frac{s}{K_{MS}} + \frac{p}{K_{MP}}} \quad (6.50)$$

---

<sup>4</sup> $R = N_A k_B$  where  $N_A$  is Avogadro's constant and  $k_B$  Boltzman's constant.

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

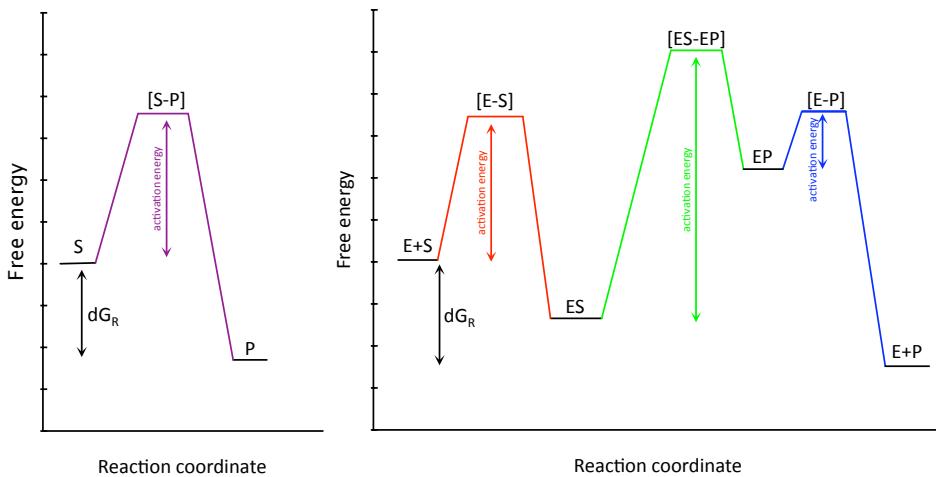


Figure 6.2: A plot of the free energy of representative states for the reaction of  $S$  to  $P$  as function of the reaction progression (reaction coordinate) when the reaction is enzyme spontaneous ( $S \rightleftharpoons P$ , left) and enzyme-catalyzed ( $E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P$ , right). In transition state theory, enzymes enhance the speed of reactions by lowering the activation energy for the reaction. Enzymes achieve 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 equilib-

rium. 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?

- 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  $K_m$ 's to 1 and  $V_{max}$ 's to 10. In the first model, you set the equilibrium constant of the second enzyme 1 and determine the  $V_{max}$  of this enzyme to have it operate 10% from thermodynamic equilibrium at steady state by judging  $\Gamma/K_{eq}$ . In the second model, set the equilibrium constant to a 100 and determine again the value for the  $V_{max}$  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  $V_{max}$  is perturbed?

## 6.6 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 extend the mechanism underlying the Michaelis-Menten kinetic rate equation with the action of an inhibitor as shown in figure 6.3. 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,

$$\begin{aligned} e_T &= e + ei + es + esi + ep + epi \\ &= e \left(1 + \frac{i}{K_{i,4}}\right) + es \left(1 + \frac{i}{K_{i,5}}\right) + ep \left(1 + \frac{i}{K_{i,6}}\right) \end{aligned} \quad (6.51)$$

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,

$$\begin{pmatrix} e \\ es \\ ep \end{pmatrix} = \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1^- - k_2^+ & k_2^- \\ 1 + \frac{i}{K_{i,4}} & 1 + \frac{i}{K_{i,5}} & 1 + \frac{i}{K_{i,6}} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \quad (6.52)$$

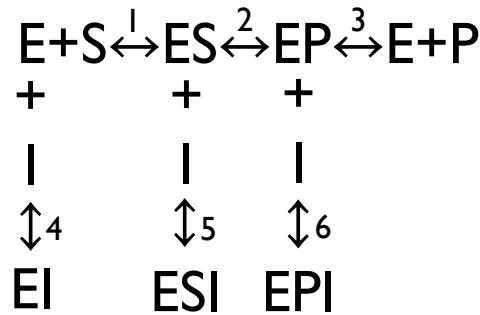


Figure 6.3: General catalytic mechanism for inhibition of an uni-uni enzyme.

The rate equation now corresponds to,

$$\begin{aligned} \frac{v}{e_{tot}} &= \frac{num_1 s - num_2 p}{const + coef_s s + coef_p p} \\ num_1 &= k_1^+ k_2^+ k_3^+ \\ num_2 &= k_1^- k_2^- k_3^- \\ const &= \left(1 + \frac{i}{K_{i,4}}\right) (k_1^- k_2^- + k_1^- k_3^+ + k_2^- k_3^+) \\ coef_s &= \left(1 + \frac{i}{K_{i,5}}\right) k_1^+ (k_2^- + k_2^+ + k_3^+) \\ coef_p &= \left(1 + \frac{i}{K_{i,5}}\right) k_3^- (k_1^- + k_2^- + k_2^+) \end{aligned} \quad (6.53)$$

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 6.41 and 6.42, the enzyme kinetic properties can be expressed in terms of the those

derived in the previous section,

$$\begin{aligned}
 V_{MAX}^{+,APP} &= \frac{V_{MAX}^+}{1 + \frac{i}{K_{i,4}}} \\
 V_{MAX}^{-,APP} &= \frac{V_{MAX}^-}{1 + \frac{i}{K_{i,4}}} \\
 K_{MS}^{APP} &= \frac{1 + \frac{i}{K_{i,4}}}{1 + \frac{i}{K_{i,5}}} K_{MS} \\
 K_{MP}^{APP} &= \frac{1 + \frac{i}{K_{i,4}}}{1 + \frac{i}{K_{i,5}}} K_{MP}
 \end{aligned} \tag{6.54}$$

On the basis of these equations different forms of inhibition can be distinguished as shown in Table 6.1. 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).

Table 6.1: 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)       | $\frac{V_{MAX}^+}{1 + \frac{i}{K_{i,5}}}$ | $V_{MAX}^+/K_{MS}$                               | $\frac{K_{MS}}{1 + \frac{i}{K_{i,5}}}$                       |

## Exercises

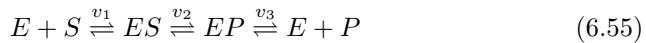
1. Is competitive or uncompetitive inhibition a more potent mechanism for inhibition?
2. Make a kinetic model of a metabolic pathway with three enzymes each catalyzing an uni-uni reaction. Make the first reaction irreversible and product-independent. Choose the other two enzymes as reversible Michaelis-Menten kinetics. Set the equilibrium constants to 100, all Km's to 1, all  $V_{max}^+$  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?

## 6.7 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 6.28),



The enzyme conservation equation equals,

$$e_T = e + es + ep \quad (6.56)$$

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,

$$e_T = e \left( 1 + \frac{s}{K_S} + \frac{p}{K_P} \right) \quad (6.57)$$

$K_S$  (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,

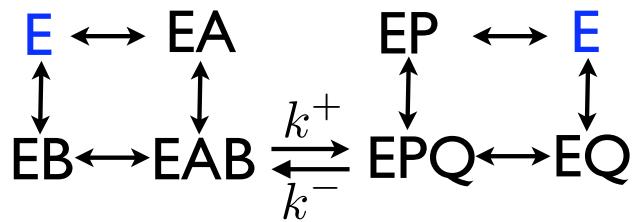
$$v = k_2^+ es - k_2^- ep \quad (6.58)$$

Using the dissociation constant definition and equation 6.57 we obtain for the rate equation,

$$v = \frac{V_{MAX}^+ \frac{s}{K_S} - V_{MAX}^- \frac{p}{K_P}}{1 + \frac{s}{K_S} + \frac{p}{K_P}} \quad (6.59)$$

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!

A.



B.

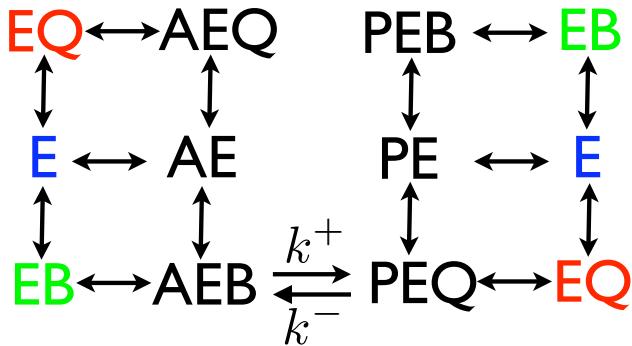


Figure 6.4: Two catalytic mechanism for a two-substrate and two-product reaction, a so-called bi-bi reaction. The equilibrium binding models will be different for the two mechanisms. Species that appear twice in the mechanisms are coloured. A. The enzyme has a catalytic site where only substrates and 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.

The power of equilibrium models derives from its straightforward derivation of multi-reactant rate equations. For instance, consider the enzyme mechanisms shown in figure 6.4. For mechanism A, we have the following conservation relation,

$$e_T = e + ea + eb + eab + ep + eq + epq \quad (6.60)$$

and for mechanism B,

$$e_T = e + ae + eb + aeb + aeq + pe + eq + peb + peq \quad (6.61)$$

Using the equilibrium binding assumption we obtain for mechanism A,

$$e_T = e \left( 1 + \frac{a}{K_a} + \frac{b}{K_b} + \frac{a \cdot b}{K_a K_b} + \frac{p}{K_p} + \frac{q}{K_q} + \frac{p \cdot q}{K_p K_q} \right) \quad (6.62)$$

whereas for mechanism B the expression can be simplified to,

$$e_T = e \left( 1 + \frac{a}{K_a} + \frac{p}{K_p} \right) \left( 1 + \frac{b}{K_b} + \frac{q}{K_q} \right) \quad (6.63)$$

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,

$$v = \frac{V_{MAX}^+ \frac{a \cdot b}{K_a K_b} - V_{MAX}^- \frac{p \cdot q}{K_p K_q}}{1 + \frac{a}{K_a} + \frac{b}{K_b} + \frac{a \cdot b}{K_a K_b} + \frac{p}{K_p} + \frac{q}{K_q} + \frac{p \cdot q}{K_p K_q}} \quad (6.64)$$

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

$$v = \frac{V_{MAX}^+ \frac{a \cdot b}{K_a K_b} - V_{MAX}^- \frac{p \cdot q}{K_p K_q}}{\left( 1 + \frac{a}{K_a} + \frac{p}{K_p} \right) \left( 1 + \frac{b}{K_b} + \frac{q}{K_q} \right)} \quad (6.65)$$

These equations have the same Haldane relationship,

$$K_{EQ} = \frac{p_{EQ} q_{EQ}}{a_{EQ} b_{EQ}} = \frac{V_{MAX}^+ K_p K_q}{V_{MAX}^- K_a K_b} \quad (6.66)$$

(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,

$$V_{MAX}^+ \frac{a \cdot b}{K_a K_b} \left( 1 - \frac{p \cdot q}{a \cdot b \cdot K_{EQ}} \right) = V_{MAX}^+ \frac{a \cdot b}{K_a K_b} \left( 1 - \frac{\Gamma}{K_{EQ}} \right) \quad (6.67)$$

## Exercises

In figure 6.5, 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

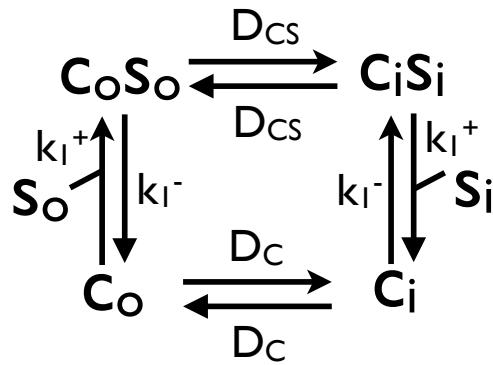


Figure 6.5: Mechanism for a membrane transporter acting as a facilitated diffusion carrier. The transport solute occurs at the external cell side with concentration,  $s_o$ , where it binds the carrier, occurring at concentration,  $c_o$ . As the carrier diffuses continuously through the membrane it sometimes occurs at the intracellular side, at concentration  $c_i s_i$ , where it can deposit the solute.

carrier through the membrane. Given those assumptions write the rate equation for the transporter in the following form,

$$v = V_{max} \frac{\frac{s_o}{K_m} - \frac{s_i}{K_m}}{1 + \frac{s_o}{K_m} + \frac{s_o}{K_m} + K_i \frac{s_i s_o}{K_m^2}} \quad (6.68)$$

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.

# Chapter 7

## Cooperative enzymes

### 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.

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 7.1 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,

$$s_{v/V_{MAX}} = K_M \frac{v/V_{MAX}}{1 - v/V_{MAX}} \quad (7.1)$$

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 hemoglobin and found a sigmoidal saturation curve, as the gray curve marked with positive cooperativity in figure 7.1. He fitted the following phenomenological equation to this relationship, now known as the Hill equation,

$$\frac{y}{y_{max}} = \frac{x^h}{K_{0.5}^h + x^h} \quad (7.2)$$

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.

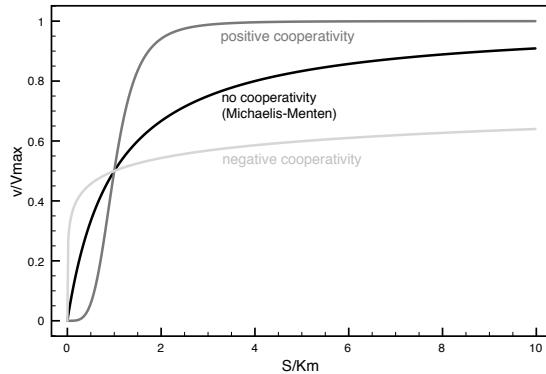


Figure 7.1: Cooperative enzymes and regulation. The saturation of three enzymes,  $v/V_{MAX}$  (their normalized rates), is plotted as function of their normalized substrate concentrations,  $S/K_M$ . They differ markedly in their sensitivity to the concentration of their substrate. The enzyme displaying 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,  $v/V_{MAX}$ .

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,

$$R = \frac{x_{0.9}}{x_{0.1}} = 81^{1/h} \quad (7.3)$$

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.

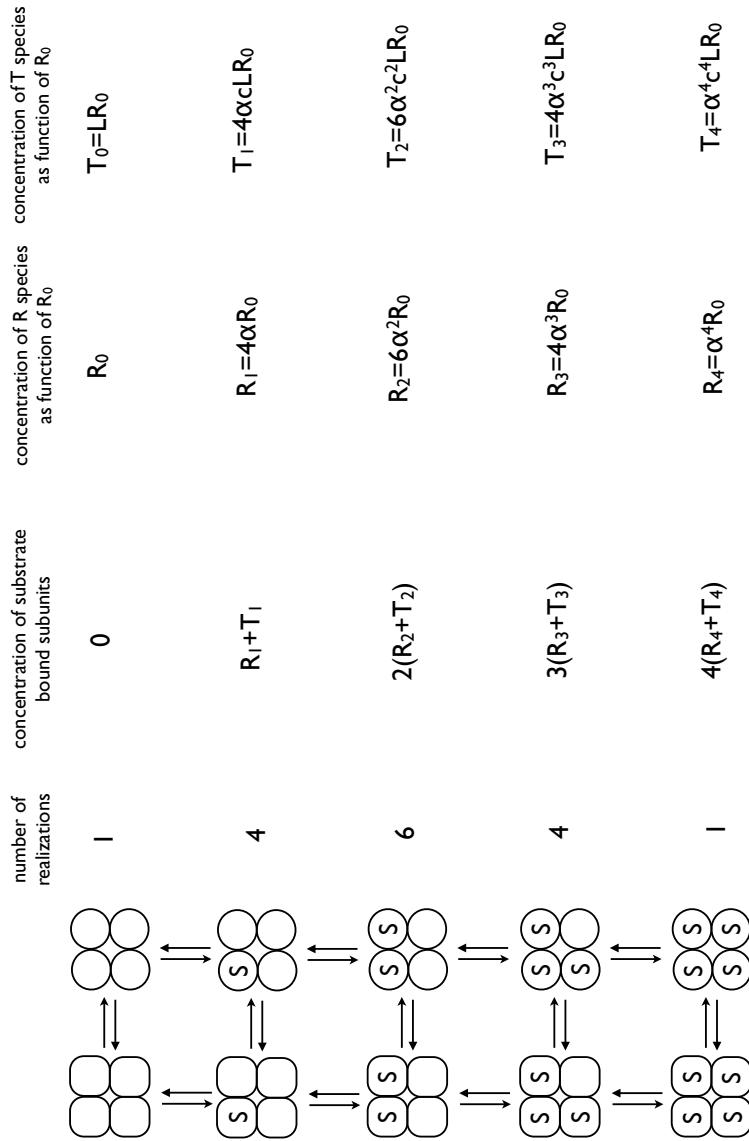


Figure 7.2: MWC scheme

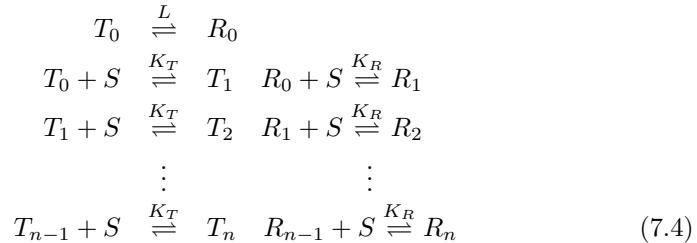
## 7.2 The Monod Wyman Changeux model for cooperative enzyme kinetics

In this section we will consider the model presented by Monod, Wyman, and Changeux (MWC) in 1965. The MWC model assumes the multimeric enzyme 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. In this section, we will use the MWC model to derive enzyme kinetics. The MWC model can be used to study GPCRs or TFs as well, which will be done in the next chapter.

An alternative model for cooperative enzymes derived by Koshland & Nemethy [28] does not make the concerted-symmetry assumption. It is therefore more general and, regrettably 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 [18] (see next section) and Popova & Selkov [31].

The following reactions are considered in the MWC model and they are all considered in thermodynamic equilibrium (hence, only the dissociation constants of the reactions are mentioned and not the rate constants),



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,

$$\begin{aligned}
 \alpha &= \frac{s}{K_R} \\
 L &= \frac{t_0}{r_0} \\
 c &= \frac{K_R}{K_T}
 \end{aligned} \tag{7.5}$$

The derivation of the MWC model can be illustrated nicely with an example of a cooperative enzyme with four subunits (figure 7.2). The activity of the enzyme is given by,

$$\begin{aligned}
 \frac{v}{V_{MAX}} &= \frac{\text{concentration of all substrate bound subunits}}{\text{concentration of all subunits}} \\
 &= \frac{1 \cdot 4 \cdot r_1 + 2 \cdot 6 \cdot r_2 + 3 \cdot 4 \cdot r_3 + 4 \cdot 1 \cdot r_4 + 1 \cdot 4 \cdot t_1 + 2 \cdot 6 \cdot t_2 + 3 \cdot 4 \cdot t_3 + 4 \cdot 1 \cdot t_4}{4(r_0 + 4r_1 + 6r_2 + 4r_3 + r_4 + t_0 + 4t_1 + 6t_2 + 4t_3 + t_4)}
 \end{aligned} \tag{7.6}$$

Studying this equation in more detail immediately shows the logic behind this equation.

Let's start with the numerator. It contains all the bound states and equals the total concentration of the substrate bounds subunits. The terms  $1 \cdot 4 \cdot R_1$  means that 1 S is bound and this multiplied by the number of forms of this state, which is  $4 - R_1$  occurs in four forms with subunit A, B, C, or D bound to S if we would label the subunits with the letters A to D. In fact, the four equals the number of combinations one can make with 1 substrate molecule and four subunits, i.e.

$$\binom{4}{1} = \frac{4!}{3!1!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)1} = 4 \quad (7.7)$$

Why does the term  $2 \cdot 6 \cdot R_2$  appear? The '2' gives the number of occupied subunits per state and the '6' derives from  $\binom{4}{2} = 6$ . The state with 2 S bound comes in 6 forms: if we denote occupied by X and empty by O then we have the following possibilities with 4 sites: XXOO, XOXO, XOOX, OXXO, OOXO, and OOXO. Indeed 6 forms. They will each have the same concentration because the four sites have the same affinity and hence the concentration of bound subunits equals the total concentrations of  $R_2$  states times the number of bound subunit per state,  $2 \cdot 6 \cdot R_2$ . The state  $R_3$  has 4 forms: OXXX, XOXX, XOXO, and XXXO and 3 bound subunits per state: thus get  $3 \cdot 4 \cdot r_3$ . So, in general for a protein with m subunits for the state of  $r_n$  we have for its numerator term,  $n \cdot \binom{m}{n} \cdot r_n$  with  $n = 1, 2, \dots, m$ .

The denominator counts the total number of subunits, regardless of whether they are occupied or not with S. Hence, every term will be the number of subunits, 4, times the number of forms of the state, and times the concentration of the state, so:  $m \cdot \binom{m}{n} \cdot r_n$  (with  $n = 0, 1, \dots, m$ ). The last equation can simplified using the definition of the dissociation constant for the binding equilibria of S to the cooperative enzyme,

$$\begin{aligned} \frac{v}{V_{MAX}} &= \frac{1 \cdot 4 \cdot r_1 + 2 \cdot 6 \cdot r_2 + 3 \cdot 4 \cdot r_3 + 4 \cdot 1 \cdot r_4 + 1 \cdot 4 \cdot t_1 + 2 \cdot 6 \cdot t_2 + 3 \cdot 4 \cdot t_3 + 4 \cdot 1 \cdot t_4}{4(r_0 + 4r_1 + 6r_2 + 4r_3 + r_4 + t_0 + 4t_1 + 6t_2 + 4t_3 + t_4)} \\ &= \frac{4 \cdot r_0 \cdot \frac{s}{K_R} + 2 \cdot 6 \cdot r_0 \cdot \frac{s^2}{K_R^2} + 3 \cdot 4 \cdot r_0 \cdot \frac{s^3}{K_R^3} + 4 \cdot 1 \cdot r_0 \cdot \frac{s^4}{K_R^4} + \text{the same for the T states}}{4 \left( r_0 + 4 \cdot r_0 \cdot \frac{s}{K_R} + 6 \cdot r_0 \cdot \frac{s^2}{K_R^2} + 4 \cdot r_0 \cdot \frac{s^3}{K_R^3} + r_0 \cdot \frac{s^4}{K_R^4} + \text{the same for the T states} \right)} \\ &= \frac{\underbrace{\alpha(1+\alpha)^3}_{(1+\alpha)^4} + \underbrace{\alpha L c (1+c\alpha)^3}_{L(1+c\alpha)^4}}{\underbrace{1 + \alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4}_{(1+\alpha)^4} + \underbrace{L + \alpha c L + 3\alpha^2 c^2 L + 3\alpha^3 c^3 L + \alpha^4 c^4 L}_{L(1+c\alpha)^4}} \\ &= \frac{\alpha(1 + \alpha)^3 + c\alpha L(1 + c\alpha)^3}{(1 + \alpha)^4 + L(1 + c\alpha)^4} \end{aligned} \quad (7.8)$$

In the derivation, from line 2 to 3 all the terms in the denominator and numerator are divided by  $r_0$  and the resulting terms  $t_0/r_0$  are then identified as the  $L$  coefficient and  $c$  is defined as  $K_R/K_T$ .

### 7.2.1 Incorporation of allosteric modifiers

Monod, Wyman, and Changeux also showed how to incorporate allosteric regulators into the MWC equation through the  $L$  coefficient. The  $L$  coefficient is defined as,

$$L = \frac{t_0}{r_0} \quad (7.9)$$

MWC considered effectors that only influence the conformation equilibria, and, hence, not the affinity of the  $R$  and  $T$  state for the reactant(-s). If all of this occurs independently then all the  $R$  and  $T$  states concentrations sum to,

$$\begin{aligned} r &= r_0 \underbrace{\left(1 + \frac{A}{K_a}\right)^n}_{A \text{ binding to } R} (1 + \alpha)^n = \text{sum of all concentration of the R state monomers} \\ t &= t_0 \underbrace{\left(1 + \frac{I}{K_i}\right)^n}_{I \text{ binding to } T} (1 + c\alpha)^n = \text{sum of all concentration of the T state monomer} \end{aligned} \quad (7.10)$$

The new conformation equilibrium constant is then denoted by  $L'$  coefficient and is defined as,

$$L' = \frac{r_0 \left(1 + \frac{A}{K_a}\right)^n}{t_0 \left(1 + \frac{I}{K_i}\right)^n} = L \cdot \frac{\left(1 + \frac{A}{K_a}\right)^n}{\left(1 + \frac{I}{K_i}\right)^n} \quad (7.11)$$

### 7.2.2 Generalisation to any number of subunits

In the general case, we then obtain,

$$\begin{aligned} \frac{v}{V_{MAX}} &= \text{fraction of subunits bound to S} \\ &= \frac{\text{total amount of subunits bound to S}}{\text{total amount of subunits}} \\ &= \frac{\underbrace{\alpha (1 + \alpha)^{n-1}}_{\text{amount of R states bound to S}} + \underbrace{L \cdot c \cdot \alpha (1 + c \cdot \alpha)^{n-1}}_{\text{amount of T state bound to S}}}{\underbrace{(1 + \alpha)^n}_{\text{total amount of R state}} + \underbrace{L \cdot (1 + c \cdot \alpha)^n}_{\text{total amount of T state}}} \end{aligned} \quad (7.12)$$

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,

$$L' = L \frac{(1 + \beta)^n}{(1 + \gamma)^n} \quad (7.13)$$

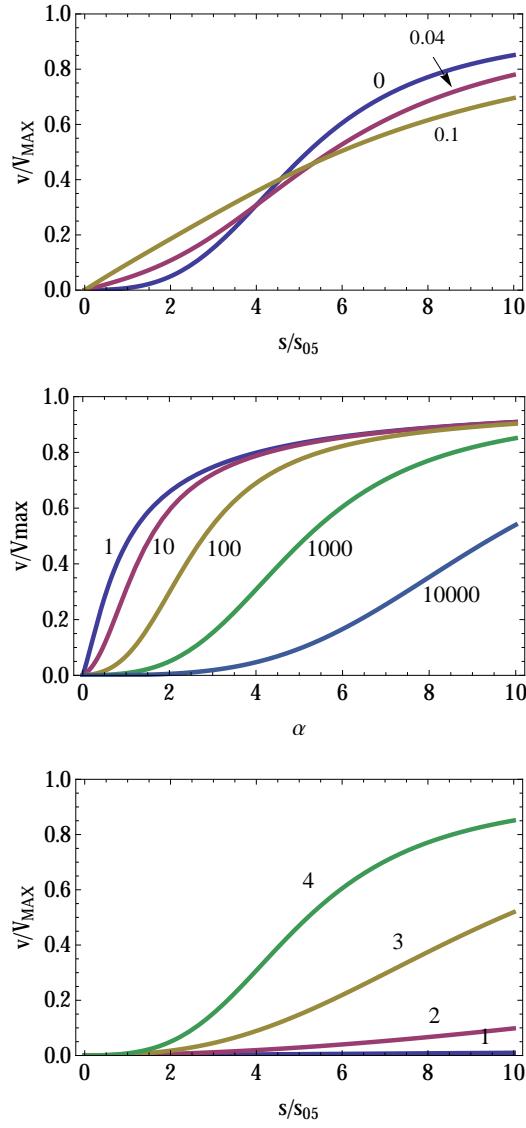


Figure 7.3: Illustrations of the consequence of cooperativity parameters on the MWC rate equation (equation 7.12). In the upper, middle and lower plot, the value of the  $c$ ,  $L$ , and  $n$  parameter were changed, respectively.

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

$$v = \frac{v}{V_{MAX}} \frac{\alpha (1 + \alpha)^{n-1} + L \cdot c \cdot \alpha (1 + c \cdot \alpha)^{n-1}}{(1 + \alpha + \beta)^n + L \cdot (1 + c \cdot \alpha + d \cdot \beta)^n} \quad (7.14)$$

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

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

$$v = V_{MAX} \frac{pep \cdot adp \left( \frac{pep}{K_{pep}} + 1 \right)^{n-1}}{K_{pep} \left( L \left( \frac{1 + \frac{atp}{K_{atp}}}{\frac{fdp}{K_{fdp}} + \frac{amp}{K_{amp}} + 1} \right)^n + \left( \frac{pep}{K_{pep}} + 1 \right)^n \right) (adp + K_{adp})} \quad (7.15)$$

The kinetic parameters are:  $K_{pep} = 0.31 \text{ mM}$ ,  $K_{adp} = 0.26 \text{ mM}$ ,  $K_{amp} = 0.2 \text{ mM}$ ,  $K_{fdp} = 0.19 \text{ mM}$ ,  $K_{atp} = 22.5 \text{ mM}$ ,  $L = 1000$ , 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 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.

- The intricate regulation of glycolysis and occurrence of the cooperative enzyme, phosphofructokinase, pyruvate decarboxylase, and purvate kinase, inspired the analysis of kinetic models of glycolysis. A glycolysis model was developed by Goldbeter and Lefever [15]. 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:  $\text{fructose - 6 - phosphate} + \text{ATP} \rightleftharpoons \text{ADP} + \text{fructose - 1, 6 - bisphosphate}$ . In figure 7.4 the network diagram of this model is shown. Here  $\gamma$  indicates ADP and  $\alpha$ , fructose-6-phosphate. PFK is an allosteric enzyme, modelled with a MWC mechanism,

$$v_2 = \sigma_M \frac{\alpha e (1 + \alpha e)^{n-1} (1 + \gamma)^n + L \theta \alpha e' (1 + \alpha e')^{n-1}}{L (1 + \alpha e')^n + (1 + \gamma)^n (1 + \alpha e)^n} \quad (7.16)$$

Where  $e = (1 + \epsilon)^{-1}$  and  $e' = (1 + \epsilon')^{-1}$  with  $\epsilon$  and  $\epsilon'$  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 7.5 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.

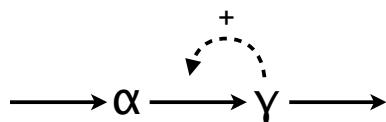


Figure 7.4: Network diagram for the simplified model developed by Goldbeter and Lefever [15] to study the role of the product activation of PFK by ADP in glycolysis. In this model, PFK was modelled according to a MWC mechanism.

### 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 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 [18]. 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 7.6). Each of the subunits catalyzes the reversible uni-uni reaction from  $S$  to  $P$ . The total enzyme concentrations obeys,

$$e_T = e + es_2 + 2esp + ep_2 \quad (7.17)$$

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

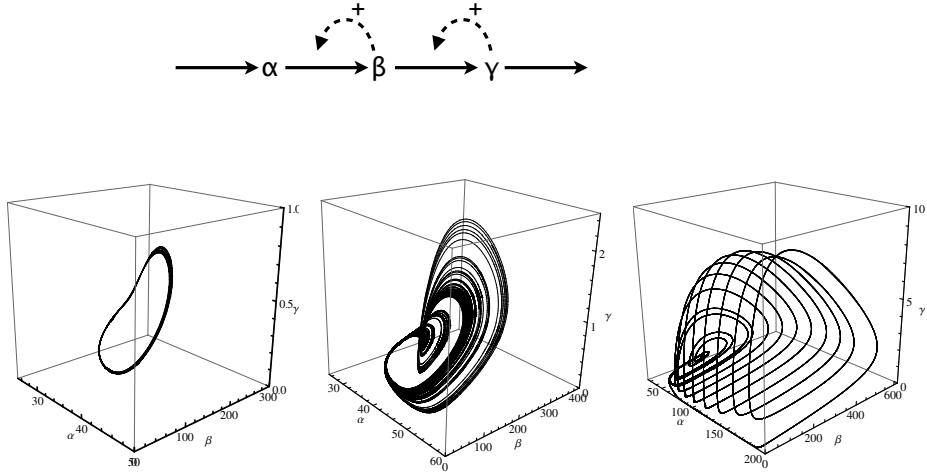


Figure 7.5: Network diagram and dynamics of an illustrative model for complex dynamics as developed by Decroly and Goldbeter [9]. The three plots with dynamics differ in the value of a first-order rate constant  $k_s$  for the degradation of  $\gamma$ . The plot of the left (for  $k_s = 1.9$  indicates 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:  $v_1 = v/K_{m1}$ ,  $v_2 = \frac{\alpha(1+\alpha)(1+\beta)^2}{L_1+(1+\alpha)^2(1+\beta)^2}$ ,  $v_3 = \frac{\beta(1+d\beta)(1+\gamma)^2}{L_1+(1+d\beta)^2(1+\gamma)^2}$ , and  $v_4 = k_s\gamma$ . The differential equations are:  $d\alpha/dt = v_1 - \sigma_1\theta$ ,  $d\beta/dt = q_1\sigma_1\theta - \sigma_2\eta$  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!

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,

$$v = k^+(2es_2 + 2esp) - k^-(2ep_2 + 2esp) \quad (7.18)$$

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

$$\begin{aligned} es_2 &= \frac{e \cdot s^2}{s_{0.5}^2} \\ esp &= \frac{e \cdot s \cdot p}{s_{0.5} p_{0.5}} \\ es_2 &= \frac{e \cdot p^2}{p_{0.5}^2} \end{aligned} \quad (7.19)$$

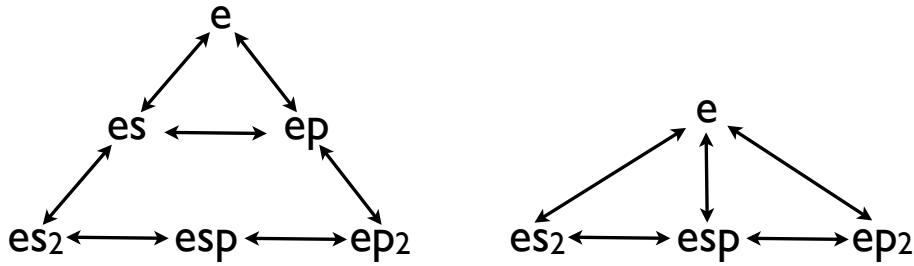


Figure 7.6: Illustration of the assumption in the derivation of the reversible Hill equation for an enzyme with two subunits each catalyzing the reversible conversion of  $S$  into  $P$ . On the left, the complete mechanism is shown and on the right the reduced mechanism as assumed for the reversible 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.

$p_{0.5}$  and  $s_{0.5}$  are at this point defined phenomenologically. The term  $s_{0.5}^2$  has to equal  $\alpha K_s^2$  with  $\alpha$  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  $\alpha$  needs to be much smaller than 1. The free enzyme concentration can now be equated,

$$e = \frac{e_T}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5}p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \quad (7.20)$$

and the other species as well, for instance,

$$es_2 = \frac{e_T}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5}p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \frac{s^2}{s_{0.5}^2} \quad (7.21)$$

The rate of the enzyme now becomes,

$$\begin{aligned}
 v &= \frac{2k^+e_T \left( \frac{s^2}{s_{0.5}^2} + \frac{s \cdot p}{s_{0.5} p_{0.5}} \right) - 2k^-e_T \left( \frac{p^2}{p_{0.5}^2} + \frac{s \cdot p}{s_{0.5} p_{0.5}} \right)}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \\
 &= \frac{2k^+e_T \frac{s}{s_{0.5}} \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right) - 2k^-e_T \frac{p}{p_{0.5}} \left( \frac{p}{p_{0.5}} + \frac{s}{s_{0.5}} \right)}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \\
 &= \frac{\left( V_{MAX}^+ \frac{s}{s_{0.5}} - V_{MAX}^- \frac{p}{p_{0.5}} \right) \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)}{1 + \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^2} \\
 &= \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left( 1 - \frac{p}{sK_{EQ}} \right) \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)}{1 + \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^2} \tag{7.22}
 \end{aligned}$$

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 7.22 to enzymes with  $n$  subunits,

$$v = \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left( 1 - \frac{p}{sK_{EQ}} \right) \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^{n-1}}{1 + \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^n} \tag{7.23}$$

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,

$$v = \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left( 1 - \frac{p}{sK_{EQ}} \right) \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^{n-1}}{\frac{1 + \left( \frac{x}{x_{0.5}} \right)^n}{1 + \beta \left( \frac{x}{x_{0.5}} \right)^n} + \left( \frac{s}{s_{0.5}} + \frac{p}{p_{0.5}} \right)^n} \tag{7.24}$$

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

## 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

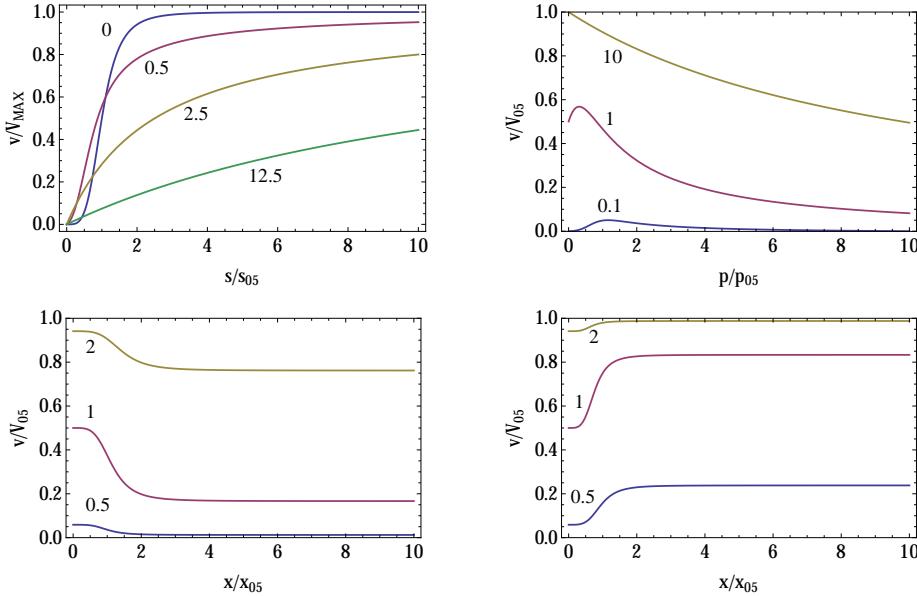


Figure 7.7: Numerical analysis of the reversible Hill rate equation (equation 7.24). In the upper left plot, product inhibition is illustrated ( $p$  was varied). The upper right plot displays product inhibition for various values of the substrate,  $s$ . The lower two plots show the effect of the  $\beta$  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.

$x_3$ . 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,  $p$ ,

$$v = \frac{\frac{V_f s}{s_{0.5}} \left(1 - \frac{s}{x_1 K_{eq}}\right) \left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^{n-1}}{\left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^n + \frac{1 + \left(\frac{p}{p_{0.5}}\right)^n}{1 + \alpha \left(\frac{p}{p_{0.5}}\right)^n}} \quad (7.25)$$

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 [19, 20].

# Chapter 8

## Examples of cooperative proteins

### 8.1 Cooperative sensors: membrane receptors and transcription factors

Receptor and transcription factors are often homodimers that switch into an active conformation upon activation. The protein then has a different 3D structure and altered affinity for other proteins and/or other kinetic activities, e.g. it becomes a kinase. For membrane receptors, activation typically involves binding of an extracellular signal. Transcription factors can either become activated in this manner or via covalent modification, e.g. by phosphorylation. The latter mechanism is likely more relevant for transcription factors. Hence, this section applies more to receptors than to transcription factors. One class of transcription factor this section is relevant for are the so-called nuclear receptors, which are found in mammals and bind small molecules, or for transcription factors that bind metabolites in other organisms.

We will consider in this section a dimeric protein. Each monomer can exist in two conformations; the relaxed, active 'R' state and the tensed, inactive 'T' state. The conformation changes occur in a concerted manner. Thus, if one monomer changes conformation, the other monomer undergoes simultaneously the same conformation change. The dissociation constant of the signal,  $S$ , for the dimeric sensor differs between the 'R' and the 'T' state and are denoted respectively by  $K_R$  and  $K_T$ . The total concentration of dimers we denote by  $e_{2T}$ , which equals,

$$e_{2T} = r_2 + 2r_2s + r_2s_2 + t_2 + 2t_2s + t_2s_2 \quad (8.1)$$

The factor '2' signifies that two forms of a single sensor state occurs, i.e. the signal can bound to either the left or the right monomer.<sup>1</sup> We can rewrite this

---

<sup>1</sup> Alternatively we could have written:  $e_{2T} = r_2 + r_2s + sr_2 + r_2s_2 + t_2 + st_2 + t_2s + t_2s_2$  to

equation using the definitions of the dissociation coefficients,

$$\begin{aligned}
 e_{2T} &= r_2 + 2\frac{s \cdot r_2}{K_R} + \frac{s^2 r_2}{K_R^2} + t_2 + 2\frac{s \cdot t_2}{K_T} + \frac{s^2 t_2}{K_T^2} \\
 &= r_2 \left( \left( 1 + 2\frac{s}{K_R} + \frac{s^2}{K_R^2} \right) + \frac{t_2}{r_2} \left( 1 + 2\frac{s}{K_T} + \frac{s^2}{K_T^2} \right) \right) \\
 &= r_2 \left( \left( 1 + 2\frac{s}{K_R} + \frac{s^2}{K_R^2} \right) + L \left( 1 + 2\frac{s}{K_T} + \frac{s^2}{K_T^2} \right) \right) \\
 &= r_2 \left( \left( 1 + \frac{s}{K_R} \right)^2 + L \left( 1 + \frac{s}{K_T} \right)^2 \right)
 \end{aligned} \tag{8.2}$$

Here we have defined  $L = \frac{t_2}{r_2}$  as the equilibrium constant for the conformational change reaction  $r_2 \rightleftharpoons t_2$ . This equation expresses the concentration of the empty active receptor state in terms of the parameters of the problem, i.e.  $L$ ,  $K_R$ ,  $K_T$ , and  $e_{2T}$ . We will use this relationship below.

### 8.1.1 Binding of the signal induces a conformational change when $K_R \neq K_T$

The total concentration of dimers in the 'R' state is given by  $r_T = r_2 + 2r_2s + r_2s_2$  and the total concentration of the dimers in the 'T' state by  $t_T = t_2 + 2t_2s + t_2s_2$ . Hence, the ratio of the 'T' over the 'R' state equals (after having followed the logic of the previous section),

$$\frac{t_T}{r_T} = L \frac{\left(1 + \frac{s}{K_T}\right)^2}{\left(1 + \frac{s}{K_R}\right)^2} = L \frac{(1 + c \cdot \alpha)^2}{(1 + \alpha)^2} \tag{8.3}$$

Here we have defined  $\alpha = \frac{s}{K_T}$ , and  $c = \frac{K_R}{K_T}$ . Figure 8.1 indicates if  $c < 1$  a conformational change of  $T \rightarrow R$  occurs and when  $c > 1$  'R' switches to 'T'. Hence, signal binding induces a conformational change! (When  $c = 1$  no conformation change is induced but then 'R' and 'T' are indistinguishable, which is not relevant.)

### 8.1.2 The fraction of bound subunits (the MWC perspective)

Monod, Wyman and Changeux (MWC) derived a model for cooperative enzymes that agrees with the previous section and they analysed the fraction of

---

make this explicit. Since we are considering a homodimer, both monomers will have the same affinity for  $S$  and distinguishing the two sensor states with one signal bound is not required.

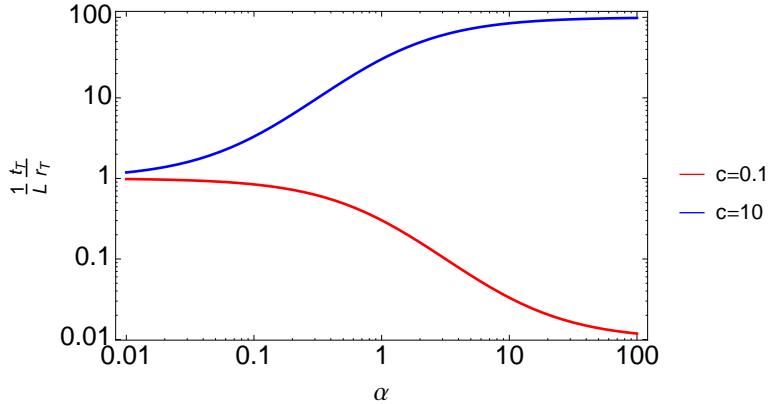


Figure 8.1: **Induction of conformational change by binding of the signal.** Relationship 8.3 is shown in a log-log plot for two values of the ratio of  $K_R/K_T$ . Depending on this ratio binding of either induces a conformation change from  $R \rightarrow T$  or vice versa. Note that the coefficient  $L$  only scales the lines up or down (it only affects the y-axis).

bound subunits,  $Y$ ,

$$\begin{aligned} Y &= \frac{\text{total concentration of bound subunits}}{\text{total concentration of subunits}} \\ &= \frac{2 \cdot 1 \cdot sr_2 + 1 \cdot 2 \cdot s_2r_2 + 2 \cdot 1 \cdot st_2 + 1 \cdot 2 \cdot s_2t_2}{2 \cdot e_{2T}} \end{aligned} \quad (8.4)$$

The logic behind this equation is the following: every sensor state that is signal bound, i.e.  $sr_2$ ,  $s_2r_2$ ,  $st_2$ , and  $s_2t_2$ , is considered and to calculate the concentration of bound subunits we need to take into consideration the number of forms of this state (e.g. two forms for the single bound states, e.g.  $sr_2$  and  $r_{rs}$ , and one forms of the doubly bound states). In the numerator of the previous equation every concentration is multiplied with two numbers: the first denotes the number of forms of the state and the second denotes the number of bound subunits in this state. In the denominator of this equation the total concentration of the subunits is given, which equals twice the concentration of the monomers. With the information from the previous subsection we can simplify

this equation further (note that the 2's drop out in the previous equation),

$$\begin{aligned}
 Y &= \frac{sr_2 + s_2r_2 + st_2 + s_2t_2}{e_{2T}} \\
 &= \frac{r_2 \left( \left( \frac{s}{K_R} + \frac{s^2}{K_R^2} \right) + L \left( \frac{s}{K_T} + \frac{s^2}{K_T^2} \right) \right)}{e_{2T}} \\
 &= \frac{r_2 \left( \frac{s}{K_R} \left( 1 + \frac{s}{K_R} \right) + L \frac{s}{K_T} \left( 1 + \frac{s}{K_T} \right) \right)}{e_{2T}} \\
 &= \frac{r_2 \left( \frac{s}{K_R} \left( 1 + \frac{s}{K_R} \right) + L \frac{s}{K_T} \left( 1 + \frac{s}{K_T} \right) \right)}{r_2 \left( \left( 1 + \frac{s}{K_S} \right)^2 + L \left( 1 + \frac{s}{K_T} \right)^2 \right)} \\
 &= \frac{\frac{s}{K_R} \left( 1 + \frac{s}{K_R} \right) + L \frac{s}{K_T} \left( 1 + \frac{s}{K_T} \right)}{\left( 1 + \frac{s}{K_R} \right)^2 + L \left( 1 + \frac{s}{K_T} \right)^2} \tag{8.5}
 \end{aligned}$$

The last equation is the MWC equation. This equation has a specific application: if the bound subunits have a specific activity, say an enzyme activity, then the MWC approach is very useful as  $Y$  multiplied by some catalytic rate constant  $k_{cat}$  and the total subunit concentration gives the total enzyme activity as function of  $s$ , which could be the substrate concentration or a regulator of the enzyme. This is the application that Monod, Wyman, and Changeux had in mind. For purposes of studying receptor or transcription factor activity, we are interested in another application of this theory. We would like to know the concentration of specific states rather than the concentration of all the bound subunits; thus, we require a dimer and state specific perspective rather than a bound subunit perspective. This is what we shall work out next.

### 8.1.3 A dimer and state perspective

Let's say that only the fully-bound relaxed state,  $R_2S_2$ , has biological activity, e.g. it activates downstream signalling proteins when we consider a membrane receptor or can bind to DNA in the case of a transcription factor. In this case, we would like to know the concentration of the  $R_2S_2$  state, which is related to the concentration of the  $R_2$  state by the following equilibrium binding relation,

$$r_2s_2 = \frac{s^2r_2}{K_R^2} \tag{8.6}$$

The concentration  $r_2$  can be expressed in terms of parameters with equation 8.2; hence, we obtain,

$$r_2s_2 = e_{2T} \frac{\frac{s^2}{K_R^2}}{\left( 1 + \frac{s}{K_R} \right)^2 + L \left( 1 + \frac{s}{K_T} \right)^2} \tag{8.7}$$

Thus, the fraction of active membrane receptors or transcription factors equals  $\frac{r_2 s_2}{e_{2T}}$  and this concentration depends on the concentration of  $S$  and affinity and conformation equilibrium of properties of the sensor.

### 8.1.4 Consideration of additional regulators

The relaxed state is the active state of the sensor and the tensed state is considered less active. Activators of the sensor will then stabilise the relaxed state whereas inhibitors will stabilise the tensed state. Hence,  $L$  will decrease due to activators and increase due to inhibitors. If the binding of the signal to the T and R state of the sensor is completely independent of the binding state of the sensor with respect to the activator and the inhibitor, the sum of all the R and T states can be obtained from,

$$\begin{aligned} r_{2T} &= r_2 \left(1 + \frac{a}{K_A}\right)^2 \left(1 + \frac{s}{K_R}\right)^2 \\ t_{2T} &= t_2 \left(1 + \frac{i}{K_I}\right)^2 \left(1 + \frac{s}{K_T}\right)^2 \end{aligned} \quad (8.8)$$

And the total dimer concentration equals,

$$\begin{aligned} e_{2T} &= r_{2T} + t_{2T} \\ &= r_2 \left(1 + \frac{a}{K_A}\right)^2 \left(1 + \frac{s}{K_R}\right)^2 + t_2 \left(1 + \frac{i}{K_I}\right)^2 \left(1 + \frac{s}{K_T}\right)^2 \\ &= r_2 \left(1 + \frac{a}{K_A}\right)^2 \left(1 + \frac{s}{K_R}\right)^2 + \underbrace{\frac{t_2 \left(1 + \frac{i}{K_I}\right)^2}{r_2 \left(1 + \frac{a}{K_A}\right)^2} \left(1 + \frac{s}{K_T}\right)^2}_{L'} \end{aligned} \quad (8.9)$$

Here we have defined the apparent  $L$  coefficient as,

$$L' = \frac{t_2 \left(1 + \frac{i}{K_I}\right)^2}{r_2 \left(1 + \frac{a}{K_A}\right)^2} = L \frac{\left(1 + \frac{i}{K_I}\right)^2}{\left(1 + \frac{a}{K_A}\right)^2} \quad (8.10)$$

With this definition, the saturation function  $Y$  becomes,

$$Y = \frac{\frac{s}{K_R} \left(1 + \frac{s}{K_R}\right) + L' \frac{s}{K_T} \left(1 + \frac{s}{K_T}\right)}{\left(1 + \frac{s}{K_R}\right)^2 + L' \left(1 + \frac{s}{K_T}\right)^2}, \quad (8.11)$$

and the active dimer state becomes,

$$r_2 s_2 = e_{2T} \frac{\frac{s^2}{K_R^2}}{\left(1 + \frac{s}{K_R}\right)^2 + L' \left(1 + \frac{s}{K_T}\right)^2} \quad (8.12)$$

This section showed how to incorporate activators and inhibitors into the MWC equations for a GPCR or TF.

### 8.1.5 Assessing the regulatory potential of $R_2S_2$

#### How to choose the affinity of the downstream process for $r_2s_2$ ?

The maximal active fraction, (i.e. when  $s \rightarrow \infty$  in equation 8.7) equals  $\left(1 + \left(\frac{K_R}{K_T}\right)^2 L\right)^{-1}$  and, therefore, only approaches 1 when  $K_T \gg K_R$  or  $L \approx 0$ . In all cases, the active sensor concentration lies within the following bounds,

$$0 \leq r_2s_2 \leq \underbrace{\frac{e_{2T}}{1 + \left(\frac{K_R}{K_T}\right)^2 L}}_{r_2s_{2max}} \quad (8.13)$$

And a downstream process regulated by  $R_2S_2$  should therefore have a half maximal response around  $\frac{1}{10} \frac{e_{2T}}{1 + \left(\frac{K_R}{K_T}\right)^2 L}$  because then the downstream process can be regulated properly over the entire concentration range of  $r_2s_2$ . To understand why, let's consider that the downstream process produces  $X$  as function of the concentration of  $R_2S_2$ ; say we have the following dependency of the steady state concentration of  $X$  on  $r_2s_2$ :  $x = x(r_2s_2)$  and let's take this dependency as, with  $x^*$  as the maximal level of  $X$ ,

$$x = x^* \frac{r_2s_2}{K_{1/2} + r_2s_2} \quad (8.14)$$

If  $r_2s_2$  ranges from  $0.1K_{1/2}$  to  $10K_{1/2}$ , the corresponding range in the output is a factor of 9 in concentration value,

$$\frac{x(10K_{1/2}) - x(0.1K_{1/2})}{x(0.1K_{1/2})} = 9 \quad (8.15)$$

So, if  $K_{1/2}$  is chosen as  $1/10r_2s_{2max}$  the downstream process can be controlled in a flexible manner.

#### How sensitive will the downstream process be for $r_2s_2$ and $s$ ?

The sensitivity of  $x$  for  $r_2s_2$  will typically be quantified from the slope of the dependency of  $x$  on  $r_2s_2$  in the following manner,

$$\frac{r_2s_2}{x} \frac{\partial x}{\partial r_2s_2} = \frac{\partial \ln x}{\partial \ln r_2s_2} \quad (8.16)$$

The sensitive of  $x$  to  $s$  then derives from  $x = x(r_2s_2(s))$  and we obtain for the sensitivity measure,

$$\frac{\partial \ln x}{\partial \ln r_2s_2} \frac{\partial \ln r_2s_2}{\partial \ln s}, \quad (8.17)$$

which equals the product of the sensitivity of  $x$  to  $r_2s_2$  and of  $r_2s_2$  to  $s$ . Note that  $r_2s_2 = r_2s_2(s)$  is given by equation 8.7.

### 8.1.6 Exercise

1. What is the advantage of a dimeric sensor over a monomeric sensor?
2. What does the exponent of '2' signify in equation 8.2?
3. Derive the Hill equation from equation 8.7.



# Chapter 9

## Stoichiometric network analysis

### 9.1 Introduction

Kinetic models require information about the initial conditions, i.e. the concentrations of the intermediate metabolites, the environment, reaction kinetics, thermodynamics, and stoichiometry. They allow for the calculation of the dynamics of the concentrations of molecules in reaction networks and their dependencies on parameters. Essential information for many biological questions, but in many cases not all of that information is available. Are there alternative modeling approaches that do not require such detailed kinetic information and can still address relevant biological questions? For metabolic networks, as we will see, a number of system properties can be found with stoichiometric network analysis (SNA).

Stoichiometric models solely consider the stoichiometry of the reactions occurring in a metabolic network. Stoichiometry can be a powerful constraint for the steady-state behavior of metabolic networks. 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.

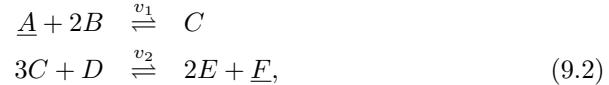
In this chapter, we will introduce some of the basics of SNA: conservation relationships, independent fluxes, flux modes, and flux space. Flux balance analysis will be explained in the next chapter.

## 9.2 The stoichiometric matrix, $\mathbf{N}$

In the previous chapters, we studied kinetic models of molecular networks. The mass balances for the variable concentrations always had the following structure,

$$\frac{d}{dt}x_i(t) = \sum_{j=1}^r n_{ij}v_j. \quad (9.1)$$

Such an equation was then determined for every variable metabolite  $X_i$  in the network. To refresh your memory: The coefficients  $n_{ij}$  are stoichiometry coefficients and denote the number of molecules of  $X_i$  that are produced (when  $n_{ij} > 0$ ), consumed (when  $n_{ij} < 0$ ) or not involved ( $n_{ij} = 0$ ) in the  $j$ -th reaction with rate  $v_j$  and  $r$  reactions occur in total. We have started studying such systems already in chapter 2. For the reaction system,



we would obtain as mass balances,

$$\begin{aligned} \frac{d}{dt}b &= -2v_1 \\ \frac{d}{dt}c &= v_1 - 3v_2 \\ \frac{d}{dt}d &= -v_2 \\ \frac{d}{dt}e &= 2v_2 \end{aligned} \quad (9.3)$$

Thus, the factor  $-3$  is the stoichiometry coefficient of metabolite  $C$  in reaction 2. We can also write this same set of mass balances in terms of a matrix and vector product,

$$\underbrace{\begin{pmatrix} \frac{d}{dt}b \\ \frac{d}{dt}c \\ \frac{d}{dt}d \\ \frac{d}{dt}e \end{pmatrix}}_{\text{rates of change}} = \underbrace{\begin{pmatrix} -2 & 0 \\ 1 & -3 \\ 0 & -1 \\ 0 & 2 \end{pmatrix}}_{\text{Stoichiometry matrix}} \cdot \underbrace{\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}}_{\text{rate vector}} \quad (9.4)$$

This is an example of the general case with  $r$  reactions and  $m$  intermediate metabolites, which can be compactly written in terms of the following differential equation,

$$\dot{\mathbf{x}}(\mathbf{t}) = \frac{d}{dt}\mathbf{x}(t) = \mathbf{N} \cdot \mathbf{v}(\mathbf{x}(t)) \quad (9.5)$$

with:  $\mathbf{x}$  as the  $m \times 1$  concentration (or state) vector,  $\mathbf{N}$  as the  $m \times r$  stoichiometric matrix, and  $\mathbf{v}$  as the  $r \times 1$  rate vector that contains the rate equations for  $r$

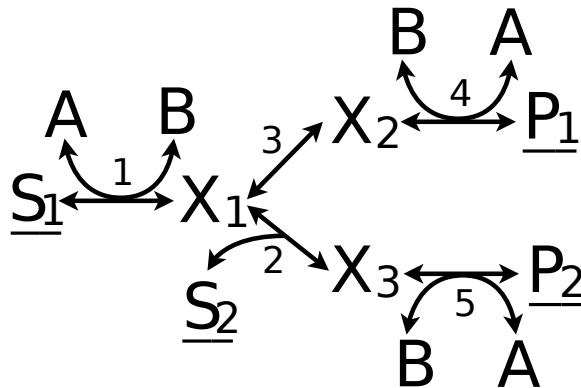


Figure 9.1: A branched metabolic pathway with five enzymes, five intermediate metabolites, and four fixed external metabolites. The intermediate metabolites  $A$  and  $B$  are being recycled.

reactions that occur in the network, which depend on the concentrations of the molecules. The “.” above  $\mathbf{x}$  means a time derivative, i.e.  $\frac{d\mathbf{x}}{dt} = \dot{\mathbf{x}}$ . We will sometime use this notational convention in this chapter.

For example, consider the network shown in Figure 9.1. The stoichiometric matrix is given by (we have indicated the names of the columns and rows here, this is in the definition of the  $\mathbf{N}$  matrix of course),

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ X_1 & 1 & -1 & -1 & 0 & 0 \\ X_2 & 0 & 0 & 1 & -1 & 0 \\ A & -1 & 0 & 0 & 1 & 1 \\ B & 1 & 0 & 0 & -1 & -1 \\ X_3 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \quad (9.6)$$

This stoichiometry matrix provides information about the number of molecules produced and consumed per unit rate. The stoichiometry coefficient were introduced in earlier chapters. So, one molecule of  $X_1$  is consumed in reaction 2, which yield one molecule of  $X_3$ . Note that the constant molecules are not denoted in the stoichiometry matrix. Multiplication of the stoichiometric matrix with the rate vector gives all the differential equation for the variable metabolite concentrations. In this chapter we will study a number of properties of the stoichiometry matrix that derive from linear algebra. In the next section we will explain the associated concepts without resorting to linear algebra yet.

## Exercises

Determine the stoichiometric matrices for the reaction systems described in the exercises of section 2.2.2.

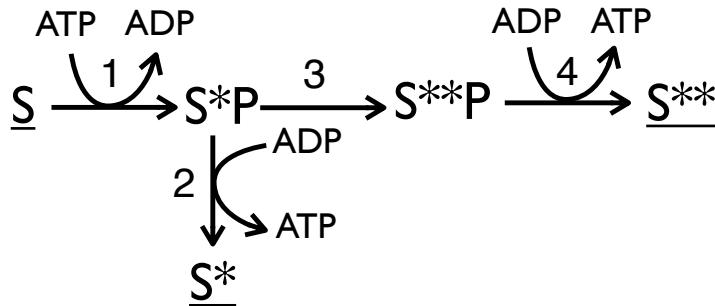


Figure 9.2: A branched metabolic pathway with four enzymes and four variable metabolites. How many reaction rates do you need to know at steady state to determine all reaction rates? Can you identify sums of concentrations in this pathway that remain fixed? Answering those questions is what this chapter is about. Note that the products of the pathway can only by isomers of  $S$  as no atoms or molecules have been added to or removed from  $S$ .

### 9.3 The principles of this chapter explained with a simple example

This chapter deals with two basic consequences of the stoichiometry of biochemical reactions inside networks; one has to do with the conserved moieties in the network (we encountered those already in chapter 2) and the other with the independent fluxes in a steady state of the network – i.e. the number of fluxes that need to be known in order to determine all fluxes at steady state. Note that in the stoichiometric world we will mostly talk about steady-states, where many rates of enzymes, e.g. along linear paths, will be equal. This rate is called the steady-state flux. As we will see later, we denote such flux with the symbol  $J$  to emphasize the steady-state nature of the rate (e.g.  $J_1$  instead of  $v_1$  denoted the rate of enzyme 1 in steady state).

Moiety conservation is important as it couples the dynamic behaviour of variables. In Figure 9.2, you see that every time ATP is consumed, ADP is produced, and vice versa. This is because only the phosphate is transferred onto the  $S$  molecule, and the adenine – the “ $A$ ” in  $ATP$  and  $ADP$  – is conserved. Hence, if ATP would decrease in time, ADP would increase, basically following as its mirror-image. This is clear from their ODE’s, as  $\frac{d\text{atp}}{dt} = -\frac{d\text{adp}}{dt}$ :

$$\begin{aligned}\frac{d\text{atp}}{dt} &= -v_1 + v_2 + v_4 \\ \frac{d\text{adp}}{dt} &= v_1 - v_2 - v_4,\end{aligned}\tag{9.7}$$

Because of this equivalence, we can choose either one of these metabolites as the independent metabolite, and the other one as the dependent metabolite;

as a consequence, we can calculate the dynamics of the dependent metabolite from the dynamics of the independent metabolites, because we know the rate of change of the dependent metabolite in terms of the rate of change of the independent metabolite.

When looking more closely at Figure 9.2, you should notice that besides "A" also phosphate "P" is not flowing into or out of the network (check whether you agree). So, these two *mieties* cannot change in total concentration, which must mean that  $atp + adp$  and  $atp + s^*p + s^{**}p$  remain constant; this is indeed correct because,

$$\begin{aligned}\frac{datp}{dt} + \frac{dadp}{dt} &= 0 \Rightarrow \underbrace{-v_1 + v_2 + v_4}_{datp/dt} + \underbrace{v_1 - v_4 - v_2}_{dadp/dt} = 0 \\ \frac{datp}{dt} + \frac{ds^*p}{dt} + \frac{ds^{**}p}{dt} &= 0 \Rightarrow \underbrace{-v_1 + v_2 + v_4}_{datp/dt} + \underbrace{v_1 - v_2 - v_3}_{ds^*p/dt} + \underbrace{v_3 - v_4}_{ds^{**}p/dt} = 0\end{aligned}$$

Thus, out of the four variable concentrations, we need to know two rates of change in order to determine all the rates of change. Therefore, if we know  $datp/dt$  then we know  $dadp/dt$  and if we know in addition  $ds^*p/dt$  we can also determine  $ds^{**}p/dt$ . Hence,  $dadp/dt$  and  $ds^{**}p/dt$  can be omitted from the network description because they are redundant. Thus, we have only to determine:

$$\begin{aligned}\frac{datp}{dt} &= -v_1 + v_2 + v_4 \\ \frac{ds^*p}{dt} &= v_1 - v_2 - v_3,\end{aligned}\tag{9.8}$$

at any time point to determine the entire dynamics of all the four variable concentrations by using the relations between the rates of change introduced above. Importantly, especially for larger networks, those relations can be written in matrix form as:

$$\begin{pmatrix} \frac{dATP}{dt} \\ \frac{dS^*P}{dt} \\ \frac{dADP}{dt} \\ \frac{dS^{**}P}{dt} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ -1 & -1 \end{pmatrix}}_{\mathbf{L} \text{ matrix}} \cdot \begin{pmatrix} \frac{datp}{dt} \\ \frac{ds^*p}{dt} \end{pmatrix}\tag{9.9}$$

Check that this is correct, i.e. that indeed, for the third row:  $adp = -atp$ . In this chapter, you will learn how to determine the **L** matrix using a linear algebra trick, which works for any network regardless of its complexity.

Apart from relations between dynamics of metabolites, there are relationships between steady-state fluxes. The equations 9.8 equal zero at steady state and represent two relationships between the reaction fluxes at steady state. We have four reaction fluxes in total. This means that we need to know two flux values, say  $J_1$  and  $J_2$ , in order to determine all four flux values at steady

state, because from the following relations (equations 9.8) we can determine the remaining two fluxes;  $J_3$  and  $J_4$ ,

$$\begin{aligned}\frac{datp}{dt} = 0 &\Rightarrow J_4 = J_1 - J_2 \\ \frac{ds^*p}{dt} = 0 &\Rightarrow J_3 = J_1 - J_2\end{aligned}$$

Thus, the number of independent variables equals the number of (linearly-independent) relationships that exist between fluxes at steady state. Thus, the number of fluxes, 4, minus the independent variables (2) equals also the number of independent fluxes (2) that need to be determined in order to determine all fluxes.

Returning to the problem at hand: In matrix form we obtain the following relation,

$$\begin{pmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{K} \text{ matrix}} \cdot \begin{pmatrix} J_1 \\ J_2 \end{pmatrix} \quad (9.10)$$

This matrix equation captures all the steady state relations between fluxes that exist in the steady state.

This section showed that we first have to determine the **L** matrix and identify the identity and the number of the independent variables then the **K** can be determined. This is a rule because the number of rates minus the number of independent variables (determined from **L**) sets the number of independent fluxes – this is the minimal number of fluxes you have to measure in order to determine all fluxes. The independent fluxes are then identified with the **K** matrix. The determination of the **L** matrix is necessary for many numeric analyses of kinetic models of molecular systems, such as stability analysis, parameter estimation and control theory. This matrix will therefore return in later chapters. The determination of the **K** matrix is required for the analysis of flux distributions through metabolic networks given experimental data or with computational techniques.

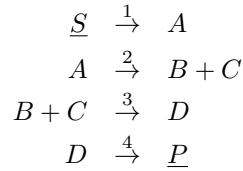
### Exercise

In these exercises the relationship between the number of reactions ( $r$ ), independent metabolites ( $m^0$ ), and the independent fluxes ( $r^0$ ) is useful, it states that:  $r - m^0 = r^0$ . This is a logical consequence of the fact that  $m^0$  independent linear relations exist between the fluxes, thus  $r - m^0$  need to be known to determine them all.

1. To see whether you have understood the previous section, determine the **L** and the **K** matrix for the following "strange" metabolic network composed

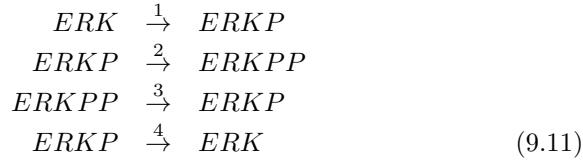
#### 9.4. DETERMINING $\mathbf{L}$ BY INSPECTION BY TRACING THE CUTTING AND PASTING OF MOLECULE PARTS

out the reactions,



$S$  and  $P$  are underlined to denote that they are fixed.

2. Determine the  $\mathbf{L}$  and the  $\mathbf{K}$  matrix for the following network involved in signal transduction,



ERK is a signalling protein in the so-called MAPK pathway of mammalian cells that has kinase activity when it is doubly phosphorylated. In this state, it translocated to the nucleus to activate nuclear transcription factors.  $ATP$ ,  $ADP$ , and  $Pi$  were considered fixed here and not shown.

3. How many independent flux occur in the two-component signalling network shown in Figure 3.8? Given this knowledge can there be more conserved moieties in addition to the total sensor and response regulator concentration?

## 9.4 Determining $\mathbf{L}$ by inspection by tracing the cutting and pasting of molecule parts by enzymes

Within molecular networks, molecules are being cut into pieces and then those pieces are pasted onto other molecules, this is all done 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 and its second nitrogen from alanine. Such parts of molecules are often 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. Their total concentrations then remain fixed in metabolic networks.

Consider the metabolic pathway depicted in Figure 9.3. In the first reaction some part of  $A$  is glued to  $S$  to give rise to the remainder of  $A$ , called  $B$ , and the newly synthesized molecule  $X$  that contains a part of  $A$  and  $S$ . 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 when we rename the species in the metabolic pathway, see Figure 9.4. Now it is a lot more clear what happens: a molecules are cut and pasted by enzymes.

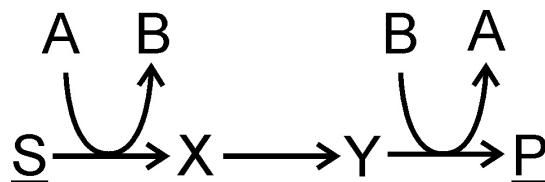


Figure 9.3: A metabolic pathway with three enzymes, four intermediate metabolites, and two fixed external metabolites ( $S, P$ ). The intermediate metabolites  $A$  and  $B$  are being recycled.

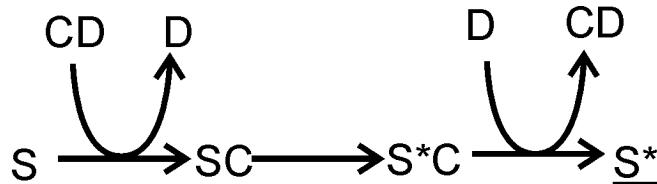


Figure 9.4: The metabolic pathway of Figure 9.3 with new names for the metabolites.

The mass balances for all the metabolites of the metabolic pathway shown in Figure 9.4 are given by,

$$\begin{pmatrix} \dot{cd} \\ \dot{sc} \\ \dot{d} \\ \dot{s^*c} \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}}_{N} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (9.12)$$

Figure 9.4 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  $c_T$  and  $d_T$  remain constant as function of time and are related to initial metabolite concentrations as they do not leave nor enter the system: the system is simply

#### 9.4. DETERMINING $\mathbf{L}$ BY INSPECTION BY TRACING THE CUTTING AND PASTING OF MOLECULE PART

stuck with the amounts it started with (denoted by the subscript 0),

$$\begin{aligned} c_T &= cd(t) + sc(t) + s^*c(t) = cd(0) + sc(0) + s^*c(0) \\ d_T &= d(t) + cd(t) = d(0) + cd(0) \end{aligned} \quad (9.13)$$

These relationships are called moiety-conservation relationships. These moiety-conservation relationships tell you that if  $d(0) = 5 \text{ mM}$  and  $cd(0) = 2.5 \text{ mM}$ ,  $d(t) + cd(t)$  at any time point will sum to  $7.5 \text{ mM}$ . This means that  $D$  and  $C$  are conserved. For this reason, the following sum of the rates of change are equal to zero (the totals are not dependent on time,  $t$ , so their derivative to  $t$  become zero),

$$\begin{aligned} 0 &= \dot{cd} + \dot{sc} + \dot{s^*c} \\ 0 &= \dot{d} + \dot{cd} \end{aligned} \quad (9.14)$$

These relationships indicate that there exist linear combinations between the rates of change, which defines the link matrix  $\mathbf{L}$  as a transformation matrix,

$$\begin{pmatrix} \dot{cd} \\ \dot{sc} \\ \dot{d} \\ \dot{s^*c} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ -1 & -1 \end{pmatrix}}_{\text{link matrix } \mathbf{L}} \begin{pmatrix} \dot{cd} \\ \dot{sc} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I} \\ \mathbf{L}_0 \end{pmatrix}}_{\text{link matrix } \mathbf{L}} \begin{pmatrix} \dot{cd} \\ \dot{sc} \end{pmatrix} \quad (9.15)$$

Here,  $\mathbf{I}$  describes the identity matrix, a matrix with ones on the diagonal and zeros elsewhere. The metabolites  $CD$  and  $SC$  are termed the independent metabolites and  $D$  and  $S^*C$  the dependent metabolites. These dependent metabolites are redundant for determining the dynamics, so the dynamics of all metabolites can be obtained from the independent metabolites as is shown in Equation (9.15).

Using the moiety-conservation relationships obtained in Equation (9.13), we can substitute for the concentrations of  $D$  and  $S^*C$  in all the rate equations,

$$\begin{aligned} d(t) &= d_T - cd(t) \\ s^*c(t) &= c_T - cd(t) - sc(t) \end{aligned} \quad (9.16)$$

The dynamics of the independent metabolites are now given by,

$$\begin{pmatrix} \dot{cd}(t) \\ \dot{sc}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}}_{\mathbf{N}_R} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (9.17)$$

Note that from  $\mathbf{N}_R$  all reference to the concentrations of the dependent metabolites have been eliminated (their rows are simply deleted).  $\mathbf{N}$  has been decomposed as (see equation 9.12),

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{pmatrix} \quad (9.18)$$

with  $\mathbf{N}_0$  as,

$$\mathbf{N}_0 = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \quad (9.19)$$

and  $\mathbf{N}_R$  as defined above.

Using the above described method involving renaming of the metabolites will allow you to derive the moiety-conservation relationships by hand. This can easily be done for systems upto the size of glycolysis. For anything bigger, linear algebra is more useful.

## 9.5 Determining the moiety conservation relations and steady-state flux relationships with Gaussian Elimination

Now that you have some intuition for the meaning of the  $\mathbf{K}$  and  $\mathbf{L}$  matrix and how to retrieve them, we will illustrate a method using linear algebra. We return again to the example of Figure 9.3. The stoichiometric matrix is given by (for clarity we show the identity of the rows and columns),

$$\mathbf{N} = \begin{pmatrix} v_1 & v_2 & v_3 \\ \dot{a} & -1 & 0 & 1 \\ \dot{x} & 1 & -1 & 0 \\ \dot{b} & 1 & 0 & -1 \\ \dot{y} & 0 & 1 & -1 \end{pmatrix} \quad (9.20)$$

In order to determine the  $\mathbf{L}$  matrix we are going to apply Gaussian elimination to this matrix using row operations (addition, subtraction, or exchange of rows). To keep track of the row operations we augment the stoichiometric matrix with an identity matrix. Due to the row operations the interpretation of the rows will change and therefore we no longer mention their identity,

$$\left( \begin{array}{ccccccc} v_1 & v_2 & v_3 & \dot{a} & \dot{x} & \dot{b} & \dot{y} \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 \end{array} \right) \quad (9.21)$$

Addition of row 1 to 3 gives for row 3,

$$\left( \begin{array}{ccccccc} v_1 & v_2 & v_3 & \dot{a} & \dot{x} & \dot{b} & \dot{y} \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 \end{array} \right) \quad (9.22)$$

## 9.5. DETERMINING THE MOIETY CONSERVATION RELATIONS AND STEADY-STATE FLUX RELATIONSHIPS

Addition of row 1, 2 and 4 gives for row 4,

$$\begin{pmatrix} v_1 & v_2 & v_3 & \dot{a} & \dot{x} & \dot{b} & \dot{y} \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \quad (9.23)$$

Addition of row 1 and 2 gives for row 2,

$$\begin{pmatrix} v_1 & v_2 & v_3 & \dot{a} & \dot{x} & \dot{b} & \dot{y} \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \quad (9.24)$$

Now we have finished the Gaussian elimination procedure because the diagonal of consists as far as possible of nonzero elements and, therefore, all the rows below the diagonal consist of zeros. Finding the moieties is now really straightforward as those are given by rows that contain zeros in all the column entries for the reaction rates, i.e.

$$\begin{aligned} \dot{a} + \dot{b} &= \dot{cd} + \dot{d} = 0 \\ \dot{a} + \dot{x} + \dot{y} &= \dot{cd} + \dot{sc} + s^* \dot{c} = 0 \end{aligned} \quad (9.25)$$

If  $A$  would be ATP and  $B$  would be ADP then the first conservation relation gives the conservation of total adenosine,  $A$ , and the other of phosphate,  $P$ . The total number of rows ( $=4$ ) minus the number of rows with zero entries in the reaction columns ( $=2$ ) gives the number of independent metabolites ( $=2$ ).

In addition, the rows with nonzero entries in the reaction columns, i.e. the rows that refer to the independent concentrations, give rise to the flux relationships in steady state,

$$\begin{aligned} -J_1 + J_3 &= 0 \\ -J_2 + J_3 &= 0 \end{aligned} \quad (9.26)$$

So, choosing  $J_3$  as independent flux makes sense. In fact, the general rule is that the reaction columns without pivots (numbers on the diagonal) identify the independent fluxes.

To go through the procedure once more, we consider a slightly more complicated problem taken from Cornish-Bowden *et al.* [8]. This problem has the following stoichiometric matrix (draw the network yourself or consult the pa-

per),

$$\mathbf{N} = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 \\ h\dot{6}p & 1 & -1 & 0 & 0 \\ \dot{tr}p & 0 & 1 & -1 & 0 \\ \dot{tr} & 0 & 1 & 0 & -1 \\ \dot{at}p & -1 & 0 & 1 & 0 \\ \dot{ad}p & 1 & 0 & -1 & 0 \\ \dot{am}p & 0 & 0 & 0 & -1 \end{pmatrix} \quad (9.27)$$

When we start the Gaussian elimination we start from

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & h\dot{6}p & \dot{tr}p & \dot{tr} & \dot{at}p & \dot{ad}p & \dot{am}p \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (9.28)$$

We have to remove the non-zero elements below the diagonal by using Gaussian elimination. We will sum row 1 and 4 to obtain for 4,

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & h\dot{6}p & \dot{tr}p & \dot{tr} & \dot{at}p & \dot{ad}p & \dot{am}p \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (9.29)$$

Then we swap for row 5, row 1-row 5, to eliminate the 1 in the first column of row 5,

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & h\dot{6}p & \dot{tr}p & \dot{tr} & \dot{at}p & \dot{ad}p & \dot{am}p \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (9.30)$$

By either subtracting or adding row 2 to row 3, 4 and 5, we remove all the

nonzero entries in columns 2 of row 3, 4 and 5,

$$\left( \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & h\dot{6}p & \dot{trp} & \dot{tr} & \dot{atp} & \dot{adp} & \dot{amp} \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \quad (9.31)$$

Next, twice the addition of row 4 to 5 eliminates the 2 in row 5 and subtracting row 4 from 6 gives for rows 5 and 6,

$$\left( \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & h\dot{6}p & \dot{trp} & \dot{tr} & \dot{atp} & \dot{adp} & \dot{amp} \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 1 \end{array} \right) \quad (9.32)$$

Now we are finally done! The moieties can now be derived from,

$$\begin{aligned} 0 &= h\dot{6}p + \dot{trp} + 2\dot{atp} + \dot{adp} \\ 0 &= -h\dot{6}p - \dot{trp} - \dot{atp} + \dot{amp} \end{aligned} \quad (9.33)$$

The first looks intelligible and concerns the conservation of phosphate:  $h\dot{6}p + \dot{trp} + 2 \cdot \dot{atp} + \dot{adp} = p_T$ . The second one is a bit more puzzling but when these two relations are added:  $\dot{adp} + \dot{atp} + \dot{amp} = 0$  appears, which indicates the conservation of adenosine:  $\dot{adp} + \dot{atp} + \dot{amp} = a_T$ .

We can also read the independent fluxes from the Gaussian elimination result by taking the rates in the nonpivoting columns as independent fluxes. This identifies  $J_4$  as independent. The Gaussian elimination result indicates that flux 1 to 3 are all equal to  $J_4$  in steady state. It also indicates  $J_5$  is zero in all steady states because all the nonzero rows in the segment of the Gaussian elimination result that refers to the reactions give the steady-state flux relationships, i.e.  $J_1 - J_2 = 0$ ,  $J_2 - J_3 = 0$ ,  $J_3 - J_4 = 0$ , and  $-J_5 = 0$ ! When a reaction flux is zero at steady state means it operates at thermodynamic equilibrium; hence reaction 5, i.e.  $\dot{atp} + \dot{amp} \rightleftharpoons 2\dot{adp}$ , operates at thermodynamic equilibrium. Therefore, the concentration ratio  $\frac{\dot{adp}^2}{\dot{atp} \cdot \dot{amp}}$  will equal the equilibrium constant of this reaction. This reaction is catalysed by the enzyme called adenylate kinase.

The linear algebra trick of this section can be done by hand, as was shown, but also by Mathematica with the command `MatrixForm[RowReduce[Join[N, IdentityMatrix[Dimensions[N][[1]]], 2]]]` with  $N$  as the stoichiometry matrix. This code is expelling in more detail below.

Often the problems will be too large and then Mathematica or another program will be used to solve it. Then, the Gaussian elimination trick as introduced

in this chapter is not the most useful; see the advanced material section at the end of this chapter.

## 9.6 Independent fluxes and the kernel matrix K

To practise a bit more with the independent fluxes at steady state, we will go through an explicit example. At steady-state all the fluxes through a linear pathway without moieties are equal.<sup>1</sup> Thus to determine all the fluxes for a linear pathway in steady-state, we only have to know the so-called independent flux. All the other fluxes are then equal to the value of this independent flux and are termed 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.

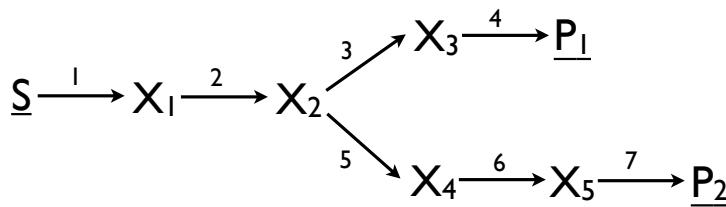


Figure 9.5: A branched metabolic pathway with five variable metabolites and seven fluxes.

In this section, we will consider networks without moiety-conservation relationships before we outline generic linear algebra methods for their determination. Let's consider the example network displayed in Figure 9.5. This network has no conserved moieties, seven fluxes, and five variable metabolites.

For those who understand linear algebra: The number of independent metabolites is given by the rank,  $m_0$ , of the stoichiometric matrix. For the rest of us, the rank of  $\mathbf{N}$  corresponds to the number of rows in  $\mathbf{N}_R$ : the number of independent metabolites in the network; the number of columns of the  $\mathbf{L}$  matrix. Thus, we have  $m_0$  relations between fluxes at steady state: i.e. the mass balances of the independent metabolites set to zero. Therefore, we have  $r - m_0$  fluxes left that we need to determine by some other means; those are the independent fluxes. Since, the network in Figure 9.5 has no conserved moieties and 5 mass balances, it has  $7-5=2$  independent fluxes. This makes perfect sense if you study the pathway in Figure 9.5 for a moment. Say you know  $J_1$  to be  $3 \text{ mM/min}$  then  $J_2$  carries the same flux at steady state. The only thing you know next is that  $J_2 = J_3 + J_5$  and  $J_3 = J_4$  and  $J_5 = J_6 = J_7$ . If you would know one more flux value, i.e.  $J_3, J_4, J_5, J_6$ , or  $J_7$ , then you would know all fluxes. This is the essence of determining the independent fluxes.

<sup>1</sup>Remember the convention that enzyme conversion rates are called fluxes at steady-state.

Thus, 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 given,

$$\begin{aligned} J_1 &= J_2 \\ J_2 &= J_3 + J_5 \\ J_3 &= J_4 \\ J_5 &= J_6 \\ J_6 &= J_7 \end{aligned} \tag{9.34}$$

These flux relationships can be captured in matrix form,

$$\begin{pmatrix} J_1 \\ J_3 \\ J_2 \\ J_4 \\ J_5 \\ J_6 \\ J_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} J_1 \\ J_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I} \\ \mathbf{K}_0 \end{pmatrix}}_{\mathbf{K}} \begin{pmatrix} J_1 \\ J_3 \end{pmatrix} \tag{9.35}$$

This equation indicates that the kernel matrix  $\mathbf{K}$  relates all flux values to the values of the independent fluxes. Each column of  $\mathbf{K}$  is a segment of the metabolic network that can attain a steady-state state by itself. These segments are called flux modes. Any linear combination of these flux modes given a steady-state flux distribution of this network.

For systems without moiety-conservation relationships 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.

### Exercise

1. Identify the flux modes of the network displayed in Figure 9.5. Construct a new valid kernel matrix  $\mathbf{K}$  that has only positive entries by taking a linear combination of the columns of the kernel matrix  $\mathbf{K}$  given in Equation 9.35.
2. In Figure 9.6 the TCA cycle is shown, which is an important pathway in the central metabolism of cells as it generates precursors for growth and electron carriers for oxidative phosphorylation, which makes ATP. The

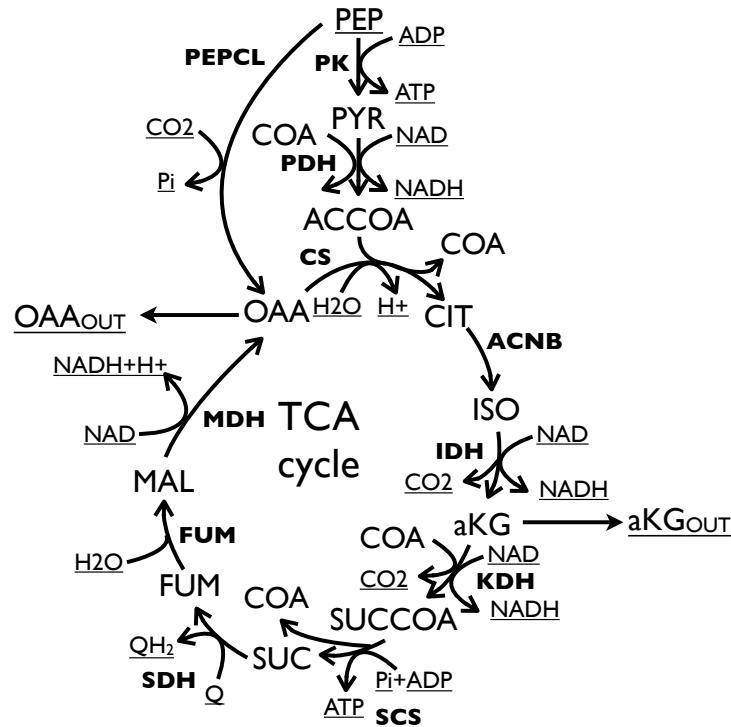


Figure 9.6: The TCA cycle. The underlined metabolites are considered fixed. Bold names denote enzyme names; PK: pyruvate kinase, PDH: pyruvate dehydrogenase, CS: citrate synthase, ACNB: aconitase B, IDH: isocitrate dehydrogenase, KDH:  $\alpha$ -ketoglutarate dehydrogenase, SCS: succinate synthase, SDH: succinate dehydrogenase, FUM: fumarase, MDH: malate dehydrogenase, and PEPCL: PEP carboxylyase. The efflux of aKG and OAA is also shown. PEP is an intermediate of glycolysis.

kernel matrix of this network is given by,

$$\mathbf{K} = \begin{pmatrix} & \text{PEPCL} & \text{KDH} & \text{KGout} \\ \text{PEPCL} & 1 & 0 & 0 \\ \text{KDH} & 0 & 1 & 0 \\ \text{KGout} & 0 & 0 & 1 \\ \text{PK} & 0 & 1 & 1 \\ \text{PDH} & 0 & 1 & 1 \\ \text{CS} & 0 & 1 & 1 \\ \text{OAAout} & 1 & 0 & -1 \\ \text{ACNB} & 0 & 1 & 1 \\ \text{IDH} & 0 & 1 & 1 \\ \text{SCS} & 0 & 1 & 0 \\ \text{SDH} & 0 & 1 & 0 \\ \text{FUMA} & 0 & 1 & 0 \\ \text{MDH} & 0 & 1 & 0 \end{pmatrix} \quad (9.36)$$

This matrix indicates that if  $PEPCL$ ,  $KDH$ , and  $KGout$  are chosen as independent fluxes all flux values can be determined from their flux values. Show that each of the columns of the  $\mathbf{K}$  matrix represents a valid steady state route of the network by drawing them in Figure 9.6. This is the basic interpretation of the nullspace of the stoichiometric matrix. However, because the nullspace is not unique, the pathways obtained directly from MATLAB or Mathematica may look counterintuitive and a linear combination of them would yield more intuitive valid steady-state flux distributions as pathway routes.

## 9.7 Determination of $\mathbf{K}$ with linear algebra

The kernel matrix  $\mathbf{K}$  equals the right nullspace of  $\mathbf{N}$ ,

$$\mathbf{NK} = \mathbf{0} \quad (9.37)$$

$\mathbf{K}$  can be obtained with a linear algebra trick. Any steady-state flux vector can be obtained from a linear combination of the columns of  $\mathbf{K}$ .

We will use the metabolic network given in Figure 9.5 to illustrate the derivation of  $\mathbf{K}$  with linear algebra. We start from the stoichiometric matrix,

$$\mathbf{N} = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ x_1 & 1 & -1 & 0 & 0 & 0 & 0 \\ x_2 & 0 & 1 & -1 & 0 & 0 & -1 \\ x_3 & 0 & 0 & 1 & -1 & 0 & 0 \\ x_4 & 0 & 0 & 0 & 0 & 1 & -1 \\ x_5 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (9.38)$$

Previously, we augmented this matrix with an identity matrix to determine the moiety conservation and the independent fluxes. If the interest is only to determine the independent fluxes then the identity matrix is not required. Next, Gaussian elimination is performed we transform this stoichiometric matrix to its reduced row echelon form (rref),

$$rref(\mathbf{N}) = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ x_1 & 1 & 0 & 0 & -1 & 0 & 0 & -1 \\ x_2 & 0 & 1 & 0 & -1 & 0 & 0 & -1 \\ x_3 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ x_4 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ x_5 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (9.39)$$

The number of nonzero rows in the row reduced echelon form of  $\mathbf{N}$  gives the rank of this matrix. No nonzero rows occurs thus the rank of this matrix is five. In other words, there are no dependent metabolites. The reactions of the columns that do not contain a pivot are the independent fluxes, as we also learned above; hence these are reaction 4 and 7 (indicated in green in the previous matrix). Every row in this matrix is now a steady state flux constraint and all fluxes can

be determined from values of  $J_4$  and  $J_7$ . Inspection of those relationships gives rise to the following conclusion,

$$\begin{pmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \\ J_5 \\ J_6 \\ J_7 \end{pmatrix} = J_4 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + J_7 \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{\text{Nullspace of } \mathbf{N}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} J_4 \\ J_7 \end{pmatrix} \quad (9.40)$$

Finally, the rows of the nullspace can be reordered to write it as  $\mathbf{K} = \begin{pmatrix} \mathbf{I} \\ \mathbf{K}_0 \end{pmatrix}$ ,

$$\begin{pmatrix} J_4 \\ J_7 \\ J_1 \\ J_2 \\ J_3 \\ J_5 \\ J_6 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}}_{\text{kernel } \mathbf{K}} \begin{pmatrix} J_4 \\ J_7 \end{pmatrix} \quad (9.41)$$

Note that this kernel matrix  $\mathbf{K}$  looks completely different than the one shown in Equation (9.35), because different dependent and independent metabolites were used. Both  $\mathbf{K}$  matrices give rise to a valid nullspace of  $\mathbf{N}$ .

### Exercises

1. Use Mathematica to derive the kernel matrix  $\mathbf{K}$  for the network displayed in Figure 9.9.
2. Determine the  $\mathbf{N}$ ,  $\mathbf{K}$ , and  $\mathbf{L}$  matrix for the four metabolic networks displayed in Figure 9.8.

## 9.8 Determining moiety conservation and independent fluxes in Mathematica

The simplest way for the exercises is to the manual procedure described above, by performing Gaussian elimination on the stoichiometry matrix augmented with an identity matrix, in Mathematica. This is done with the following Mathematica input:

`MatrixForm[RowReduce[Join[N, IdentityMatrix[Dimensions[N][[1]]], 2]]]`  
with  $\mathbf{N}$  as the stoichiometry matrix. What does this mean?

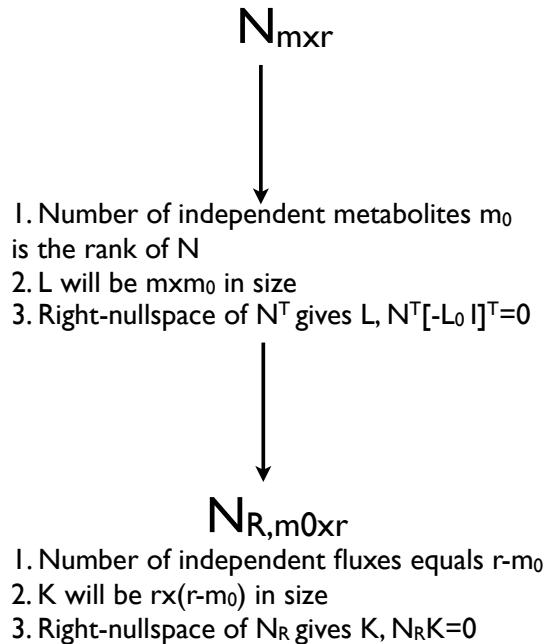
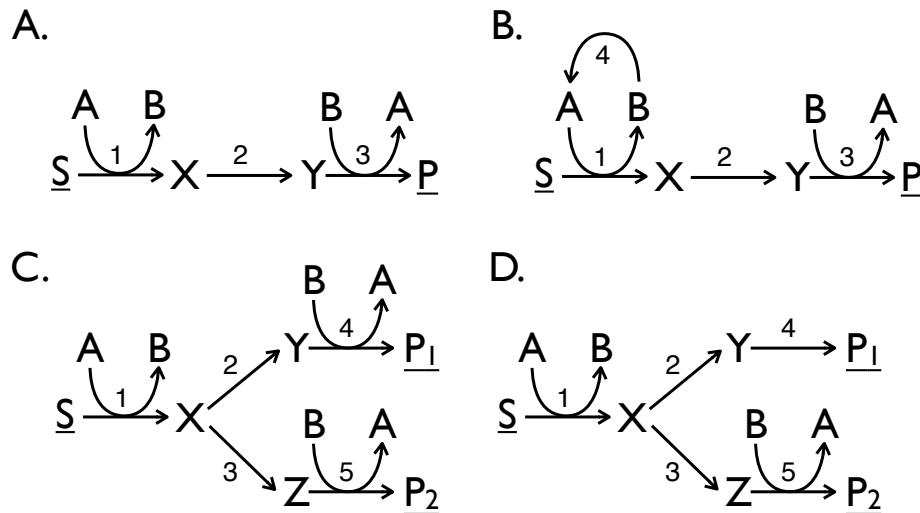
Figure 9.7: Workflow for **L** and **K** determination.

Figure 9.8: Four metabolic networks with that differ in independent fluxes and metabolic intermediates.

1. `IdentityMatrix[Dimensions[N][[1]]]` means make an identity matrix of size given by the number of rows of the stoichiometric matrix. The number of rows of the stoichiometric matrix equals `Dimensions[N][[1]]`.
2. `Join[N, IdentityMatrix[Dimensions[N][[1]]], 2]` augment the identity matrix behind the **N** matrix (without the "2" the identity matrix is stacked under the **N** matrix but we want it after **N**).
3. `RowReduce[Join[N, IdentityMatrix[Dimensions[N][[1]]], 2]]` performs the row reduction.
4. `MatrixForm[A]` displays a matrix **A** in human readable format.

The outcome is then what you could also have obtained by hand. Note that these approaches do not give unique results – different choices of independent fluxes of concentrations are often possible – and therefore the Mathematica result may look different than your result (but is always correct). Note that `MatrixForm[RowReduce[N]]` allows you to identify the independent fluxes as was done above in section 9.7.

## 9.9 Advanced material

### 9.9.1 Relations between N and L

In the general case, we obtain for equation (9.5) a decomposed version with the independent and dependent intermediates separated if dependent intermediates indeed exist (it may look complicated but study it a moment and you will see the logic),

$$\dot{\mathbf{x}} = \mathbf{N}\mathbf{v} \quad (9.42)$$

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{\mathbf{x}}^I \\ \dot{\mathbf{x}}^D \end{pmatrix} \quad (9.43)$$

$$\begin{pmatrix} \dot{\mathbf{x}}^I \\ \dot{\mathbf{x}}^D \end{pmatrix} = \begin{pmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{pmatrix} \mathbf{v} = \mathbf{L} \mathbf{N}_R \mathbf{v} = \begin{pmatrix} \mathbf{I} \\ \mathbf{L}_0 \end{pmatrix} \mathbf{N}_R \mathbf{v} \quad (9.44)$$

$$\dot{\mathbf{x}}^I = \mathbf{N}_R \mathbf{v} \quad (9.45)$$

$$\dot{\mathbf{x}}^D = \mathbf{L}_0 \dot{\mathbf{x}}^I \quad (9.46)$$

with:  $\mathbf{x}^I$  and  $\mathbf{x}^D$  as the concentration vector of independent and dependent metabolites, respectively. Here we have used the definition of the **L** matrix.

Identification of the moiety-conservation relationships can be done by realising that the following equation should be correct (obtained by integrating equation 9.46),

$$\mathbf{x}^D - \mathbf{L}_0 \mathbf{x}^I = \mathbf{T} \quad (9.47)$$

When we again consider the metabolic network shown in Figure 9.1 the link matrix  $\mathbf{L}$  is given by,

$$\mathbf{L} = \begin{pmatrix} \mathbf{I} \\ \mathbf{L}_0 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & A \\ X_1 & 1 & 0 & 0 \\ X_2 & 0 & 1 & 0 \\ A & 0 & 0 & 1 \\ B & 0 & 0 & -1 \\ X_3 & -1 & -1 & -1 \end{pmatrix} \quad (9.48)$$

This means that  $X_1$ ,  $X_2$ , and  $A$  are the independent metabolites. The concentrations of the remaining dependent metabolites are linearly related to the independent metabolites through,

$$\mathbf{L}_0 = \begin{pmatrix} X_1 & X_2 & A \\ B & 0 & 0 & -1 \\ X_3 & -1 & -1 & -1 \end{pmatrix} \quad (9.49)$$

The reduced stoichiometric matrix now corresponds to the rows of the independent metabolites of the  $\mathbf{N}$  matrix,

$$\mathbf{N}_R = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 \\ X_1 & 1 & -1 & -1 & 0 & 0 \\ X_2 & 0 & 0 & 1 & -1 & 0 \\ A & -1 & 0 & 0 & 1 & 1 \end{pmatrix} \quad (9.50)$$

Using Equation (9.47) we obtain the moiety-conservation relationships,

$$\begin{aligned} \text{constant}_1 &= A(t) + B(t) = A(0) + B(0) \\ \text{constant}_2 &= A(t) + X_1(t) + X_2(t) + X_3(t) \\ &= A(0) + X_1(0) + X_2(0) + X_3(0) \end{aligned} \quad (9.51)$$

So, all the definitions in equation 9.46 can be obtained when  $\mathbf{L}$  is known.

### Exercises

The metabolic network displayed in Figure 9.9 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.

### 9.9.2 Derivation of the $\mathbf{K}$ and $\mathbf{L}$ matrix using linear algebra for any $\mathbf{N}$ matrix

In order to determine the  $\mathbf{K}$  and  $\mathbf{L}$  matrices for any  $\mathbf{N}$  matrix we have to follow the scheme displayed in Figure 9.7. From above, we conclude from equation (9.46) that the following has to be correct,

$$\mathbf{L}_0 \mathbf{N}_R = \mathbf{N}_0 \quad (9.52)$$

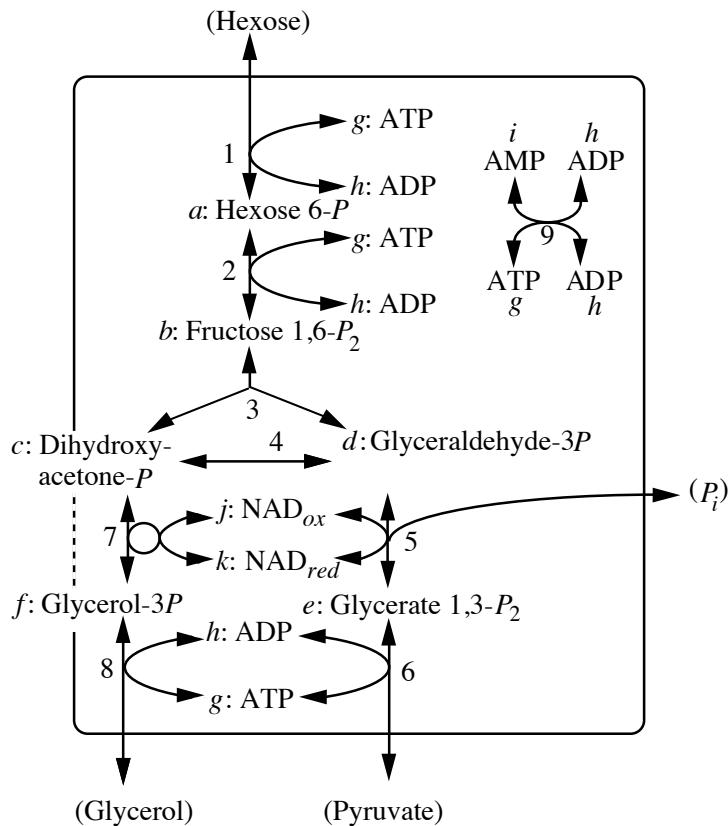


Figure 9.9: Simplified representation of the glycolysis as it occurs in the glycosome of *Trypanosomes*. Besides pyruvate, ATP will be formed out of ADP outside of the glycosome in reaction 6.

This equation can be rearranged to yield,

$$\begin{pmatrix} -\mathbf{L}_0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{pmatrix} = \mathbf{0} \quad (9.53)$$

and therefore,

$$\begin{pmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{pmatrix}^T \begin{pmatrix} -\mathbf{L}_0 & \mathbf{I} \end{pmatrix}^T = \mathbf{0} \quad (9.54)$$

This equation states that the right nullspace of the transpose of  $\mathbf{N}$  (the left nullspace of  $\mathbf{N}$ ) equals  $(-\mathbf{L}_0 \quad \mathbf{I})^T$ . Remember from linear algebra that the left null space of a matrix  $\mathbf{A}$  is a matrix  $\mathbf{B}$  such that  $\mathbf{B} \cdot \mathbf{A} = \mathbf{A}^T \cdot \mathbf{B}^T = 0$ . The right null space is defined accordingly. The dimensions of  $\mathbf{L}$  are therefore  $m \times m_0$  (see equation (9.48)).

The **K** is then obtained by directly taking the right nullspace of **N** and identification of the independent fluxes. The number of independent fluxes equals the number of columns of **K**. The reordering of the rows of **K** and **L** such that stacked matrices of an identity matrix on top of another matrix may in some cases row and column reordering by hand or computationally.



# Chapter 10

## Flux balance analysis

### 10.1 Steady-state flux space and constraint-based modeling

In the previous chapter we have seen how to analyze 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 possibilities 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.

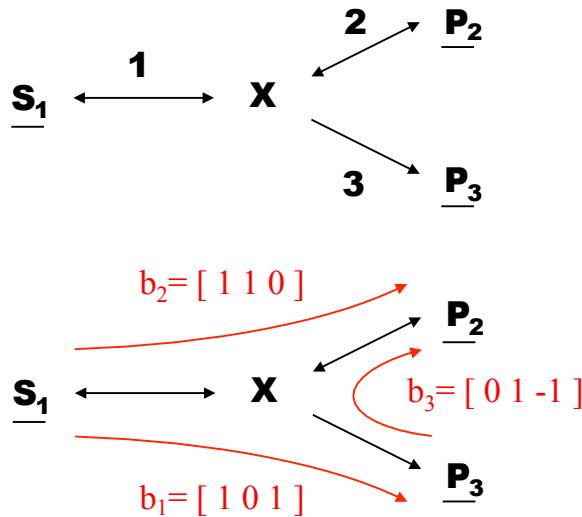


Figure 10.1: Simple three-enzyme pathway with a branch.

A simple branched pathway is depicted in Figure 10.1, with  $\mathbf{N} = [1, -1, -1]$ . The rank of  $\mathbf{N}$  is one and since there are three rates, the dimension of the right nullspace of  $\mathbf{N}$  is two which is equivalent to two independent fluxes. If we take  $\mathbf{v}_1$  and  $\mathbf{v}_2$  as independent fluxes,  $\mathbf{K}$  equals,

$$\mathbf{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \quad (10.1)$$

The columns of  $\mathbf{K}$  can be interpreted as flux modes as indicated in Figure 10.1. These are pathways running from  $S_1$  to  $P_3$  ( $b_1$ ) and  $P_3$  to  $P_2$  ( $b_3$ ), respectively. They are base vectors that span the steady-state solution space, which in this case is a plane in 3D-space as is illustrated in Figure 10.2. Note that we have no constraints on the values that reactions can take, i.e., the plane stretches out to infinity in both directions.

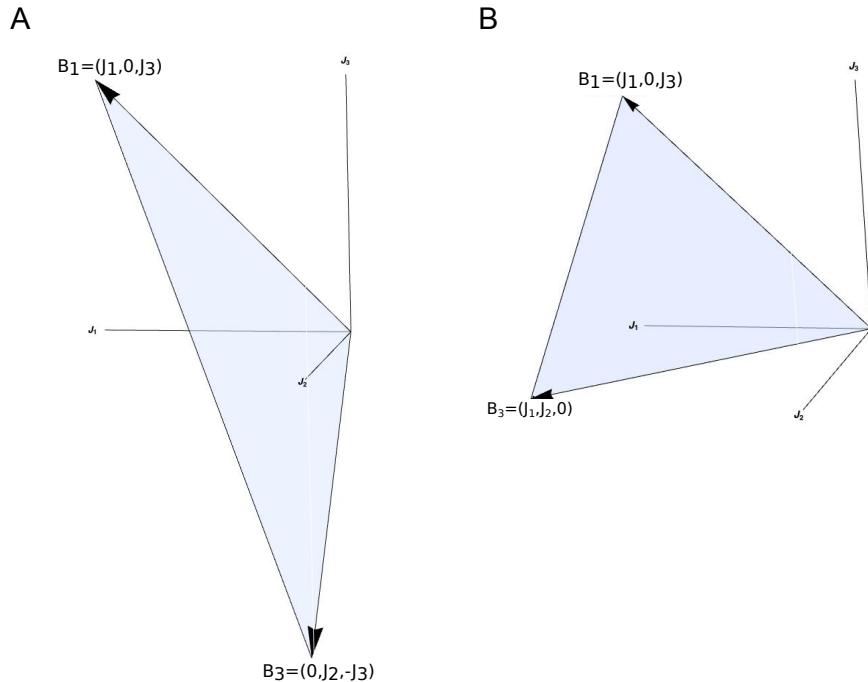


Figure 10.2: Right nullspace of the simple branched pathway shown in Figure 10.1 spanned by two different sets of basis vectors. Note that they span the same plane which stretches out in all directions if there are no additional constraints: any linear combination of the basis vectors is a point in this nullspace.

Suppose, however, that the third reaction is irreversible under physiological conditions such that the first constraint on the reaction network is  $v_3 > 0$ .

This means that the mass action ratio of  $X_1$  and  $P_3$  cannot compensate within reasonable bounds of  $[X_1]$  and  $[P_3]$  for a large equilibrium constant of the third reaction. 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 right nullspace of  $\mathbf{N}$ . Alternatively, adding the first to the second column from  $\mathbf{K}$  given in Equation (10.1) results in,

$$\mathbf{K} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (10.2)$$

This is 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  $\mathbf{K}$  equally well describes the unconstrained right nullspace of  $\mathbf{N}$  as did  $\mathbf{K}$  of Equation (10.1).

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. These constraints reduce the number of feasible base vectors to describe the right nullspace. 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 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 constraint-based modeling reviews to visualize the flux space. This solution space is under non-negativity a flux cone, some sort of ice cream cone spanned by base vectors of  $\mathbf{K}$  (see Figure 10.3). 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 shown in Figure 10.1 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 10.2.

Apart from constraints on the direction of reactions, there are also constraints possible on the maximal rates of certain enzymes. These constraints can also be represented as planes that reduce the solution space of possible steady-state flux distributions. An example of these constraints is given in Figure 10.4 with  $0 < J_1 < J_{1,max}$  and  $0 < J_2 < J_{2,max}$ . Constraints on the direction of fluxes and on their rates are within constraint-based modeling collectively called capacity constraints. The maximal rates of enzymes are needed to bound the solution space into a closed polytope, because the solution space is unbounded without proper capacity constraints. This is not biologically possible and bounded solution spaces are required to ask questions 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-conservations and hence reduce the number of variables that determine the dynamic behavior 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

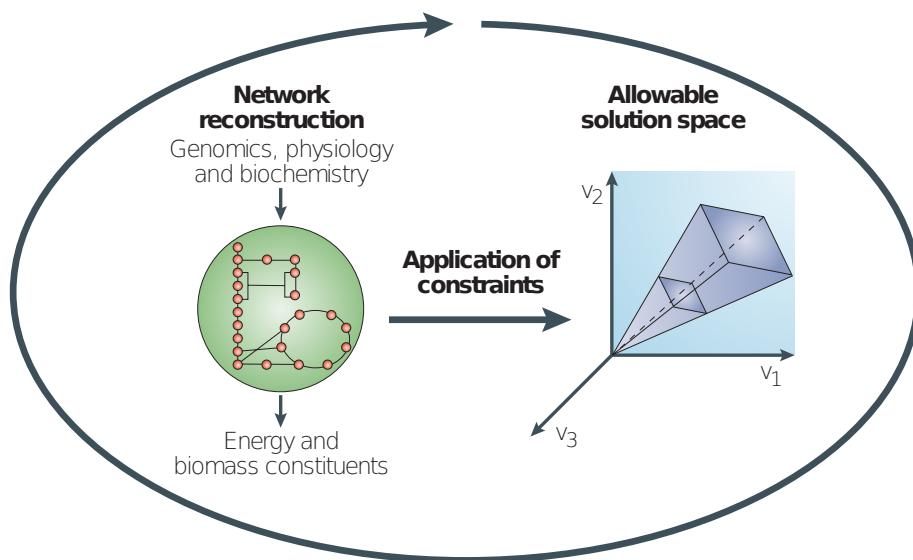


Figure 10.3: Popular representation of the flux cone in 3D. Picture taken from Price et al. (2004) Nat Rev Micro 2, 886.

the kinetic parameters.<sup>1</sup> The kinetic parameters, as we will see, determine the specific steady-state that will be reached in the solution space. These kinetic parameters will determine together with the initial conditions the trajectory towards the steady-state and the control structure in the steady-state (see also Figure 10.5).

The basics of common modeling strategies and the difference between them is exemplified in Figure 10.5. A metabolic network is shown in Figure 10.5 A where a source can be converted into biomass and by-product through several chemical conversions, which involves three intracellular metabolites,  $C_1$ ,  $C_2$ , and  $C_3$ , and two biomass precursors,  $P_1$  and  $P_2$ . 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. As a consequence, these variables can be treated as constant at each point in time (pseudo steady-state assumption through timescale separation).

In metabolic flux analysis and in constraint-based modeling, steady-states are assumed which postulates that all metabolites are balanced and therefore  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$ . This results in a set of homogeneous equations. The set of all possible solutions to this set of equations is called the solution space (see Figure 10.5 C), which mathematically corresponds to the right nullspace of the stoichiometric matrix. The set of equations in  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  is usually

---

<sup>1</sup>Given fixed source and sink concentrations and assuming for now an unique and stable solution to the set of ODE's of the model.

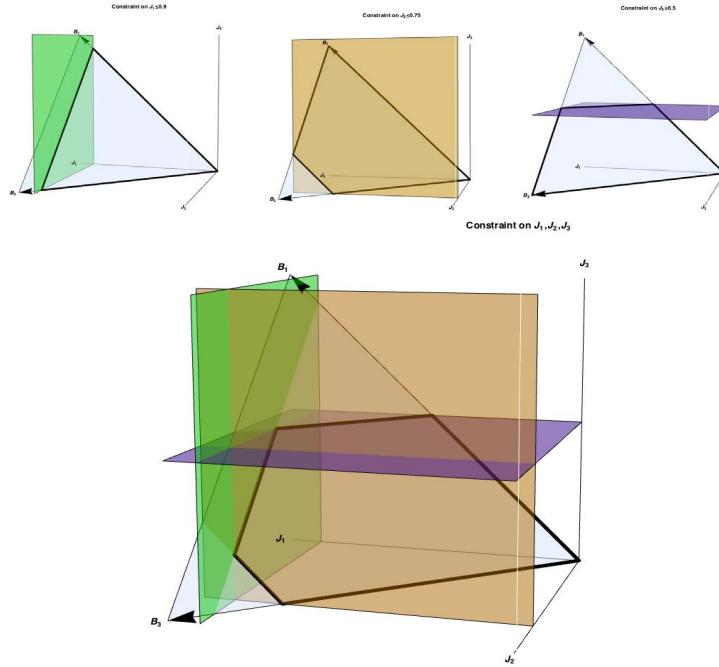


Figure 10.4: Flux cone of simple branched pathway of Figure 10.1 as the result of capacity constraints applied to the original nullspace. Constraints on  $J_1 \leq 0.9$  (green plain),  $J_2 \leq 0.75$  (orange plain), and  $J_3 \leq 0.5$  (purple plain) bound the solution space of this pathway.

underdetermined, resulting in a large solution space. Also in this example, there are five metabolites and seven reactions (five unknowns and seven equations). The solution space is therefore two-dimensional.

We can solve  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  by measuring enough unknown fluxes (reaction rates in steady-state) so that the system becomes determined (or preferably overdetermined). Alternatively, <sup>13</sup>C-labeling studies can be used to estimate fluxes. Subsequently, the state the system is in, a point in the solution space (see Figure 10.5 D), 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.

A predicted solution to  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  is found by assuming that the system fulfills 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 colored vectors shown in Figure 10.5 E. These red colored vectors are sometimes referred to as the ‘line of optimality’, corresponding to optimal biomass formation or by-product formation. To find a particular maximum, one or more

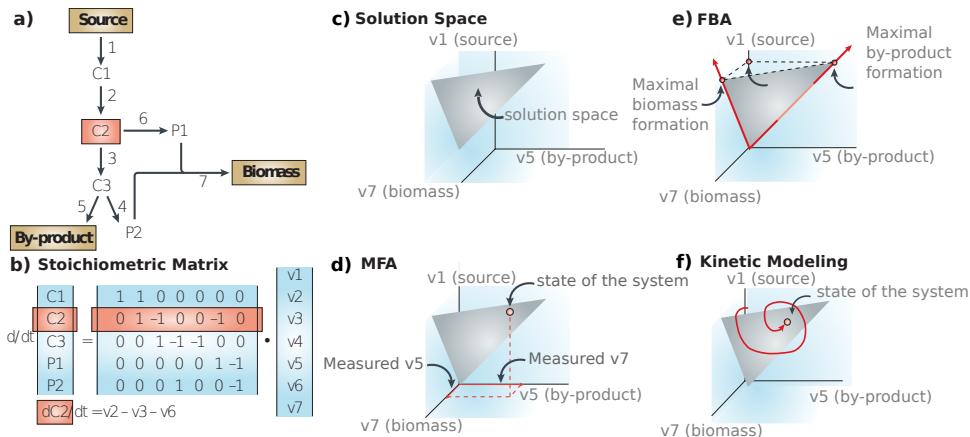


Figure 10.5: A) Metabolic network visualization. B) Stoichiometric network of the metabolic network model. One can derive for each compound in the network a differential equation describing its dynamics as a function of the reaction rates. An example is given for metabolite  $C_2$ . C) Steady-state approach:  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$ . 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. D) Metabolic flux analysis (MFA).  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  is solved by measuring enough unknown fluxes (reaction rates in steady-state) so that the system becomes determined (or preferably overdetermined). Alternatively,  $^{13}\text{C}$ -labeling studies can be used to estimate fluxes. For this reason, the state the system is in, a point in the solution space, is estimated based on experimental data. E) Flux balance analysis (FBA). A predicted solution to  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  is found by assuming that the system fulfills an optimality condition, and therefore the state of the system could be predicted by finding the optimal flux distribution for that objective function. F) Kinetic modeling. One might make a full kinetic model where all kinetic parameters are collected. Consequently, one can numerically solve  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  with  $v = (C, p)$ ; 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. Figure taken from [39].

capacity constraints are required, otherwise the solution to the optimization will be unbounded (the vector goes to infinity). At a given source capacity constraint the corresponding optimal solutions for growth and by-product formation are indicated as red dots.

Alternatively, one can numerically solve  $\dot{\mathbf{X}}(t) = \mathbf{Nv} = \mathbf{0}$  if a full kinetic model is available where all kinetic parameters are collected. 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.

But also without any kinetics, stoichiometric analysis have been popular, and also successful, in analyzing metabolic networks. In metabolic engineering, stoichiometric network analysis is used to deduce flux distributions from measured data. The idea is simple: if the dimension of the solution space is five, meaning there are five independent fluxes that together determine all dependent fluxes, one needs to just measure those five 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 kernel matrix  $\mathbf{K}$  in which all independent fluxes can be measured experimentally. This activity is often referred to as metabolic flux analysis.

## 10.2 Flux Balance Analysis

In the genomics era, especially the stoichiometric analysis of large genome-scale metabolic networks have been useful to turn sequenced genomes into mathematical models representing the complete metabolic network of the sequenced organism. The approaches for making such metabolic reconstructions fall outside the scope of this book. 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 analysis of such networks. Even the steady-state solution space of these networks is hugely complex, but there are a number of analysis that have proven to be useful.

The most popular and influential one is called flux balance analysis (FBA). With FBA linear optimization is used to find flux distributions that maximize or minimize a certain *objective function* considering limited resources (capacity constraints). This objective function is itself a flux or a linear combination of

fluxes. The optimization problem reads:

$$\begin{aligned} & \max Z \\ & \text{subject to} \\ & \mathbf{Nv} = \mathbf{0} \text{ (steady-state constraint, often called "mass-balance" constraints)} \\ & a_i < v_i < b_i \text{ for all } v_i \text{ elements of } \mathbf{v} \text{ (capacity constraints)} \end{aligned} \tag{10.3}$$

All mathematical functions in the model are required to be linear functions. For this reason, the contribution of each reaction to  $Z$  is proportional to the influence of this reaction in the objective function. Note that if  $a_i$  is zero, the corresponding reaction is 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 linear optimization does we will make use of the same example as in the previous section, the network of Figure 10.1. The solution space is drawn as a projection on the  $(J_2, J_3)$  plane in Figure 10.6. Important is the mass-balance line  $J_2 + J_3 = J_1$ . By constraining  $J_1$  to  $\max(J_1)$  the solution space is already bounded. For instance, given that  $J_1 = 10$  maximization of  $J_3$  gives  $J_3 = 10$  and  $J_2 = 0$  while maximization of  $J_2$  gives  $J_2 = 10$  and  $J_1 = 0$ .

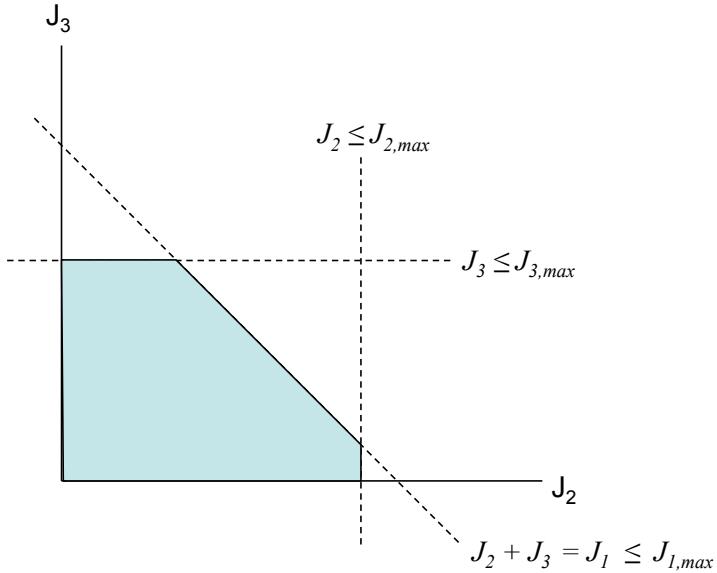


Figure 10.6: 2D projection of the solution space of the network displayed in Figure 10.1. Indicated are a mass balance constraint ( $J_2 + J_3 = J_1$ ) and two capacity constraints on  $J_2$  and  $J_3$

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

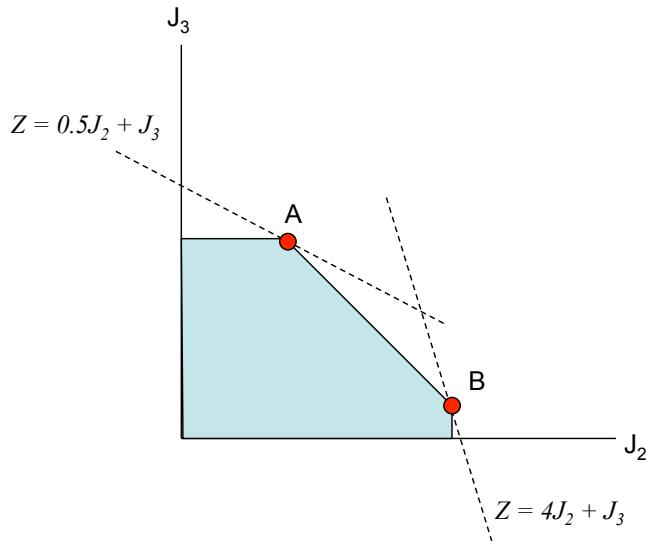


Figure 10.7: FBA solution of the network displayed in Figure 10.1 with an objective function with different weights on  $J_2$  and  $J_3$ .

we can immediately see what flux distribution would maximize a certain objective function,  $Z$ . We can express this objective function 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 10.7 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 10.8) (iii) an infinite number of solutions: if  $w_2$  is zero and  $w_3$  is nonzero, any point between the y-axis and point A would be a solution. There is of course 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.

In the case of 500 to 1000 reactions, we have to resort to computers rather than inspection to find the optima. Efficient algorithms for these type of linear optimization problems have been developed collectively called linear programming. A FBA problem with more than 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 modeling. The constraints come from physiology 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 and what is the composition of the cell in terms of i.e. proteins, lipids, RNA, DNA, and carbohydrates.

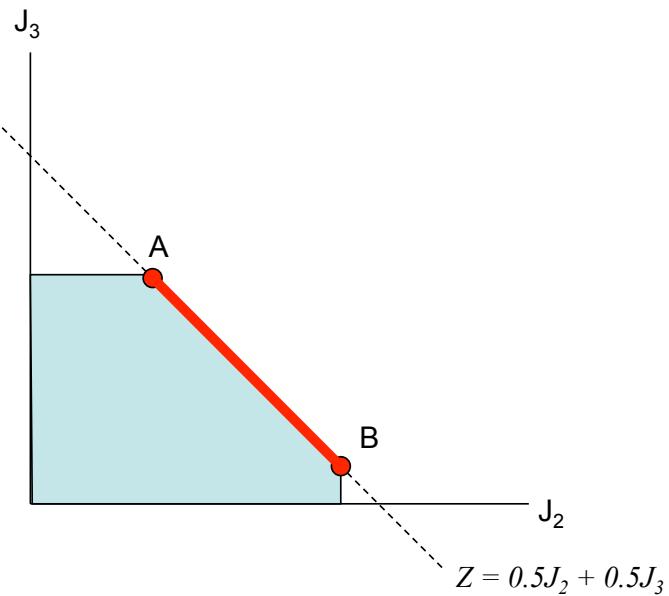


Figure 10.8: FBA solution of the network displayed in Figure 10.1 with an objective function with equals weights on  $J_2$  and  $J_3$ .

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 optimization 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 a FBA solution is called flux variability analysis (FVA). It should follow any FBA solution of interest. FVA is formulated as:

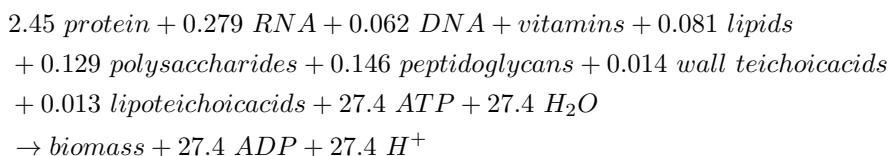
$$\begin{aligned}
 & \max / \min v_i \text{ for all } v_i \text{ elements in } \mathbf{v} \\
 & \text{subject to} \\
 & \mathbf{Nv} = \mathbf{0} \text{ (steady-state constraint, often called "mass-balance" constraints)} \\
 & a_i < v_i < b_i \text{ for all } v_i \text{ elements of } \mathbf{v} \text{ (capacity constraints)} \\
 & Z = Z_{\max} \text{ FBA result}
 \end{aligned} \tag{10.4}$$

The last constraint ensures that the maximal and minimal fluxes through each individual flux is evaluated at the optimal value of the objective function. FVA involves many rounds of FBA, two times the number of reactions in the network. Subsequently, the span of a reaction is defined as the maximal minus the minimal flux value: if the flux through a reaction is completely fixed by the maximal value of  $Z$ , its span is zero. In the example of Figure 10.6, 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.

### 10.3 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, but not necessarily from the perspective of yeast. In 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 optimize growth rate with FBA? To answer this question, we first need to define the objective function, growth rate, within the FBA format. Within stoichiometric modeling growth is modeled as a sink of biomass components (protein, DNA, RNA, lipid, and carbohydrates) that reflects the biomass composition of the cell. By the way of illustration, the objective function for the lactic acid bacterium *Lactobacillus plantarum* is,



The stoichiometric coefficients have unit mmol gDW<sup>-1</sup>. Note the use of 27.4 ATP to form biomass, which reflects 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<sup>-1</sup>), the unit of this reaction is different from the other reactions in the network. Convention is to express fluxes as mmol h<sup>-1</sup> gDW<sup>-1</sup>. Consequently, the unit of the biomass formation rate is h<sup>-1</sup>, the specific growth rate. So it appears indeed that we optimize growth rate, but this is not true. 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, thus 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) which is further illustrated in Figure 10.9.

Optimizing yield or rate is biologically a completely different thing. One can be fast at the expense of efficiency (yield) and this strategy may win the battle of survival of the fittest. Under other conditions such as poor energy resources, maximizing rate may be the best strategy. Importantly, stoichiometric analysis can only predict optimal yields, not *a priori* rates. To turn concentrations into rates, one needs kinetics.

## 10.4 Sensitivity analysis

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

$$r_i = \frac{\partial Z}{\partial b_i} \quad (10.5)$$

For instance, what is change in  $Z$  if we can increase the glucose uptake rate of yeast by a certain percentage. Another example is given in Figure 10.10 where the capacity constraint of  $J_3$  is increased.

Reduced costs are always zero if the FBA solution does not hit the minimal or maximal constraint, thus reduced costs indicate reactions that are somehow constraining the objective function. In many applications, these are input or output fluxes, e.g. the glucose input flux will often constrain the maximal growth rate. Those numbers are therefore interesting if one is interested in medium optimization.

The other sensitivity coefficient is the shadow price, which 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 10.11.

## Exercises

1. Consider the metabolic network given in Figure 10.1. Maximize  $Z = 3J_1 + 2J_2$  subject to  $0 \leq J_1 \leq 4$ ,  $0 \leq J_2 \leq 6$ , and  $3J_1 + 2J_2 \leq 18$ .
  - Draw the solution space in 2D similar to Figure 10.6
  - Is the solution unique? Why (not)?
2. Draw examples of different network topologies where at least five fluxes can be estimated from only two measurements. Explore with what structure external flux measurements do not suffice to solve the set of fluxes.
3. In Figure 10.5 a metabolic network is shown. Assume all reactions are irreversible.
  - Construct a kernel matrix  $\mathbf{K}$  by inspection and by linear algebra (section 9.9.2) 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
  - A flux of  $v_1$  of  $10 \pm 2 \text{ mmol h}^{-1} \text{ gDW}^{-1}$  was measured, and a biomass production of  $4 \text{ h}^{-1}$ . Calculate the predicted span of the byproduct formation flux.
4. In Figure 10.12 a metabolic network is shown with parallel pathways. Assume all reactions are irreversible.
  - The input flux was measured to be  $10 \text{ mmol h}^{-1} \text{ gDW}^{-1}$ . 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.

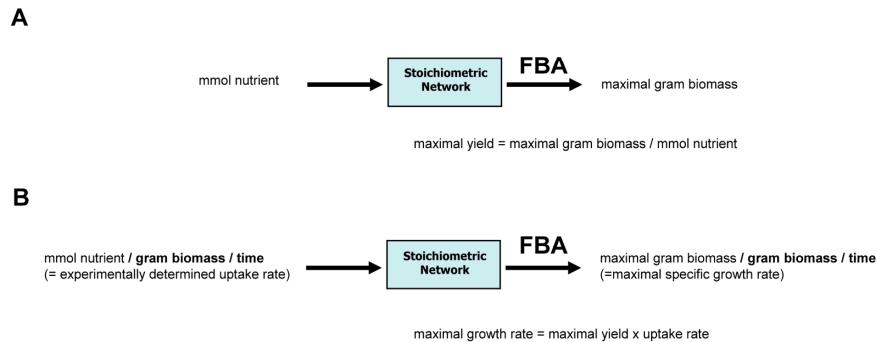


Figure 10.9: A) A stoichiometric network can be used, with FBA, for optimization of maximal yield of biomass on a certain nutrient. B) by providing an experimentally measured input rate (capacity constraint in constraint-based modeling terms), FBA predicts a specific growth rate. The two situations are, 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.

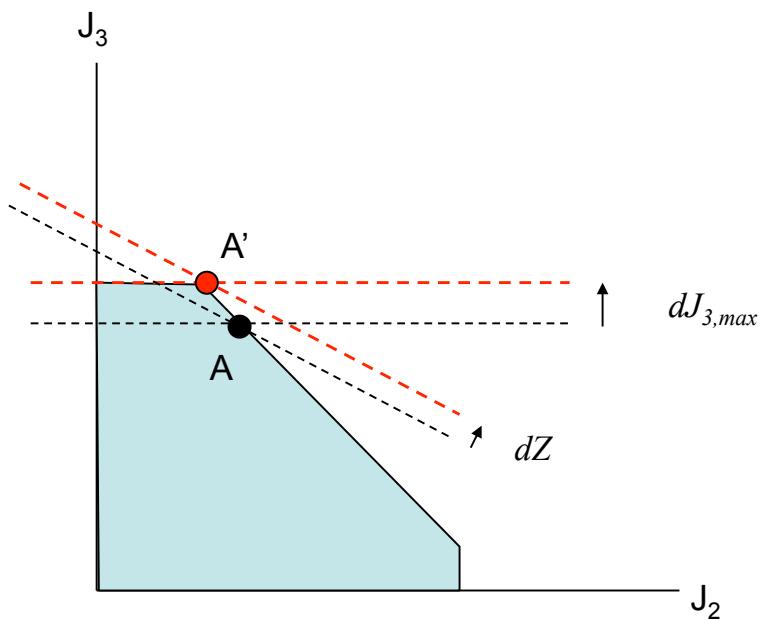


Figure 10.10: Illustration of reduced costs

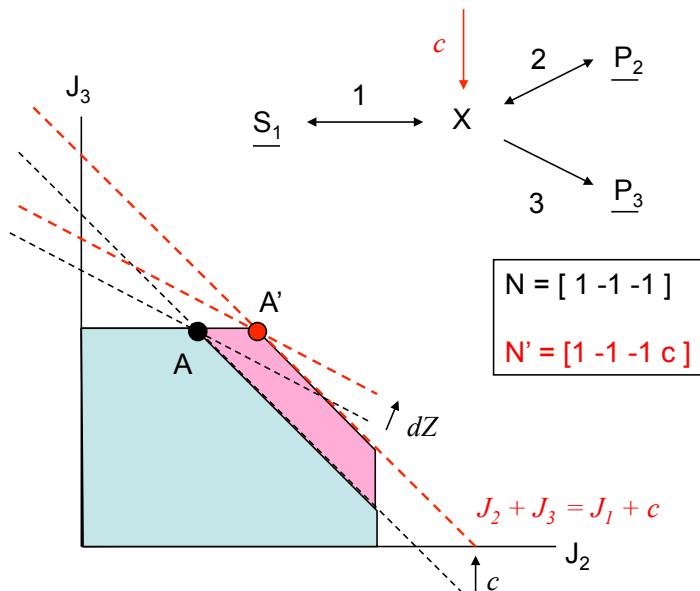


Figure 10.11: Illustration of shadow prices

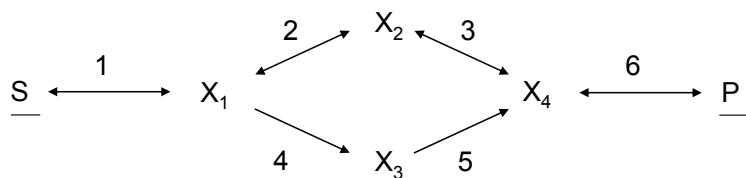


Figure 10.12: Metabolic network with parallel pathways



# Chapter 11

## Dynamics of molecular systems

### 11.1 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 mass-action 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 its 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 system 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,

$$\frac{dx}{dt} = v_s(x) - v_d(x) = f(x) \quad (11.1)$$

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  $\delta x$ . If after some time it returns to  $x_s$  we call the steady state stable and otherwise unstable. So we want to know whether  $\delta x$  converges to zero (stability) or not (instability). For this we have to use equation (11.1) with  $x$  replaced by  $x_s + \delta x$  and evaluate

$$\frac{d(x_s + \delta x)}{dt} = f(x_s + \delta x) \quad (11.2)$$

Here the definition of a derivative comes to the rescue, for small enough values of  $\delta x$ , this definition states that

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

You have seen the prime notation  $f'(x)$  in high school. We write  $\partial$  and not  $d$  because reaction rates may depend on other concentrations and parameters as well. The right hand side is then a *partial* derivative. Note that in the partial fraction notation we don't write  $(x)$  after the differential quotient. We could but we won't if it is clear what is being meant.

Note that we wrote  $\approx$  and not  $=$ . In fact we have

$$f(x_s + \delta x) = f(x_s) + f'(x_s)\delta x + o(\delta x) \quad (11.3)$$

where  $o(\delta x)$  is a quantity which is negligible<sup>1</sup> compared to  $\delta x$  if  $\delta x \rightarrow 0$ , and therefore not specified any further.

Thus we obtain<sup>2</sup> for equation (11.2),

$$\frac{d\delta x}{dt} = \frac{d(x_s + \delta x)}{dt} = f(x_s + \delta x) \approx \underbrace{f(x_s)}_{=0} + f'(x_s)\delta x = (v'_s(x_s) - v'_d(x_s))\delta x \quad (11.4)$$

Changing to partials  $\delta$  again in the notation:

$$\frac{d\delta x}{dt} = \left( \frac{\partial v_s}{\partial x} - \frac{\partial v_d}{\partial x} \right) \delta x \quad (11.5)$$

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  $\lambda$ . To understand the concept of  $\lambda$  as a stability measure we solve the previous differential equation (or just guess the solution, skip the 3 lines below then)

$$\begin{aligned} \frac{d\delta x}{\delta x} &= \lambda dt \\ \int_{\delta x(0)}^{\delta x(t)} \frac{d\delta x}{\delta x} &= \int_0^t \lambda dt \\ \ln \delta x(t) - \ln \delta x(0) &= \lambda t \\ \delta x(t) &= \delta x(0)e^{\lambda t} \end{aligned} \quad (11.6)$$

---

<sup>1</sup>Think of  $(\delta x)^2$

<sup>2</sup>Since  $\frac{d(x_s + \delta x)}{dt} = \frac{dx_s}{dt} + \frac{d\delta x}{dt} = \frac{d\delta x}{dt}$

From the last equation you can see that the perturbation caused by applying  $\delta x(0)$  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  $\lambda$  parameter is called an eigenvalue. Why? Because we may trivially write the right hand side of (11.5) as  $A\delta x$  with  $A$  a  $1 \times 1$  matrix whose only entry is  $\lambda$ , and  $\lambda$  is then an eigenvalue of  $A$ .

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. In general when a steady state becomes unstable several phenomena can occur. We will consider two such phenomena: i. the system jumps off towards a distant steady state (this often happens in the simplest but most dangerous of all bifurcations: the saddle-node bifurcation in which an unstable and a stable steady state come together and disappear; associated with bistability) or ii. the system starts to display regular oscillations (a Hopf bifurcation<sup>3</sup>). 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, but we won't see that yet in scalar equations.

We note that there is another bifurcation that may occur. It is called the pitchfork bifurcation, which occurs when, varying the parameter, the steady state splits up in multiple (typically three) steady states.

### Exercise

- 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). Get used to this terminology:

- Product inhibition means  $\frac{\partial v_s}{\partial x} < 0$
- Product activation means  $\frac{\partial v_s}{\partial x} > 0$
- Substrate activation means  $\frac{\partial v_d}{\partial x} > 0$
- Substrate inhibition means  $\frac{\partial v_d}{\partial x} < 0$

If such a derivative is large the effect is called strong.

- 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?

---

<sup>3</sup>This does not happen for scalar equations

- Can a system with product activation and substrate activation become unstable?
  - Can a system with product inhibition and substrate inhibition negative become unstable?
3. 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?)
4. Consider the following system  $\frac{dx}{dt} = 5 + \underbrace{\frac{20x^5}{1+x^5}}_{\text{synthesis rate}} - \underbrace{15x}_{\text{degradation rate}}$ . 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?
5. Consider equations 11.6, does a system with a more negative  $\lambda$  return to steady state faster or slower than one with a larger (but also negative)  $\lambda$  value?
6. What happens to  $\delta x(t)$  as a function of time for a system with a positive  $\lambda$  and what happens in the case of a negative  $\lambda$ ?

## 11.2 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 11.1 shows 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),

$$\frac{dx}{dt} = \underbrace{5 + \frac{20x^5}{1+x^5}}_{\text{synthesis rate}} - \underbrace{15x}_{\text{degradation rate}} \quad (11.7)$$

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 11.7). 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 11.1). 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 11.1: 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

$$\lambda = \frac{\partial}{\partial x} \left( 5 + \frac{20x^5}{1+x^5} \right) - \frac{\partial}{\partial x} 15x \quad (11.8)$$

This you can do with the derivative function (D) in Mathematica; this gives (test this!),

$$\lambda = -15 - \frac{100x^9}{(1+x^5)^2} + \frac{100x^4}{1+x^5} \quad (11.9)$$

To determine whether the steady state  $x_s = 0.34$  is stable we computed the  $\lambda$  for this value of  $x$ , which gives: -13.7. So this state is stable,  $\lambda$  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  $\lambda = -8.6$ .

## 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 11.1 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 like

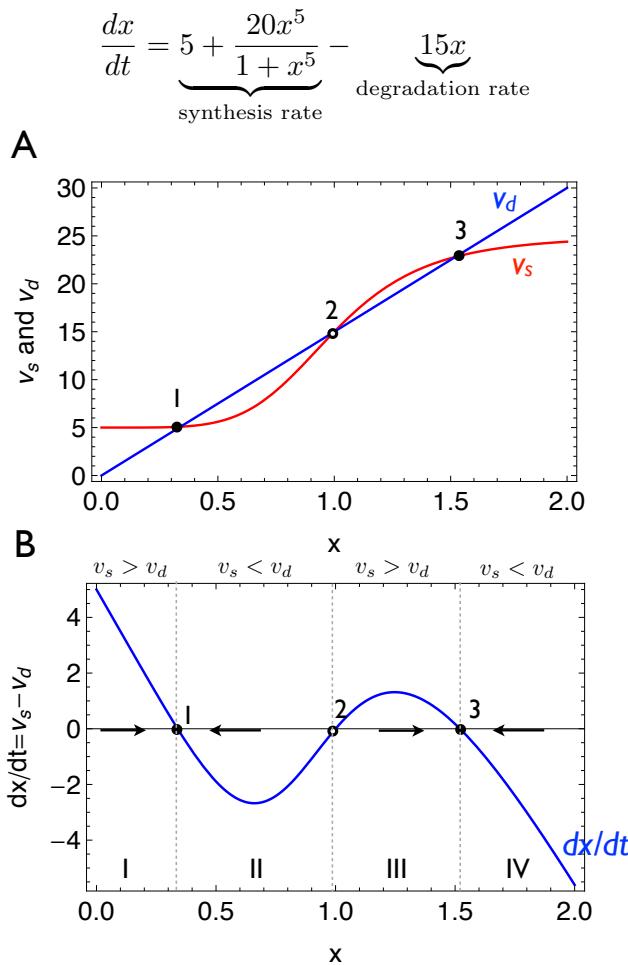


Figure 11.1: **Explanation of bistability.** We consider a system that has autocatalytic synthesis and linear degradation. In A the rates of synthesis (red line) and degradation (blue line) are drawn as function of  $x$ . Three intersections occur labelled 1, 2, and 3 and those are the steady states 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.

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  $y$ ?

### 11.2.1 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  $K$ ,

$$\frac{dx}{dt} = 5 + \frac{20x^5}{K^5 + x^5} - 15x \quad (11.10)$$

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 11.2, 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 11.3 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 Z 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 XPPaut. With a little effort a continuation algorithm can be programmed in Mathematica that does parameter scans and make S/Z-shaped bifurcation curves.

### 11.2.2 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

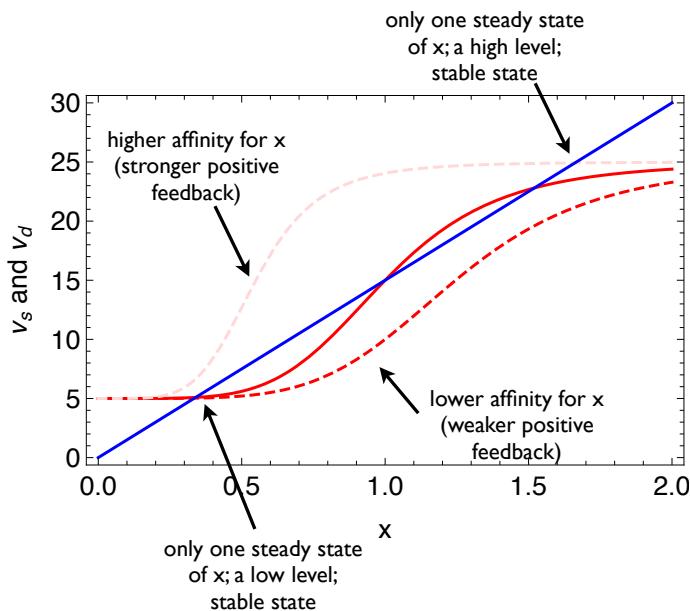


Figure 11.2: **Emergence and disappearance of bistability as function of a parameter.** This plot refers to the same model as analyzed in figure 11.1. Here we consider three different positive feedback strengths of  $x$  on its own synthesis rate - the two dashed lines. We modulate the feedback strength by changing the affinity of the synthesis process for  $x$ , i.e. the 1 in equation 11.7. 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.

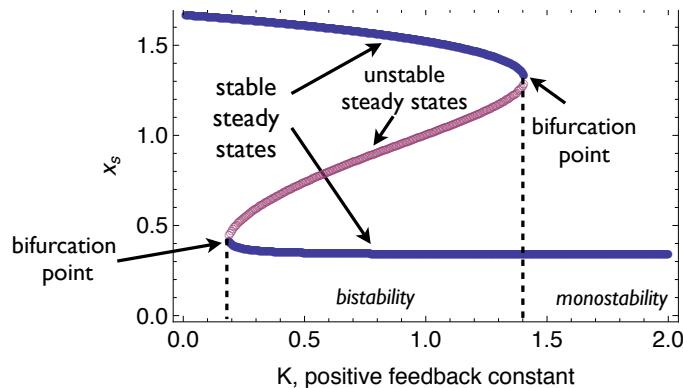


Figure 11.3: A **bifurcation diagram indicating bistability**. For different values of the feedback parameter  $K$  the steady state were determined. Each steady state was checked for stability: stable steady state are denoted by the blue line and unstable steady state are denoted by the purple line. 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  $K$ .

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 11.3 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 remain on the lower branch of steady states until we hit the other bifurcation point; see Figure 11.4. 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.

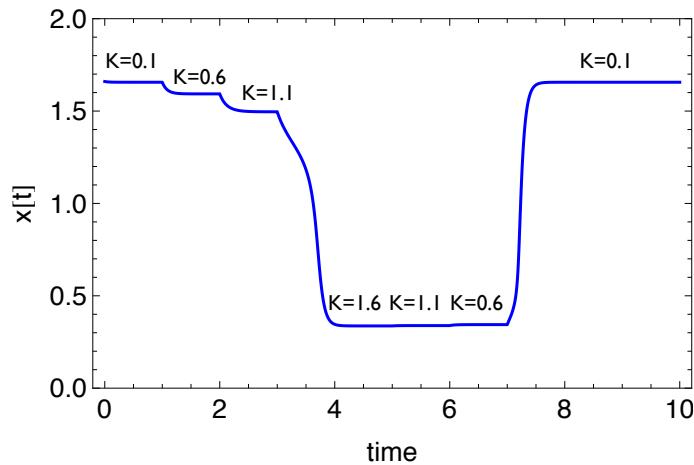


Figure 11.4: **Illustration of hysteresis (memory).** Bistable systems can have different states for the set of parameters depending on their history. This is illustrated in this figure where the dynamics of  $X$  is simulated upon a stepwise decrease of  $K$  followed by a stepwise increase in this parameter. This figure can be understood if figure 11.3 is taking into account because the systems "walks" over this "mirrored" 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  $x$ .

### 11.2.3 Intermezzo: the Goldbeter-Koshland function

Many reaction networks involve blocks like



Here  $R$  is some protein,  $RP$  its phosphorylated form and  $S$  some signal. The concentrations of  $R$  and  $RP$  are written as  $R$  and  $R_P$ . The ODE's are

$$\frac{dR_P}{dt} = \frac{k_1 \cdot S \cdot R}{K_{m1} + R} - \frac{k_2 R_P}{K_{m2} + R_P} = -\frac{dR}{dt}. \quad (11.12)$$

At steady state we have, under the assumption that the total concentration  $R_T$  is constant,

$$\frac{k_1 SR}{K_{m1} + R} = \frac{k_2 R_P}{K_{m2} + R_P}, \quad R + R_P = R_T. \quad (11.13)$$

Introducing dimensionless quantities:

the fraction  $x_p = \frac{R_P}{R_T}$  of phosphorylated RP,

the fraction  $x = \frac{R}{R_T}$  of unphosphorylated R,

and scaled parameters

$$u = k_1 S, \quad v = k_2, \quad J = \frac{K_{m1}}{R_T}, \quad K = \frac{K_{m2}}{R_T}, \quad (11.14)$$

equilibrium means that

$$\frac{ux}{J+x} = \frac{vx_p}{K+x_p}, \quad x + x_p = 1. \quad (11.15)$$

Note that  $u$  contains the signal  $S$ . The solution may be written as

$$x = G(v, u, K, J) = \quad (11.16)$$

$$\frac{2Jv}{(K+1)u + (J-1)v + \sqrt{(K+1)^2u^2 + 2(JK-K-J-1)uv + (J+1)^2v^2}},$$

and

$$x_p = G(u, v, J, K) = \quad (11.17)$$

$$\frac{2Ku}{(K-1)u + (J+1)v + \sqrt{(K+1)^2u^2 + 2(JK-K-J-1)uv + (J+1)^2v^2}}.$$

Here  $G$  is the Goldbeter-Koshland function, a function of 4 dimensionless variables.

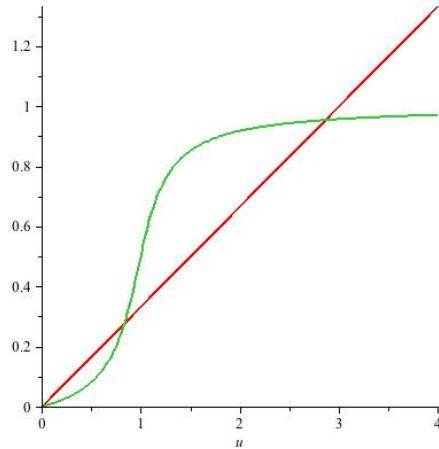


Figure 11.5:  $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ : the graph of  $x_p(u)$  as a function of  $u$  is sigmoidal and has multiple intersections with the graph of  $ku$ .

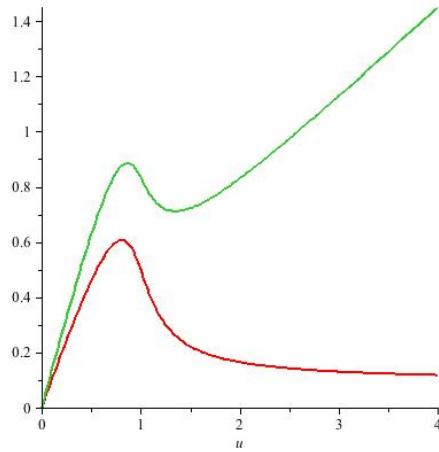


Figure 11.6:  $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ : the graph of  $ux(u)$  as a function of  $u$  has a maximum, the graph of  $ux(u) + ku$  has a maximum and a minimum.

1. You can sketch the graphs of both rates in (11.15), and combine them in one picture, remembering that  $x + x_p = 1$ . Start with a diagram of the degradation rate  $\frac{vx_p}{K+x_p}$  plotted versus  $x_p$ . On the horizontal axis you put  $x = 0$  where  $x_p = 1$  and  $x = 1$  where  $x_p = 0$ , and draw the graph of the  $x$ -dependent synthesis rate  $\frac{ux}{J+x}$  from right to left. Note that in the resulting picture the windows  $0 < x < 1$  and  $0 < x_p < 1$  coincide. You can vary  $u$  and have various synthesis rate curves in the picture. Explain that for each choice of positive values  $u, v, J, K$ , the two graphs intersect in one point only. This identifies the GK-equilibrium, which we will want to consider as a function of the scaled signal  $u$ , i.e.

$$x = x(u), \quad x_p = x_p(u)$$

2. Explain why

$$G(u, v, J, K) + G(v, u, K, J) = 1. \quad (11.18)$$

3. Verify that in equilibrium

$$\frac{R_P}{R_T} = G(k_1 S, k_2, \frac{K_{m1}}{R_T}, \frac{K_{m2}}{R_T}) \quad (11.19)$$

$$\frac{R}{R_T} = 1 - \frac{R_P}{R_T} = G(k_2, k_1 S, \frac{K_{m2}}{R_T}, \frac{K_{m1}}{R_T}). \quad (11.20)$$

4. A simple criterium you can verify is:

$$JK < 1 + J \Leftrightarrow x_p \text{ versus } u \text{ is a convex-concave sigmoidal graph} \quad (11.21)$$

See Fig. 11.5. Thus, if  $J$  and  $K$  are small, the GK-curve is certainly sigmoidal, and quite steep in fact, with a sharp transition from 0 to 1 at  $u \approx v$ , as you can see from the fact that  $x = x_p = \frac{1}{2}$  corresponds to  $\frac{u}{v} = \frac{1+2J}{1+2K}$ .

5. Also of interest, see Fig. 11.6, verify that

$$JK < 1 + K \Leftrightarrow ux(u) \text{ has a (single) maximum} \quad (11.22)$$

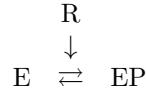
6. GK-formula from solving a quadratic: assume all parameters  $u, v, J, K$  positive, substitute  $x = 1 - x_p$  in the reaction balance in (11.15), derive a quadratic equation for  $\frac{1}{x_p}$  and solve it. You should arrive at

$$x_p = \frac{2uK}{v - u + vJ + uK + \sqrt{(v - u + vJ + uK)^2 - 4(v - u)K}},$$

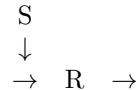
which may be rewritten as (11.17). In fact you get 2 solutions, while there is only one physical solution in the relevant window  $0 \leq x_p \leq 1$ . You have to pick the right sign in front of the square root in the solution formula to get it right.

### 11.2.4 A one-way switch

You will now examine what happens if you combine a GK-block



with a signal-response block



through mutual activation. The degradation rate of R in the signal-response block is

$$v_{degradation} = v_d = k_2 R,$$

and the synthesis rate of R is

$$v_{synthesis} = v_s = k_0 + k_1 S.$$

of signal strength. Without any other signals, the time dependent concentration  $R = R(t)$  of R satisfies

$$\frac{dR}{dt} = v_s - v_d = k_0 + k_1 S - k_2 R, \quad (11.23)$$

and a stationary balance is given by

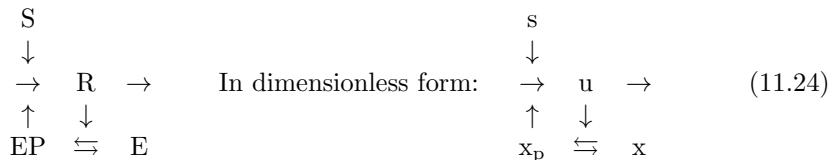
$$R = R_{ss} = \frac{k_0 + k_1 S}{k_2}.$$

Thus, the signal response is just a linear function of  $S$ .

One possibility to combine this signal-response block with a GK-block is to let the response R act a signal in the GK-block, and, simultaneously, have EP enhance<sup>4</sup> the synthesis of R by changing  $v_s$  to

$$v_s = k_0 + k'_0 E_P + k_1 S.$$

The new term in  $v_s$  is proportional to the concentration  $E_P$  of EP. Both S and EP stimulate synthesis. The reaction diagram is




---

<sup>4</sup> Positive feedback between R and EP.

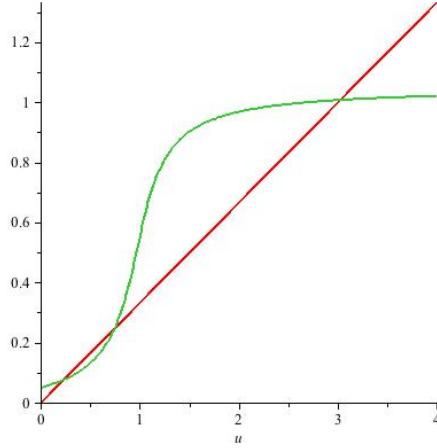


Figure 11.7:  $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ ,  $s = \frac{1}{20}$ : the graph of the synthesis rate  $x_p(u) + s$  as has multiple intersections with the graph of the degradation rate  $ku$ .

Assuming the time scale for E $\rightleftharpoons$ EP to be fast,  $E_P$  is set equal to its GK-equilibrium, so that the synthesis rate of R becomes

$$v_s = k_0 + k'_0 E_P(R) + k_1 S,$$

in which  $E_P(R)$  is the steady state concentration of EP corresponding to the balance

$$\frac{k_3 R E}{K_{m3} + E} = \frac{k_4 E_P}{K_{m4} + E_P}, \quad E + E_P = E_T,$$

as explained in Section 11.2.3. The ODE (11.23) is modified accordingly as

$$\frac{dR}{dt} = v_s - v_d = k_0 + k'_0 E_P(R) + k_1 S - k_2 R \quad (11.25)$$

With

$$s = \frac{k_0 + k_1 S}{k'_0 E_T}, \quad k = \frac{k_2}{k_3 k'_0 E_T}, \quad \tau = k_3 k'_0 E_T t,$$

equation (11.25) can be rewritten as

$$\frac{du}{d\tau} = s + x_p(u) - ku,$$

for  $u = k_3 R$ .

The steady states are solutions of

$$s + x_p(u) = ku. \quad (11.26)$$

The left and right sides in (11.26) are the *normalised synthesis and degradation rates* of  $u$ , see Fig. 11.7. Writing  $s$  as a function of  $u$ , the corresponding graph

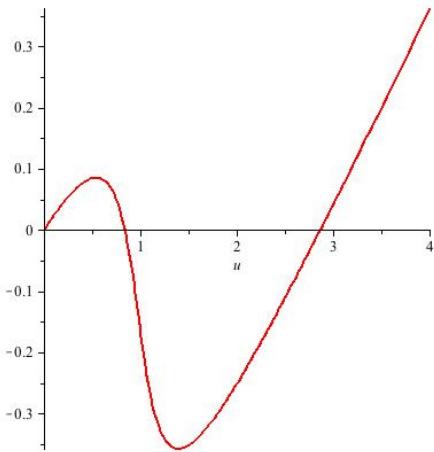


Figure 11.8:  $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ : the graph of  $s$  versus  $u$ .

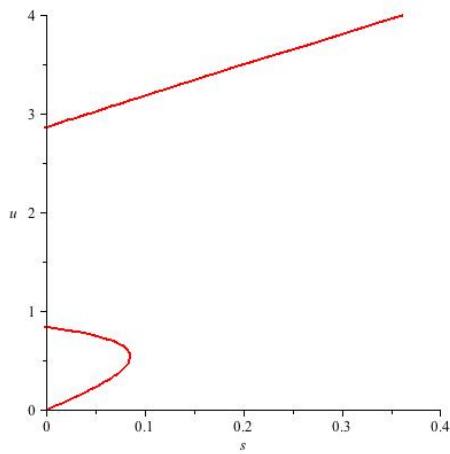
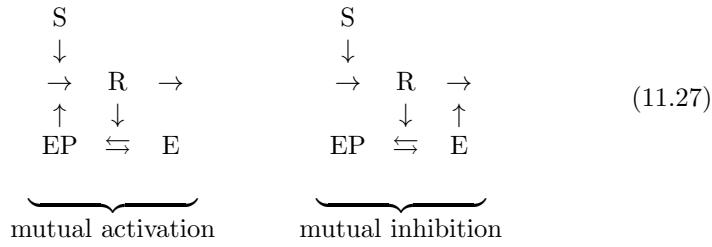


Figure 11.9: On one-way switch ( $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ ): the graph of  $u$  versus  $s$ .

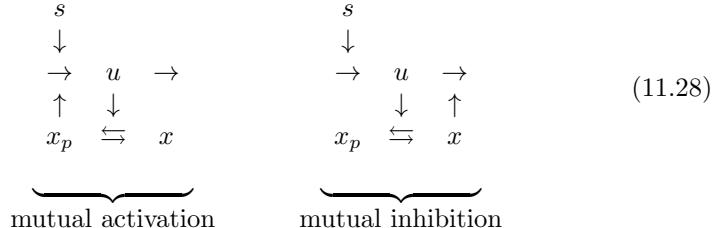
tells you for each  $u$  which signal  $s$  you need to have  $u$  as a steady state, see Fig. 11.8. Note though that  $s$  cannot be negative. Interchanging  $s$  and  $u$  and omitting  $s < 0$  from the diagram we obtain a bifurcation diagram which is called a one-way switch, see Fig. 11.9. Explain this in detail by examining what happens if you start from  $s = 0$  and turn  $s$  up, and then turn  $s$  down again. Indicate in the diagram which states are stable and which states are unstable.

### 11.2.5 A two-way switch

Whereas Section 11.2.4 combines a GK-block and a signal-response block in which EP and R are mutual activating, we can also combine them as mutual inhibiting through R and E. In terms of S, R, E and EP the difference is clear if we line up the diagrams:



In dimensionless form, with  $x$ ,  $x_p$ ,  $u$  and  $s$ :



Compared to the single signal-response block, which has

$$v_s = k_0 + k_1 S, \quad v_d = k_2 R,$$

the change is in  $v_d$ . This degradation rate becomes

$$v_d = k_2 R + k'_2 E_{ss}(R) R = (k_2 + k'_2 E_{ss}(R)) R,$$

in which the new term is  $k'_2 ER$ , with  $E$  set equal to its equilibrium state  $E_{ss}(R)$ . Drawing  $v_s$  en  $v_d$  in one diagram, with varying values of  $S$ , you see that the equilibria of

$$\frac{dR}{dt} = v_s - v_d = k_0 + k_1 S - k_2 R - k'_2 E_{ss}(R) R \tag{11.29}$$

are given by

$$k_2 + k'_2 E_{ss}(R) = R(k_0 + k_1 S). \tag{11.30}$$

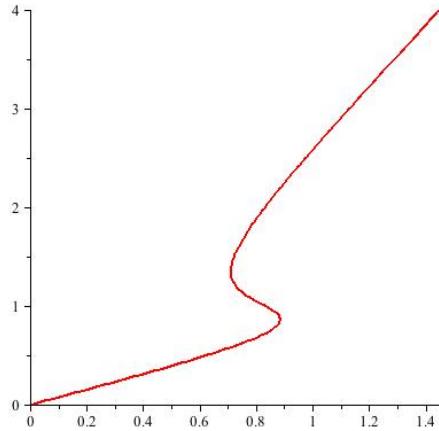


Figure 11.10: Two-way switch ( $K = J = \frac{1}{10}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ ): the graph of  $s$  versus  $u$ .

Verify that, with

$$s = k_3 \frac{k_0 + k_1 S}{k'_2 E_T}, \quad k = \frac{k_2}{k'_2 E_T}, \quad \tau = k'_2 E_T t.$$

you get

$$\frac{du}{d\tau} = s - ux(u) - ku,$$

as the ODE for  $u = k_3 R$ . Conclude that the bifurcation diagram is given by

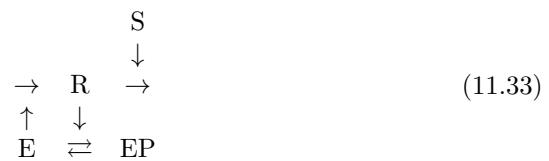
$$s = ux(u) + ku. \quad (11.31)$$

$$s = ux(u) + ku. \quad (11.32)$$

Explain why the bifurcation diagram in Fig. 11.10 is called a two-way switch. Indicate again which solutions are stable and which are unstable. The  $s$  window where multiple steady states co-exist can be quite large, see Fig. 11.11.

### 11.2.6 Homeostasis

Another way to combine a GK-block with a signal response block is



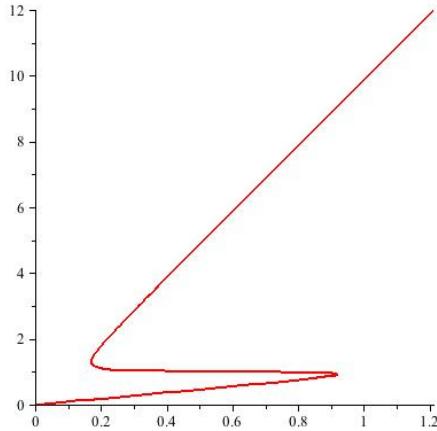


Figure 11.11: Two-way switch ( $K = J = \frac{1}{100}$ ,  $v = 1$ ,  $k = \frac{1}{10}$ ): the graph of  $s$  versus  $u$ .

The differential equation is

$$\frac{dR}{dt} = k_0 E(R) - k_2 S R, \quad (11.34)$$

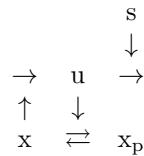
and can be rewritten as

$$\frac{du}{d\tau} = x(u) - su,$$

for  $u = k_3 R$  with

$$s = \frac{k_2 S}{k_0 k_3 E_T}, \quad \tau = k_0 k_3 E_T t,$$

corresponding to



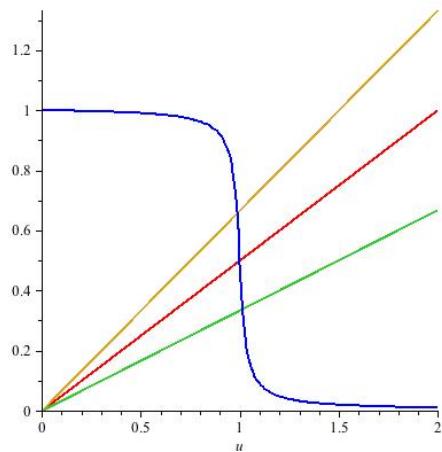


Figure 11.12: Homestasis ( $K = J = \frac{1}{100}$ ,  $v = 1$ ): A synthesis rate with different linear degradation rates.

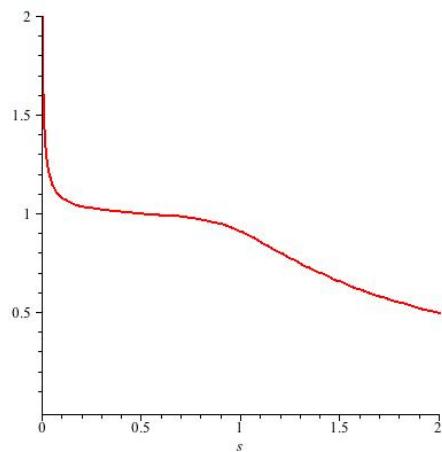


Figure 11.13: Homestasis ( $K = J = \frac{1}{100}$ ,  $v = 1$ ,  $k = \frac{1}{3}$ ): the graph of  $s$  versus  $u$ .

### 11.3 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.

### 11.4 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,

$$\begin{aligned}\frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= g(x, y)\end{aligned}\tag{11.35}$$

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<sup>5</sup>,



We consider the concentration of  $A$  and  $B$  fixed and the mass balances for  $X$  and  $Y$  are now given by,

$$\begin{aligned}\frac{dx}{dt} &= k_1^+ a - k_1^- x + k_3 x^2 y = f(x, y) \\ \frac{dy}{dt} &= k_2 b - k_3 x^2 y = g(x, y)\end{aligned}\tag{11.37}$$

---

<sup>5</sup>This section follows closely a section in *Mathematical Biology* by J D Murray

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,

$$\begin{aligned} 0 &= f(x_s, y_s) \\ 0 &= g(x_s, y_s) \end{aligned} \quad (11.38)$$

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 11.37 to zero allows us to solve for the steady state concentration of  $X$  and  $Y$ , those are (I used Mathematica to find those),

$$\begin{aligned} x_s &= \frac{ak_1^+ + bk_2}{k_1^-} \\ y_s &= \frac{k_2 b (k_1^-)^2}{(ak_1^+ + bk_2)^2 k_3} \end{aligned} \quad (11.39)$$

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  $\delta t$  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. It 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.<sup>6</sup>

---

<sup>6</sup>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. ☺

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 11.37 to zero and solve for  $y$ . We then obtain for the nullclines,

$$\begin{aligned} y = f^{-1}(x) &= \frac{k_1^- x - ak_1^+}{k_3 x^2} \\ y = g^{-1}(x) &= \frac{bk_2}{k_3 x^2} \end{aligned} \quad (11.40)$$

As a sanity check we can determine the intersection point, where the system is at steady state,

$$\begin{aligned} \frac{k_1^- x_s - ak_1^+}{k_3 x_s^2} &= \frac{bk_2}{k_3 x_s^2} \Rightarrow \\ k_1^- x_s - ak_1^+ &= bk_2 \Rightarrow \\ x_s &= \frac{bk_2 + ak_1^+}{k_1^-} \end{aligned} \quad (11.41)$$

And indeed this agrees with our earlier findings.

In figure 11.14 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 is 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 11.15 indicates that the kinetic parameters  $k_1^+$  is an interesting parameter to study bifurcations. In figure 11.16 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.

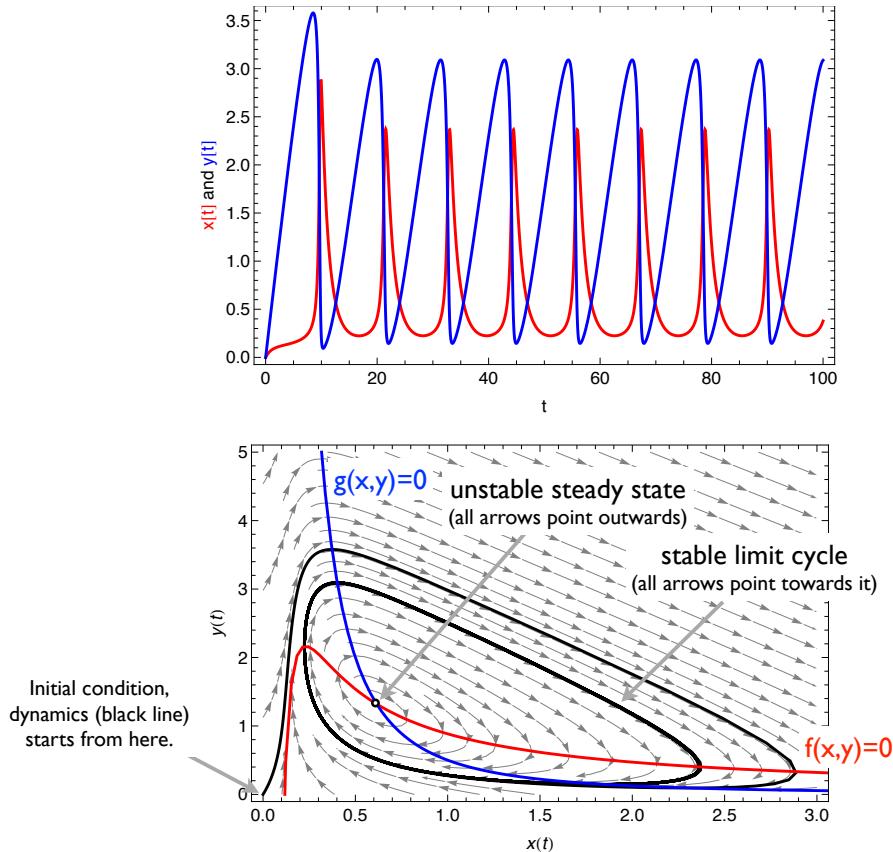


Figure 11.14: **Explanation of phase plane analysis.** We consider the dynamics of the molecules  $X$  and  $Y$  that engage in the chemical reactions given in equation 11.36. We choose  $k_1^+ a = \frac{1}{5\sqrt{3}}$ ,  $k_1^- = 1$ ,  $k_2 = 0.5$  and  $k_3 = 1$  to obtain figures A and B. In A the 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  $dy/dt = 0$ .

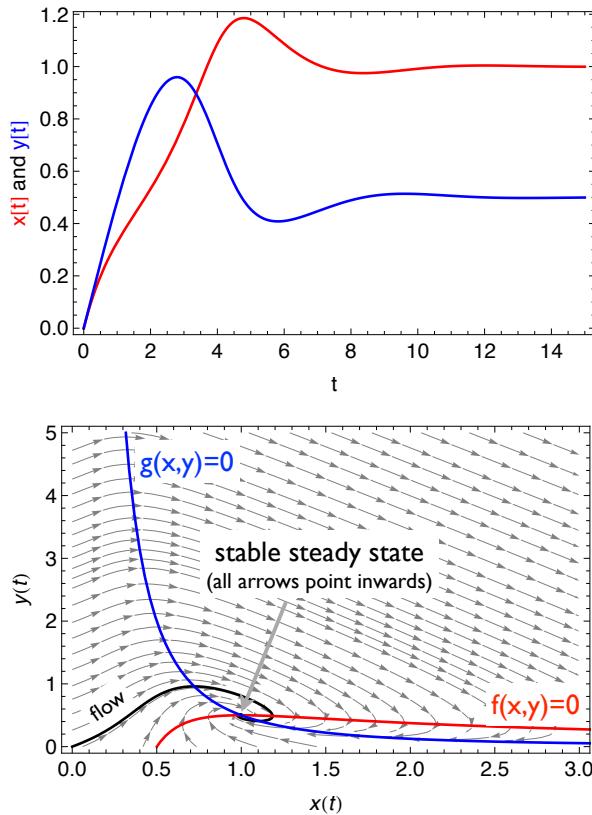


Figure 11.15: **Disappearance of oscillations upon parameter change.** A change in kinetic parameter  $ak_1^+$  to 0.5 leads to disappearance of oscillatory behavior and the steady state become stable. The black line indicates the dynamics ("flow") of the system as shown in the upper figure.

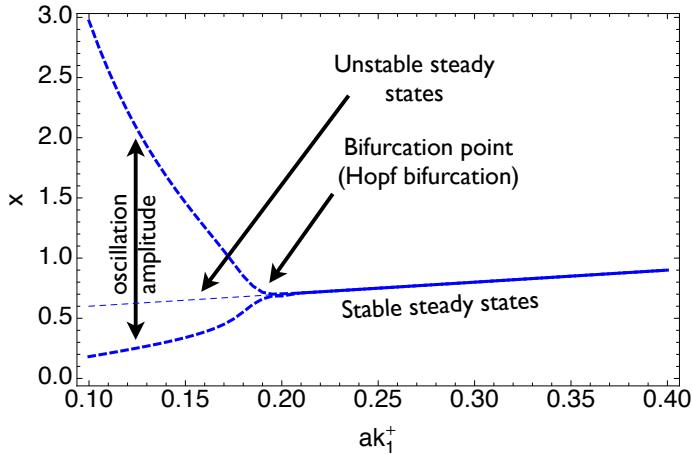


Figure 11.16: **Bifurcation diagram.** The kinetic parameter  $ak_1^+$  was varied and the stability of the steady states were checked. Below a critical value of this parameter oscillations were found and above this bifurcation point stable (non-oscillating states were found). The thick-dashed lines indicate the amplitude of the oscillations. A similar figure can be obtained for  $y$  as function of  $ak_1^+$ .

## 11.5 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 otherwise 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,

$$\begin{aligned} \frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= g(x, y) \end{aligned} \quad (11.42)$$

The steady state is defined as the combination of concentrations  $(x_s, y_s)$  such that,

$$\begin{aligned} 0 &= f(x_s, y_s) \\ 0 &= g(x_s, y_s) \end{aligned} \quad (11.43)$$

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,

$$\begin{aligned}\frac{dx_s}{dt} + \frac{d\delta x}{dt} &= f(x_s, y_s) + \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y \\ \frac{dy_s}{dt} + \frac{d\delta y}{dt} &= g(x_s, y_s) + \frac{\partial g}{\partial x} \delta x + \frac{\partial g}{\partial y} \delta y\end{aligned}\quad (11.44)$$

All these derivatives are evaluated at  $x = x_s$  and  $y = y_s$ . This set of equations can be simplified to,

$$\begin{aligned}\frac{d\delta x}{dt} &= \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y \\ \frac{d\delta y}{dt} &= \frac{\partial g}{\partial x} \delta x + \frac{\partial g}{\partial y} \delta y\end{aligned}\quad (11.45)$$

This set of equation is often written in terms of vectors and matrices,

$$\left( \begin{array}{c} \frac{d\delta x}{dt} \\ \frac{d\delta y}{dt} \end{array} \right) = \underbrace{\left( \begin{array}{cc} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{array} \right)}_{\text{Jacobian matrix}} \cdot \left( \begin{array}{c} \delta x \\ \delta y \end{array} \right)\quad (11.46)$$

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):

$$\begin{aligned}\frac{dx}{dt} &= k_1^+ a - k_1^- x + k_3 x^2 y \\ \frac{dy}{dt} &= k_2 b - k_3 x^2 y\end{aligned}\quad (11.47)$$

(d) The Brusselator:

$$\begin{aligned}\frac{dx}{dt} &= k_1 a + k_2 x^2 \cdot y - k_3 b \cdot x - k_4 x \\ \frac{dy}{dt} &= k_3 b \cdot x - k_2 x^2 y\end{aligned}\quad (11.48)$$

(e) Determine the steady state of the Brusselator model in terms of kinetic parameters.

### 11.5.1 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,

$$\begin{aligned}\text{trace : } T &= \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} \\ \text{determinant : } D &= \frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}\end{aligned}\quad (11.49)$$

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),

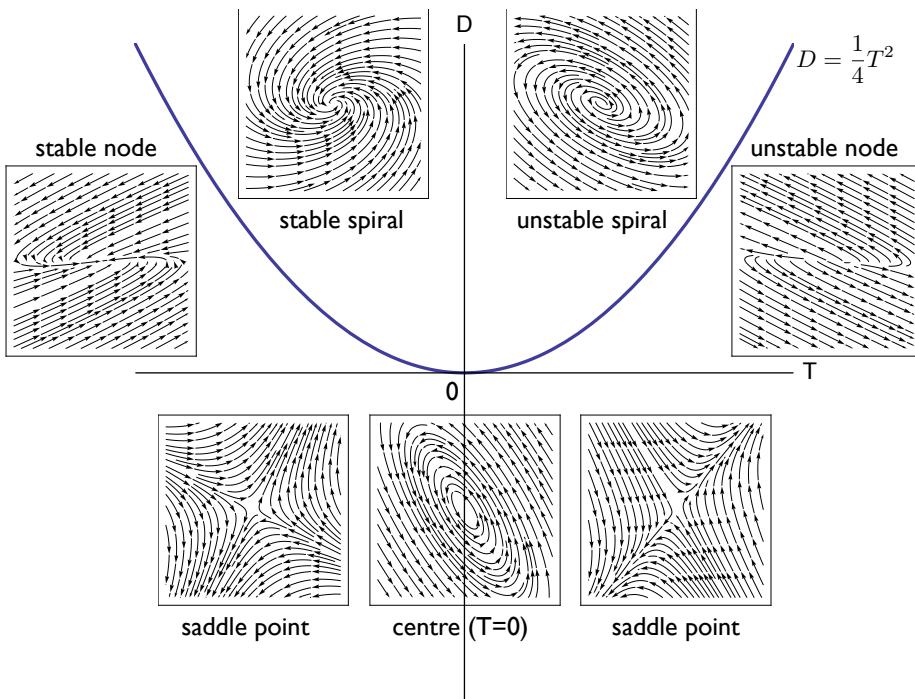
$$\begin{aligned}\lambda_1 &= \frac{1}{2}(T + \sqrt{(T^2 - 4D)}) \\ \lambda_2 &= \frac{1}{2}(T - \sqrt{(T^2 - 4D)})\end{aligned}\quad (11.50)$$

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 11.17). 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 11.17 by working out an example<sup>7</sup>. The

---

<sup>7</sup>Here we follow the appendix in *Mathematical models in molecular and cellular biology* edited by L.A. Segel.



**Figure 11.17: 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.

example is given by,

$$\begin{aligned}\frac{dx}{dt} &= x(x(1-x) - y) = f(x, y) \\ \frac{dy}{dt} &= k(x - 1/\mu)y = g(x, y)\end{aligned}\quad (11.51)$$

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 11.18.

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,

$$\mathbf{M} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x - 3x^2 - y & -x \\ ky & k(x - \mu^{-1}) \end{pmatrix} \quad (11.52)$$

We choose  $k = 1$  and  $\mu = 1.8$ . Evaluation of the jacobian matrix at the three steady states gives:

$$\begin{aligned}(0, 0) \Rightarrow \mathbf{M} &= \begin{pmatrix} 0 & 0 \\ 0 & -0.56 \end{pmatrix} \Rightarrow D = 0, T = -0.56 \\ (0.56, 0.25) \Rightarrow \mathbf{M} &= \begin{pmatrix} -0.062 & -0.56 \\ 0.25 & 0 \end{pmatrix} \Rightarrow D = 0.13, T = -0.062 \\ (1, 0) \Rightarrow \mathbf{M} &= \begin{pmatrix} -1 & -10 \\ 0 & 0.44 \end{pmatrix} \Rightarrow D = -0.44, T = -0.56\end{aligned}\quad (11.53)$$

Inspection of figure 11.17 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.

### 11.5.2 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

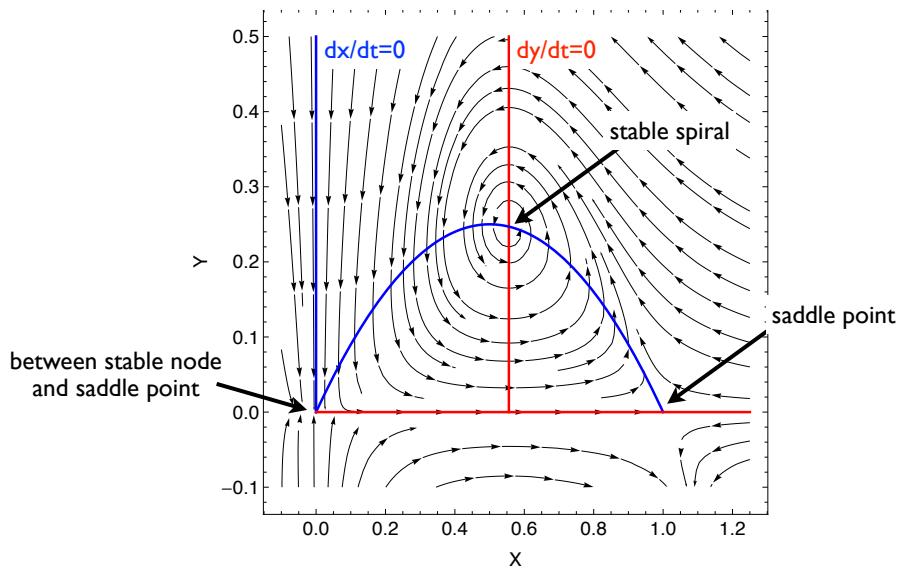


Figure 11.18: **Stability analysis of a complicated example system.** The blue and red lines correspond to the nullcline that results from  $dy/dt = 0$  and  $dx/dt = 0$ . Three intersections are found between the nullclines so three steady states occur. The arrows already hint at the nature of those 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 11.17.

- (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.

## Chapter 12

# Introduction to metabolic control analysis

### 12.1 Introduction

Cells adapt their physiological strategy upon environmental changes. This may involve network rewiring, changes in rates of metabolic reactions, 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 subsequent processing by signaling, metabolic and gene networks.

Remarkably, to some environmental influences cells do not respond at all, even though some of their processes are sensitive to those disturbances, whereas to others they respond sensitively. How can cells achieve perturbation-specific sensitivity and robustness? We shall see that 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.

Another central question is what cells aim to achieve when they adjust themselves. Evolutionary reasoning can help here. One would expect cellular adaptation – self-regulation is meant here and not genetic mutations – to be in the direction of improved performance, i.e. fitness. So, when you have an hypothesis about the fitness objective of a cellular subsystem you can appreciate the cellular adjustments in enzyme concentrations from this perspective. In this way, it can be investigated to what extent cells are capable of optimising their own performance by exploiting gene networks that sense deviations from optimal performance and aim to minimise those deviations.

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. Thus some

## 210CHAPTER 12. INTRODUCTION TO METABOLIC CONTROL ANALYSIS

concentrations have to be 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, which means that the benefit of a physiological adaptation should at least compensate for the cost. Another complication are trade-offs. A trade-offs occurs when a network mutation enhances a specific performance, while at the same time causing a reduction of another performance. Thus, cells are continuously tinkered by natural selection to achieve sophisticated constraint multi-objective control and optimization tasks.

These reflections lead to several questions,

1. How can we change specific behaviour of molecular networks, using changes in enzyme levels? Which enzymes levels should be changed and by what extent?
2. What is the function allosteric regulation in molecular networks?
3. How can homeostasis, robustness, sensitivity, and optimisation of metabolic performance be achieved?
4. What are the limits of gene networks with respect to steering molecular networks?
5. Can we perceive gene networks as systems that try to optimise performance of the system they control?

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 networks (ultra-)sensitive and robust to changes in their environment

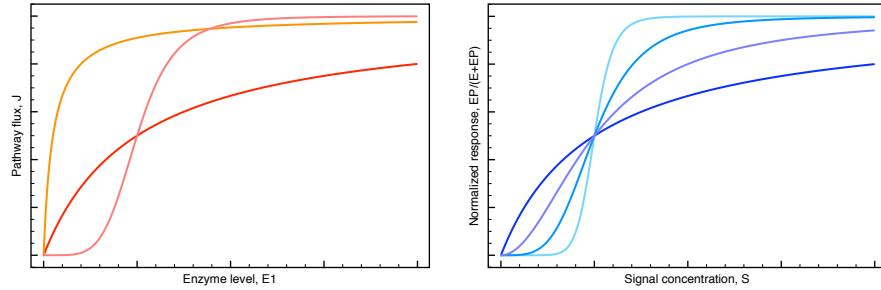


Figure 12.1: **Different responses of steady-state system properties of a molecular network, i.e. a metabolic flux ( $J$ ) and the covalent modification fraction of a protein  $EP/(E+EP)$ , as function of a physiological parameter, such as an enzyme or signal level.** Networks can generate different input/output characteristics depending on their structure and parameterizations. 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.

## 12.2 Cellular self-regulation

A free-living cell is an autonomous unit that given extracellular information, its intracellular state, and its regulatory programs – encoded on its genome and shaped by natural selection – adapts itself to its environment to sustain its fitness and outcompete competitors. In Figure 12.2 the basic design of metabolic self regulation is shown. Gene regulation, transcription, and translation leads to changes of enzyme concentrations under the influence of metabolite-binding transcription factors. So, upon a change in the environment, e.g. the concentration of a nutrient, product or some toxin, enzyme rates are changed leading to changes in the concentration of the metabolites of  $X$  and  $Z$  that then cause changes in gene regulation.

The kinetic models discussed in the previous allow for a complete description of the self-regulation process in terms of underlying biochemical kinetics,

1. **Transcription factor activation** Let's for instance consider the activation of the transcription by  $X$ . We have to consider binding, conformation change, and that only one conformation is active. This means we would arrive at,

$$t_{1x} = t_{1,T} \frac{\frac{x}{K_{R,X}}}{1 + \frac{x}{K_{R,X}} + L \left(1 + \frac{x}{K_{T,X}}\right)} \quad (12.1)$$

Here we used a simple MWC model, with  $L$  as the conformation equilib-

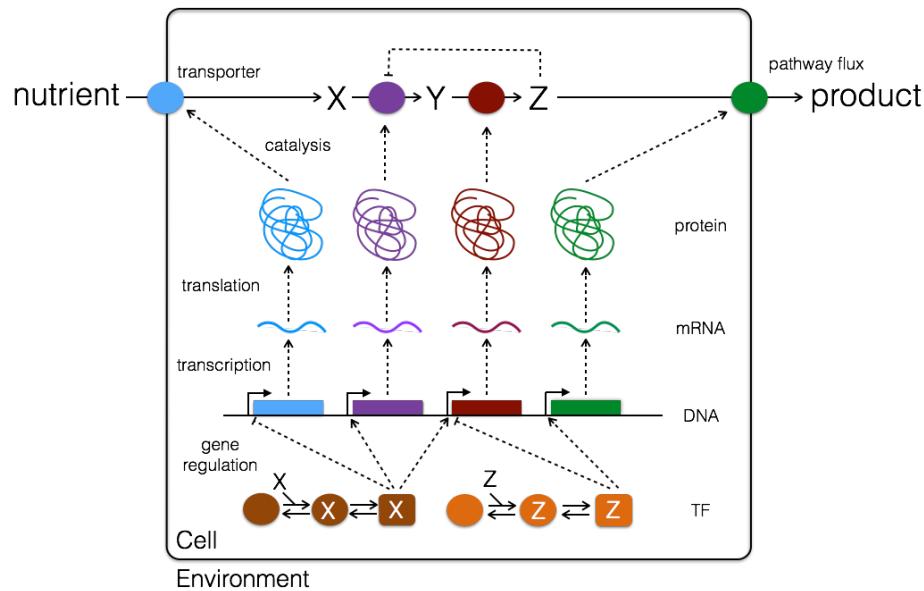


Figure 12.2: **An overview of the gene control of a metabolic pathway by transcription factors that are regulated by metabolic intermediates: “metabolism self-regulation”.** All four proteins occurring in metabolism, two transporters and two enzymes, are synthesised from their mRNA. The mRNA results from transcription regulation by two transcription factors that each bind a different metabolite, either  $X$  or  $Z$ . Note that we have added the degradation of the mRNA and the proteins. The synthesis and degradation of the transcription factors have also been omitted.

rium constant, the  $K$ 's as dissociation constants for the two conformations, and  $t_{1,T}$  as the total transcription factor concentration. We would arrive at a similar relation for the second transcription factor.

2. **Transcription rate regulation** We have four promoters, so four transcription rates. We consider the third promoter that is regulated by  $t_{1x}$  and  $t_{2z}$ . The former activated and the latter inhibits. For instance, we could therefore arrive at the following rate equation for transcription of gene 3,

$$v_3^t = k_3 \frac{\frac{t_{1x}}{K_1}}{1 + \frac{t_{1x}}{K_1} + \frac{t_{2z}}{K_2}} \quad (12.2)$$

The apparent rate constant  $k_3$  depends on the concentration of RNA polymerase, availability of nucleotides and ATP, and auxiliary proteins, such as a sigma factors if prokaryotes are considered or TBP if yeast is considered. So, the rate of change in the corresponding mRNA concentration,

$m_3$ , would become,

$$\frac{dm_3}{dt} = v_3^t - k_{d,3}^m m_3 \quad (12.3)$$

with  $k_{d,3}^m$  as the degradation constant of the mRNA.

3. **Translation** Translation requires ribosomes, tRNAs, and amino acids. When we assume those to occur in excess then the differential equation describes the protein concentration becomes quite straightforward,

$$\frac{de_3}{dt} = k_{t,3} m_3 - k_{d,3}^e e_3 \quad (12.4)$$

4. **Enzymatic catalysis** When the enzyme is produced and folded it will carry out its task: the catalysis of reaction 3. The rate of this reaction would for instance be given by a reversible Michaelis-Menten equation,

$$v_3 = k_{cat,3} e_3 \frac{\frac{y}{K_y} \left(1 - \frac{z}{y \cdot K_{eq}}\right)}{1 + \frac{y}{K_y} + \frac{z}{K_z}} \quad (12.5)$$

5. **Steady-state pathway flux** Eventually, all this regulation leads to a steady state such that,

$$J = v_1(e_1, s_s, x_s) = v_2(e_2, x_s, y_s) = v_3(e_3, y_s, z_s) = v_4(e_4, z_s, p) \quad (12.6)$$

The nutrient concentration is denoted by  $s$  and that of the product by  $p$ . The subscript "s" denotes the steady state consternation of the metabolites. When a small change in the concentration of the nutrient is made that the change in flux becomes,

$$dJ = \frac{\partial v_1}{\partial e_1} de_1 + \frac{\partial v_1}{\partial x} dx + \frac{\partial v_1}{\partial s} ds \quad (12.7)$$

(Identical equations apply for the changes in the rate of the remaining reactions.) Rearranging the previous equation leads to a useful view,

$$1 = \underbrace{\frac{\partial v_1}{\partial e_1} \frac{de_1}{dJ}}_{\text{Gene regulation}} + \underbrace{\frac{\partial v_1}{\partial x} \frac{dx}{dJ} + \frac{\partial v_1}{\partial s} \frac{ds}{dJ}}_{\text{Metabolic regulation}} \quad (12.8)$$

The gene and metabolic regulation term indicate the importance of those two modes of regulation for the observed change in flux.

So far, the description of the mechanistic aspect of self-regulation that answers the "How" question. However, the "Why" question is not addressed. Why does the system work like this? Why is this self-regulation parameterisation beneficial for the cell? What does the cell aim to achieve with this mode of regulation? What is the fitness objective of the cell? These aspects of understanding the system is very different from the more reductionistic biochemical view on self regulation, and in many ways they are also more interesting. They require however a completely different approach than biochemistry.

### 12.3 System properties

Cells change their physiological states upon external stimuli. Changes in pathway fluxes, covalent modification levels of signaling proteins, rates of gene expression are all responses involved in adaptive behavior. Each of these system properties respond in a specific manner to a specific stimulus. Figure 12.1 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 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 analytically – as one equation – in most cases.

Consider for instance the simplest pathway imaginable to illustrate some of the concepts behind MCA (network A in figure 12.3),



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,

$$\frac{d}{dt}x = v_1 - v_2 = k_1^+s - k_1^-x - k_2^+x + k_2^-p \quad (12.10)$$

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

$$x_s = \frac{k_1^+s + k_2^-p}{k_1^- + k_2^+} \quad (12.11)$$

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

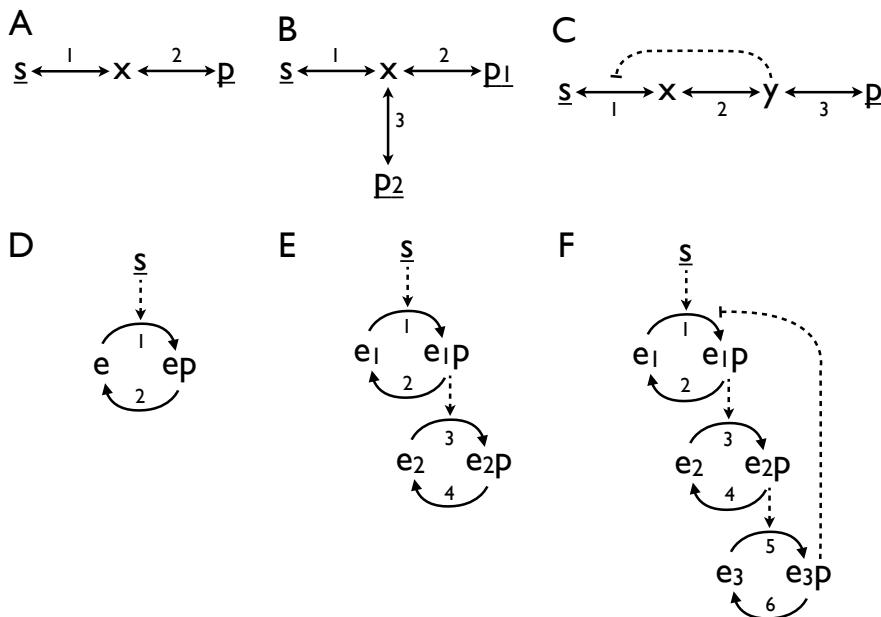
$$x_s = x_s(k_1^+, k_1^-, s, k_2^+, k_2^-, p), \quad (12.12)$$

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 12.11. 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.<sup>1</sup> Do not let the name "metabolic" control analysis fool you; it's application is by now far more broad than its 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 12.3: 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.

<sup>1</sup>A related theory is biochemical systems theory developed by Michael Savageau.

## 12.4 MCA the basics

Before we dive into MCA, we will first give you a flavour of the basics of MCA. MCA aims to relate enzyme behaviour and molecular network behaviour. At the network level, properties emerge that derive from the concerted activity of all the molecules in the system. Examples are the steady-state fluxes, concentrations of molecules, time-scales, and bifurcations. Ideally we would like to understand how individual molecules, such as regulatory proteins and enzymes, determine those properties but, since all molecules are partially determining these properties, this is not a trivial task. MCA takes the following approach. It distinguishes local, enzymatic properties called elasticity coefficients defined at the level of individual reactions and global, systemic properties called control and response coefficients defined at the level of the entire molecular network. The local properties can be determined in isolation of the entire system by studying an enzyme in isolation. This is not possible for the systemic properties as those require the consideration of the entire system. The aim of MCA is to express the global properties in terms of the local properties using a few basic relationships deriving from a mathematical model of a biochemical reaction network in terms of mass balances.

### 12.4.1 Response of a metabolic pathway to a change in an enzyme concentration

To become familiar with the concepts of MCA it is instructive to consider how a metabolic pathway responds to a change in the concentration of one of its enzymes. We consider a simple pathway of two enzymes,



This system we model simply as,

$$\frac{dx}{dt} = \underbrace{0.55 - 0.08x}_{v_1} - \underbrace{e_2 \frac{x}{1+x}}_{v_2} \quad (12.14)$$

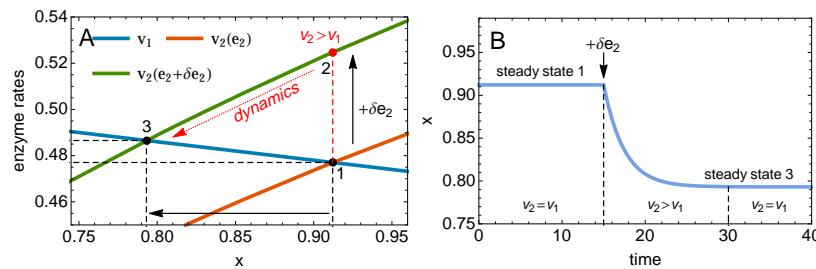
We start at a steady state at  $e_2 = 1$  and perturb this concentration by 10% to obtain a new enzyme concentration of  $e_2 = 1.1$ . The metabolic pathway will respond to this perturbation, display dynamics, and settle to a new steady state. This shown in Figure 12.4. The starts at steady state 1 where  $v_1 = v_2$  and  $e_2 = 1$  then  $e_2 \rightarrow 1.1$  and  $v_2 > v_1$  (state 2 in Figure 12.4A). Because  $v_2 > v_1$ ,  $x$  will drop such that  $v_1$  increases and  $v_2$  decreases. This continues until a new steady state is reached, state 3 in Figure 12.4, where again  $v_1 = v_2$  but now at  $e_2 = 1.1$ . In the new steady state, the flux has increased from  $\approx 0.47$  to  $0.486$  and the steady state concentration of  $X$  has dropped from  $0.912$  to  $0.792$ . So, a change of 10% in the enzyme concentration gave a % change in the flux of  $(0.486 - 0.47)/0.47 \times 100\% \approx 3\%$  and in the concentration

of  $(0.792 - 0.912)/0.912 \times 100\% \approx -13\%$ . So,

$$\frac{\% \text{ change in flux}}{\% \text{ change in enzyme}} = \frac{3\%}{10\%} = 0.3 \quad (12.15)$$

$$\frac{\% \text{ change in concentration}}{\% \text{ change in enzyme}} = \frac{-13\%}{10\%} = -1.3 \quad (12.16)$$

MCA expresses such relative changes in fluxes and concentration in terms of enzyme properties. Figure 12.4 already suggests which properties of the enzymes are important to determine those changes: it has to do with the slopes of the enzyme rates as function of  $x$ . For instance, if the blue line would have been steeper than enzyme 1 would have been more strongly inhibited and the change in the flux would have been greater while the change in the concentration would have been smaller. If, the blue line would have been flat then no change in the flux would have occurred, only in the concentration of  $X$ . MCA quantifies those slopes with elasticity coefficients, which will be explained in the next section.



**Figure 12.4: Dynamic response of a metabolic pathway to a change in one of its enzyme concentrations.** Rate characteristics of the enzymes (A). When the system is at steady state 1, additional enzyme 2 is added to the system. This causes  $v_2$  to exceed  $v_1$  (state 2 (red dot) in A) and as a result the concentration of  $X$  (B) will drop until  $v_1 = v_2$  again (state 3 in A). Note that  $v_1$  increases and  $v_2$  decreases when  $X$  drops in concentration (A).

Finally, MCA always uses normalised responses – akin to percentage changes. The advantages of this becomes clear with an example. Say the flux change is 0.1 mM/min. Only when the steady state at which this occurs has a small enough flux will this change be significant: if the flux is 1 mM/min then a 10% has occurred but if it is 100 mM/min then only 0.1%. In the latter case, the change is negligible. So, relative changes are the appropriate measure for quantification of system responses.

#### 12.4.2 Elasticity coefficients: local properties defined at the level of single reactions

An elasticity coefficient is the normalised sensitivity of the rate of a reaction, be it spontaneous or enzyme-catalysed, for a reactant or effector supplied at a

particular concentration; for reaction  $j$  and molecule concentration  $x_i$  we have for the elasticity coefficient,

$$\epsilon_{x_i}^{v_j} = \frac{\partial v_j}{\partial x_i} \frac{x_i}{v_j} = \frac{\partial \ln v_j}{\partial \ln x_i} \quad (12.17)$$

Here we have used the relationship  $d \ln x = \frac{1}{x} dx$ ; hence  $\frac{1}{x} = \frac{\partial \ln x}{\partial x}$ , which follows from the definition of a derivative, i.e.  $df(x) = \frac{\partial f(x)}{\partial x} dx$ . Note that this coefficient is defined – and measurable – when the reaction occurs at a certain rate, which only requires the reaction in isolation and its reactants (and effectors) supplied at a specific concentration.

For instance, for a reversible mass-action kinetics equation,  $v = k^+ s - k^- p = k^+ s \left(1 - \frac{k^- p}{k^+ s}\right) = k^+ s \left(1 - \frac{p}{K_{eq} \cdot s}\right)$  we obtain,

$$\epsilon_s^v = \frac{\partial v}{\partial s} \frac{s}{v} = \frac{k^+ s}{k^+ s - k^- p} = \frac{1}{1 - \frac{k^- p}{k^+ s}} = \frac{1}{1 - \frac{p}{s \cdot K_{eq}}} \quad (12.18)$$

With  $K_{eq} = \frac{p_e}{s_e} = \frac{k^+}{k^-}$  as the equilibrium constant of the reaction. For the elasticity coefficient towards the product concentration we obtain,

$$\epsilon_p^v = \frac{\partial v}{\partial p} \frac{p}{v} = \frac{-k^- p}{k^+ s - k^- p} = \frac{-\frac{k^- p}{k^+ s}}{1 - \frac{k^- p}{k^+ s}} = \frac{-\frac{p}{K_{eq} \cdot s}}{1 - \frac{p}{K_{eq} \cdot s}} \quad (12.19)$$

Thus, the elasticity coefficient for the substrate is positive – it activates the rate – and the elasticity coefficient of the product is negative – it inhibits the rate. Note that the elasticity coefficients are not constants, they depend on the actual concentrations of the reactants and effectors! Hence, when an enzyme is embedded in some molecular network than its elasticity coefficients vary with the state of the system, i.e. all the concentrations in the network.

For the reversible Michaelis-Menten rate equation,

$$\begin{aligned} v &= \frac{V_M^+ \frac{s}{K_s} - V_M^- \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} = \frac{V_M^+ \frac{s}{K_s} \left(1 - \frac{p}{s \cdot K_{eq}}\right)}{1 + \frac{s}{K_s} + \frac{p}{K_p}} = \frac{f(s, p)}{g(s, p)} \\ f(s, p) &= V_M^+ \frac{s}{K_s} - V_M^- \frac{p}{K_p} \\ g(s, p) &= 1 + \frac{s}{K_s} + \frac{p}{K_p} \end{aligned} \quad (12.20)$$

we obtain for the elasticity coefficient for the substrate,

$$\begin{aligned}
 \epsilon_s^v &= \frac{\partial \ln v}{\partial \ln s} = \frac{\partial \ln \frac{f}{g}}{\partial \ln s} = \frac{\partial \ln f}{\partial \ln s} - \frac{\partial \ln g}{\partial \ln s} = \frac{\partial f}{\partial s} \frac{s}{f} - \frac{\partial g}{\partial s} \frac{s}{g} \\
 &= \frac{V_M^+}{K_s} \frac{s}{V_M^+ \frac{s}{K_s} - V_M^- \frac{p}{K_p}} - \frac{1}{K_s} \frac{s}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \\
 &= \underbrace{\frac{1}{1 - \frac{p}{s \cdot K_{eq}}}}_{\text{thermodynamic part}} - \underbrace{\frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}}}_{\text{kinetic part}}
 \end{aligned} \tag{12.21}$$

We can distinguish the following regimes,

- An irreversible enzyme  $K_{eq} \rightarrow \infty$ : the thermodynamic part becomes equal to 1 and we obtain for the substrate elasticity coefficient,

$$\epsilon_s^v = 1 - \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} = \frac{1 + \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \tag{12.22}$$

Thus the elasticity coefficient decreases with  $s$ ; the enzyme becomes less sensitive for  $S$  when it is more saturated with it. This should make intuitive sense as the dependency of the rate on  $s$  becomes less pronounced – smaller slope  $\frac{\partial v}{\partial s}$  (when  $s$  is large).

- When  $s \gg \frac{p}{K_{eq}}$  the same story applies.
- $p = 0$  then the enzyme functions in an irreversible Michaelis-Menten regime,

$$\epsilon_s^v = \frac{1}{1 + \frac{s}{K_s}} = \frac{K_s}{K_s + s} \tag{12.23}$$

### Exercise

1. Determine the elasticity coefficient of a reversible Michaelis-Menten rate equation for its product.
2. Determine the elasticity coefficient of an irreversible Michaelis-Menten rate equation for its substrate.

#### 12.4.3 Flux control coefficients defined at the level of the entire network

The other set of coefficients of MCA are the systemic coefficients and one example of those are the flux control coefficients defined as,

$$C_{v_i}^{J_k} = \frac{\partial \ln J_k}{\partial \ln v_i} = \frac{\partial \ln J_k}{\partial \ln e_i} \tag{12.24}$$

This coefficient refers to the  $k$ -th steady state flux,  $J_k$ . For a steady state all the enzymes in the network are required; hence this is a systemic property<sup>2</sup>. The flux control coefficient quantifies the fractional change in the flux, i.e.  $d \ln J_k = \frac{dJ_k}{J_k}$ , given a fractional change in the  $i$ -th enzyme activity,  $d \ln v_i$  or, equivalently, in its concentration  $d \ln e_i$ . One could think of a control coefficient in the following manner: the % change in the flux,  $J_k$ , equals  $C_{v_i}^{J_k}$  when the enzyme level,  $e_i$  is change by 1%.

With MCA, control coefficients can be equated in terms of elasticity coefficients using the summation and connectivity theorems of MCA; below we derive them and here we state them for the two-enzyme linear pathway  $S \rightleftharpoons X \rightleftharpoons P$ ,

$$\begin{aligned} \text{summation theorem: } & C_1^J + C_2^J = 1 \\ \text{connectivity theorem: } & C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} = 0 \end{aligned} \quad (12.25)$$

Solving for  $C_1^J$  and  $C_2^J$  gives expressions in terms of elasticity coefficients,

$$\begin{aligned} C_1^J &= \frac{-\epsilon_x^{v_2}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \\ C_2^J &= \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \end{aligned} \quad (12.26)$$

It is convenient to study them in the following manner,

$$\begin{aligned} C_1^J &= 1 - C_2^J \\ C_2^J &= \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} = \frac{1}{1 - \frac{\epsilon_x^{v_2}}{\epsilon_x^{v_1}}} = \frac{1}{1 + \alpha} \\ \alpha &= -\frac{\epsilon_x^{v_2}}{\epsilon_x^{v_1}} \end{aligned} \quad (12.27)$$

The elasticity coefficient ratio  $\alpha = -\frac{\epsilon_x^{v_2}}{\epsilon_x^{v_1}}$  is generally positive because generally  $\epsilon_x^{v_1} < 0$  (product inhibition) and  $\epsilon_x^{v_2} > 0$  (substrate activation). Thus  $0 \leq C_1^J \leq 1$  and  $0 \leq C_2^J \leq 1$ . Figure 12.5 illustrates this type of analysis.

We can distinguish the following regimes:

1. Enzyme 1 has all flux control:  $C_1^J = 1$ ,  $C_2^J = 0$  when  $\alpha \rightarrow \infty$  and thus  $\epsilon_x^{v_1} = 0$ . Thus, enzyme 1 is product insensitive and acts a pump. The flux through the network is now set only by the kinetic properties of enzyme 1 and the concentration of  $S$ ,
2. Enzyme 1 and 2 have equal flux control  $C_1^J = C_2^J = 0.5$ . This occurs when  $\alpha = 1$ ; thus, when enzyme 1 and 2 are equally sensitive to  $x$ . "There is no dictator in a perfect democracy",

---

<sup>2</sup>In contrast to an elasticity coefficient as this quantity can be determined from an experiment on the isolated enzyme.

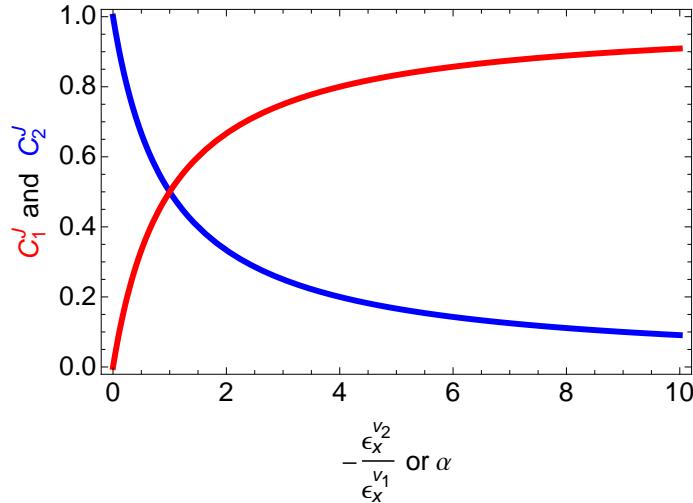


Figure 12.5: **The basics of MCA and flux control illustrated for the system:**  $S \rightleftharpoons X \rightleftharpoons P$ . On the x-axis the ratio of the relevant elasticity coefficients is plotted. This ratio is larger than 1 when enzyme 2 is more sensitive to  $x$  than enzyme 1; then, the plot indicates that the flux control of enzyme 1 is higher than that of enzyme 2. Enzyme 1 has less flux control when enzyme 1 is more sensitive to  $x$  and then the ratio is smaller than 1. When the ratio equals 1 both enzyme share the control equally:  $C_1^J = C_2^J = 0.5$ .

3. Enzyme 2 has all flux control:  $C_1^J = 0$ ,  $C_2^J = 1$  when  $\alpha \rightarrow 0$  and thus  $\epsilon_x^{v_2} = 0$  or  $\epsilon_x^{v_1} \gg \epsilon_x^{v_2}$  (strong feedback inhibition). Thus, enzyme 1 is much more sensitive to  $x$  than enzyme 2. For instance, when enzyme 2 acts in its saturated regime; hence,  $J \approx V_{max,2}$  and  $J \ll V_{max,1}$ ,
  
  
  
  
  
  
4. Flux control is shared between enzyme 1 and 2. The enzyme that has most control is the least sensitive to  $x$ : thus  $C_1^J > C_2^J$  if  $-\epsilon_x^{v_1} < \epsilon_x^{v_2}$  or  $C_2^J > C_1^J$  if  $\epsilon_x^{v_2} < -\epsilon_x^{v_1}$ .

Given the basic biochemical facts that enzymes are reversible and product sensitive, MCA predicts that generally flux control will be distributed. Hence, rate limiting steps are not predicted to occur by MCA; despite the strong intuitive expectations that we – and the scientific literature – sometimes have about those systems.

#### 12.4.4 The rate of an enzyme catalysed reaction, dependency on enzyme concentration and general elasticity equations (skip on a first read)

From enzymology we know that the rate of enzyme has a general structure,

$$\begin{aligned} v &= \overbrace{k_{cat} \cdot e \cdot \prod_i \frac{s_i}{K_{s,i}} \cdot \left(1 - \frac{\prod_k p_k}{\prod_i s_i \cdot K_{eq}}\right)}^{\substack{f(\mathbf{s}, \mathbf{p}) \\ V_{max}^+ \cdot \text{mass-action term}}} \underbrace{g(\mathbf{x})}_{\text{mechanism specific}} \\ &= \left(V_{max}^+ \cdot \prod_i \frac{s_i}{K_{s,i}} - V_{max}^- \cdot \prod_k \frac{p_k}{K_{p,k}}\right) \cdot g(\mathbf{x}) \end{aligned} \quad (12.28)$$

The term  $g(\mathbf{x})$  is enzyme mechanism specific (e.g. ordered, random, ping-pong, equilibrium-binding, quasi-steady state etc.) and may depend in addition to concentration of reactants depend on effector concentrations (collected in the state vector  $\mathbf{x}$ ). Thus, the rate of enzyme catalysed reaction is a linear function of the amount of enzyme  $v = e \cdot \omega$ ; as a consequence,

$$d \ln v = \frac{\partial v}{\partial e} \frac{e}{v} d \ln e = \omega \cdot \frac{e}{e \cdot \omega} d \ln e = d \ln e \quad (12.29)$$

Thus, all the following definitions are identical as was assumed above,

$$d \ln J = C_v^J d \ln v = C_v^J d \ln e = C_e^J d \ln e = C_v^J d \ln e \quad (12.30)$$

The substrate elasticity coefficient also has a general form,

$$\begin{aligned} \epsilon_{s_k}^v &= \frac{\partial \ln f \cdot g}{\partial \ln s_k} = \frac{\partial \ln f}{\partial \ln s_k} + \frac{\partial \ln g}{\partial \ln s_k} = \frac{\partial f}{\partial s_k} \frac{s_k}{f} + \frac{\partial g}{\partial s_k} \frac{s_k}{g} \\ &= \frac{V_{max}^+}{K_{s,k}} \prod_{i, i \neq k} \frac{s_i}{K_{s,i}} \frac{s_k}{V_{max}^+ \cdot \prod_i \frac{s_i}{K_{s,i}} \cdot \left(1 - \frac{\prod_k p_k}{\prod_i s_i \cdot K_{eq}}\right)} + \frac{\partial \ln g}{\partial \ln s_k} \\ &= \frac{1}{\left(1 - \frac{\prod_k p_k}{\prod_i s_i \cdot K_{eq}}\right)} + \frac{\partial \ln g}{\partial \ln s_k} \end{aligned} \quad (12.31)$$

The product elasticity coefficient becomes,

$$\begin{aligned} \epsilon_{p_m}^v &= \frac{\partial f}{\partial p_m} \frac{p_m}{f} + \frac{\partial \ln g}{\partial \ln p_m} \\ &= -\frac{V_{max}^-}{K_{p,m}} \prod_{k, k \neq m} \frac{p_k}{K_{p,k}} \frac{p_m}{V_{max}^- \cdot \prod_i \frac{s_i}{K_{s,i}} - V_{max}^- \cdot \prod_k \frac{p_k}{K_{p,k}}} + \frac{\partial \ln g}{\partial \ln s_k} \\ &= \frac{-V_{max}^- \cdot \prod_k \frac{p_k}{K_{p,k}}}{V_{max}^+ \cdot \prod_i \frac{s_i}{K_{s,i}} - V_{max}^- \cdot \prod_k \frac{p_k}{K_{p,k}}} + \frac{\partial \ln g}{\partial \ln s_k} \\ &= \frac{-\frac{\prod_k p_k}{K_{eq} \prod_i s_i}}{1 - \frac{\prod_k p_k}{K_{eq} \prod_i s_i}} + \frac{\partial \ln g}{\partial \ln s_k} \end{aligned} \quad (12.32)$$

For any effector  $y$  we have

$$\epsilon_y^v = \frac{\partial \ln g}{\partial \ln y} \quad (12.33)$$

## 12.5 The machinery of MCA: response, control and elasticity coefficients

The coefficients of metabolic control analysis will be introduced using network A in figure 12.3 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 12.6. 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 12.6? 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.

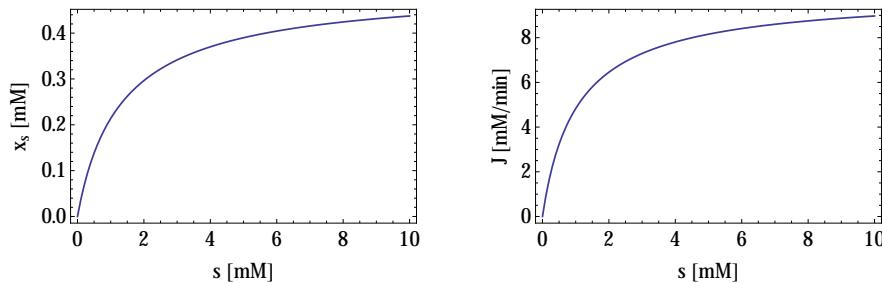


Figure 12.6: The steady state concentration of  $X$  and the steady-state flux  $J$  for network A in figure 12.3 as function of the fixed concentration of the pathway substrate,  $s$ . Both enzyme follow irreversible product-sensitive Michaelis-Menten kinetics with parameters:  $V_{max,1} = 10 \text{ mM min}^{-1}$ ,  $K_{M,1,s} = 1 \text{ mM}$ ,  $K_{M,1,x} = 3 \text{ mM}$ ,  $V_{max,2} = 50 \text{ mM min}^{-1}$ , and  $K_{M,2,x} = 2 \text{ mM}$ . 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 12.6 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 generally consider a response coefficient as,

$$R_s^{x_s} = \frac{\partial \ln x_s}{\partial \ln s} = \frac{s}{x_s} \frac{\partial x_s}{\partial s} \quad (12.34)$$

MCA relates this systemic coefficient to enzyme properties. For notational convenience we will write the steady state concentration of  $x$ ,  $x_s$ , simply as  $x$ . Figure 12.6 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  $s$ ,

$$0 = v_1(s, x(s)) - v_2(x(s)) \quad (12.35)$$

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 12.6, we differentiate the previous equation with respect to  $s$ ,

$$0 = \left( \frac{\partial v_1}{\partial s} + \frac{\partial v_1}{\partial x} \frac{\partial x}{\partial s} - \frac{\partial v_2}{\partial x} \frac{\partial x}{\partial s} \right) ds \quad (12.36)$$

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

$$\begin{aligned} 0 &= \left( \frac{\partial v_1}{\partial s} + \frac{\partial v_1}{\partial x} \frac{\partial x}{\partial s} - \frac{\partial v_2}{\partial x} \frac{\partial x}{\partial s} \right) ds \\ &= \left( \frac{s}{v_1} \frac{\partial v_1}{\partial s} + \frac{x}{v_1} \frac{\partial v_1}{\partial x} \frac{s}{x} \frac{\partial x}{\partial s} - \frac{x}{v_2} \frac{\partial v_2}{\partial x} \frac{s}{x} \frac{\partial x}{\partial s} \right) \frac{ds}{s} \quad (\text{note: } v_1 = v_2 = J) \\ &= \left( \frac{\partial \ln v_1}{\partial \ln s} + \frac{\partial \ln v_1}{\partial \ln x} \frac{\partial \ln x}{\partial \ln s} - \frac{\partial \ln v_2}{\partial \ln x} \frac{\partial \ln x}{\partial \ln s} \right) d \ln s \end{aligned} \quad (12.37)$$

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,  $\frac{\partial \ln v_i}{\partial \ln x}$ , or a fixed external reactant,  $\frac{\partial \ln v_i}{\partial \ln s}$ . 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  $\frac{\partial \ln v_1}{\partial \ln x}$ .<sup>3</sup> 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

---

<sup>3</sup>In some literature, an elasticity coefficient to an external metabolite or fixed concentrations has been termed a 'π' elasticity, we not do this here.

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 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 12.37 can be written in terms of MCA notation,

$$0 = \epsilon_s^{v_1} + \epsilon_x^{v_1} R_s^x - \epsilon_x^{v_2} R_s^x \quad (12.38)$$

and solved for the response coefficient,

$$R_s^x = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \epsilon_s^{v_1} \quad (12.39)$$

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 12.39 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),

$$\frac{\partial \ln x}{\partial \ln v_1} = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \quad (12.40)$$

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,

$$v_1(\lambda) = \lambda v_1 \quad (12.41)$$

at a reference value of  $\lambda = 1$ . At the moment  $\lambda$  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  $\lambda$  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  $\lambda$  parameter is nothing mysterious. We can now determine the effect of the rate of a reaction on system properties through control coefficients. The control coefficient on the steady-state concentration of  $X$  can be obtained by differentiating with respect to  $\lambda$ ,

$$0 = v_1(\lambda, x_s(\lambda)) - v_2(x_s(\lambda)) \quad (12.42)$$

and scaling this equation,

$$0 = \left( \frac{\partial \ln v_1}{\partial \ln \lambda} + \frac{\partial \ln v_1}{\partial \ln x} \frac{\partial \ln x}{\partial \ln \lambda} - \frac{\partial \ln v_2}{\partial \ln x} \frac{\partial \ln x}{\partial \ln \lambda} \right) d \ln \lambda \quad (12.43)$$

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,

$$C_1^x = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \quad (12.44)$$

This is exactly the relationship we were searching for (equation 12.40).

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 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 12.6, set  $s$  to  $2 \text{ mM}$ . 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 12.6. 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.

## 12.6 Control coefficients for a linear pathway

In the previous section we have determined the concentration control coefficient of the first reaction (equation 12.44). The concentration control coefficient for the second reaction can be found in the same manner,

$$C_2^x = \frac{1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \quad (12.45)$$

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,

$$C_1^X + C_2^X = 0 \quad (12.46)$$

The interpretation of equation 12.46 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,

$$\sum_i^r C_i^{X_j} = 0 \quad (12.47)$$

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,

$$d \ln J_k = \frac{\partial \ln J_k}{\partial \ln v_i} d \ln v_i = C_{v_i}^{J_k} d \ln v_i \quad (12.48)$$

We have already concluded that,

$$J_1 = v_1(e_1, x_s(e_1)) \quad (12.49)$$

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,

$$d \ln J = \left( \frac{\partial \ln v_1}{\partial \ln e_1} + \frac{\partial \ln v_1}{\partial \ln x} \frac{\partial \ln x}{\partial \ln e_1} \right) d \ln e_1 \quad (12.50)$$

So we have,

$$C_1^J = \epsilon_{e_1}^{v_1} + \epsilon_x^{v_1} C_1^X = 1 + \epsilon_x^{v_1} C_1^X \quad (12.51)$$

You can easily verify that  $\epsilon_{e_1}^{v_1} = 1$  because  $v_1$  is a linear function of  $e_1$ .<sup>4</sup> We have already determined the concentration control coefficient, the flux control coefficient then becomes,

$$C_1^J = \frac{-\epsilon_x^{v_2}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} = \frac{1}{1 - \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}}} \quad (12.52)$$

---

<sup>4</sup>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.<sup>5</sup> For the flux control coefficient of the second enzyme we obtain,

$$C_2^J = \epsilon_{e_2}^{v_2} + \epsilon_x^{v_2} C_2^X = 1 + \epsilon_x^{v_2} C_2^X = \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \quad (12.53)$$

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

$$C_1^J + C_2^J = 1 \quad (12.54)$$

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  $-\frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}} < 1$ .<sup>6</sup> 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 translates 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  $\alpha$ . So a simultaneous change of the rates by a factor  $\alpha$  causes the flux to change with a factor of  $\alpha$  too. Hence, the summation theorem for flux control coefficients (equation 12.54). The summation theorem for flux control coefficients can be generalized to networks of any complexity,

$$\sum_i^r C_i^{J_k} = 1 \quad (12.55)$$

## Exercises

1. Determine the elasticity coefficient for the reversible Michaelis-Menten mechanism (equation 6.42) with respect to its substrate and product. Write it as a difference between a term that contains the mass-action ratio,  $\Gamma/K_{eq}$  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?

---

<sup>5</sup>Explain why.

<sup>6</sup>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,

$$\begin{aligned} C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} &= 0 \\ C_1^X \epsilon_x^{v_1} + C_2^X \epsilon_x^{v_2} &= -1 \end{aligned} \quad (12.56)$$

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

4. Use the connectivity theorems (equations 15.59) to show that the enzyme, which is sensitive to a metabolite, that only influences this enzyme and no other, 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.

### 12.6.1 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,

$$\sum_k^r C_k^{X_l} \epsilon_{X_j}^{v_k} = -\delta_j^l, \quad \delta_j^l = 1 \text{ if } j = l \text{ else } 0 \quad (12.57)$$

$$\sum_k^r C_k^{J_l} \epsilon_{X_j}^{v_k} = 0 \quad (12.58)$$

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

$$\sum_k^r C_k^{X_l} = 0 \quad (12.59)$$

$$\sum_k^r C_k^{J_l} = 1 \quad (12.60)$$

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

### Exercises

- Determine the flux control coefficients of a 3-enzyme linear pathway without feedback. Do the same for the pathway displayed in 12.3C. What is the effect of the feedback on the flux control coefficients?
- Which enzymes have most flux control in a 3-enzyme linear pathway if the second enzyme is not sensitive to its product?

3. A transient time,  $\tau$ , 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.

## 12.7 Control shifts due to changes in enzyme concentrations

### 12.7.1 Irreversible enzymes generally loose flux control when increased in concentration

Consider an irreversible, product-sensitive enzyme,  $k$ , that is slowly increased in concentration,  $e_k$ . At low concentrations it will control on flux, as it may be amongst the group of enzymes with the lowest  $V_{MAX}$  values. The flux through the enzyme will increase as well as the flux through the enzymes it is directly linked to via its substrates and products. The enzyme(-s) directly preceding the enzyme  $k$  will increase in flux due to relief of product inhibition as enzyme  $k$  can cope with a lower substrate concentration due to its enhanced concentration. The enzyme(-s) that directly follow  $k$  increase in flux because their substrate concentration increases, which is the product of  $k$ . Again, enzyme  $k$  can cope with additional product inhibition as its concentration and the extend of its flux increase will depend on its elasticity coefficients to its substrate and product. At some enzyme concentration, enzyme  $k$  no longer has flux control; it now has the highest  $V_{MAX}$  in the system and is in great excess: thus,  $J/V_{MAX,k} << 1$  and  $J$  essentially remains fixed when  $e_k$  increases and its substrate concentration continues to drop and its product concentration rises. In the limit of close to zero substrate concentration and close to infinity product concentration the elasticity coefficient of the enzyme  $k$  for  $S$  and  $P$  become respectively  $-1$  and  $1$ ; assuming  $v_k = \frac{V_{max,k}s/K_s}{1+s/K_s+p/K_p}$ . And hence, the enzyme before enzyme  $k$  that  $S$  as a product has become insensitive to  $S$ , because this concentration has dropped far below the  $K_M$  of this enzyme. Hence, in the simplest case  $S$  only influences enzyme  $k$  and the connectivity theorem  $C_k^J \epsilon_S^{v_k} = 0$  in dictates that  $C_k^J = 0$ . Also, in more complicated cases enzyme  $k$  will generally have lost all flux control.

### 12.7.2 Reversible enzymes generally loose flux control when increased in concentration and then operate closer to thermodynamic equilibrium

Now consider enzyme  $k$  to be a reversible enzyme, again as a response to an increase in its concentration its substrate concentration will drop and its product concentration will rise to allow for an increase in flux through its neighbouring enzymes. This means that the forward rate of enzyme  $k$  reduces while the backward increases; as a result enzyme  $k$  starts to operate closer to thermody-

namic equilibrium. Its distance from thermodynamic equilibrium is captured by  $1 - \frac{p}{s \cdot K_{eq}}$ , which equals 0 at equilibrium. While  $J/V_{max,k} \rightarrow 0$ , as a result of an enzyme  $k$  increase,  $\frac{p}{s} \rightarrow K_{eq}$ . This means that  $\epsilon_S^{v_k} \rightarrow \infty$  and  $\epsilon_P^{v_k} \rightarrow -\infty$ . The connectivity theorem  $C_{k-1}^J \epsilon_S^{v_{k-1}} + C_k^J \epsilon_S^{v_k} = 0$  then indicates that  $C_k^J \rightarrow 0$  as  $\epsilon_S^{v_{k-1}}$  will be generally be small and negative (provided this enzyme does not operate close to equilibrium) and typically  $0 \leq C_{k-1}^J \leq 1$ . This does mean that  $C_{k-1}^J \rightarrow 0$  because  $C_k^J \epsilon_S^{v_k} \approx 0 \cdot \infty \neq 0$ .

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

### 12.8.1 Control coefficients

Consider figure 12.3B, 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,

$$0 = v_1(e_1, x(e_1)) - v_2(x(e_1)) - v_3(x(e_1)) \quad (12.61)$$

We can take the derivative of this equation to  $e_1$ ,

$$0 = \frac{\partial v_1}{\partial e_1} + \frac{\partial v_1}{\partial x} \frac{\partial x}{\partial e_1} - \frac{\partial v_2}{\partial x} \frac{\partial x}{\partial e_1} - \frac{\partial v_3}{\partial x} \frac{\partial x}{\partial e_1} \quad (12.62)$$

and scale it,

$$0 = v_1 \frac{\partial \ln v_1}{\partial \ln e_1} + v_1 \frac{\partial \ln v_1}{\partial \ln x} \frac{\partial \ln x}{\partial \ln e_1} - v_2 \frac{\partial \ln v_2}{\partial \ln x} \frac{\partial \ln x}{\partial \ln e_1} - v_3 \frac{\partial \ln v_3}{\partial \ln x} \frac{\partial \ln x}{\partial \ln e_1} \quad (12.63)$$

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,

$$C_1^X = \frac{-1}{\epsilon_x^{v_1} - \frac{J_2}{J_1} \epsilon_x^{v_2} - \frac{J_3}{J_1} \epsilon_x^{v_3}} \quad (12.64)$$

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 12.44 if we consider,

$$\frac{J_2}{J_1} \epsilon_x^{v_2} + \frac{J_3}{J_1} \epsilon_x^{v_3} \quad (12.65)$$

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 12.51. Rewriting the result gives,

$$C_1^J = \frac{-\left(\frac{J_2}{J_1}\epsilon_x^{v_2} + \frac{J_3}{J_1}\epsilon_x^{v_3}\right)}{\epsilon_x^{v_1} - \left(\frac{J_2}{J_1}\epsilon_x^{v_2} + \frac{J_3}{J_1}\epsilon_x^{v_3}\right)} \quad (12.66)$$

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.

### 12.8.2 Low-flux branches and branch point control

The flux control coefficient of the second reaction on the third is given by,<sup>7</sup>

$$C_2^{J_3} = \frac{-1}{\frac{J_3}{J_2} + \frac{\epsilon_X^{v_2}}{\epsilon_X^{v_3}} - \frac{J_1\epsilon_X^{v_1}}{J_2\epsilon_X^{v_3}}} \quad (12.67)$$

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  $\approx 1$  for  $x$ , we obtain,

$$C_2^{J_3} \approx \frac{-J_2}{J_3 + J_1|\epsilon_X^{v_1}|} \quad (12.68)$$

You may wonder whether these conditions are exotic. They are for many branches actually quite realistic [41, 29]. Consider a branch point where the branching is regulated by the  $K_M$ 's of the enzyme after the branch point; enzymes 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

$$C_2^{J_3} \approx -J_2/J_3 \quad (12.69)$$

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,

$$C_2^{J_3} \approx \frac{\epsilon_X^{v_3}}{\epsilon_X^{v_2} + |\epsilon_X^{v_1}|} \quad (12.70)$$

---

<sup>7</sup>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.<sup>8</sup> 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 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  $J_3 \ll J_2$ . Under what conditions becomes  $C_2^{J_3} < -1$ ? This phenomenon is called branch-point ultrasensitivity.

## 12.9 A metabolic pathway with negative feedback: homeostasis

We will study in this section network C in figure 12.3. Using the summation and connectivity theorems for flux control,

$$\begin{aligned} C_1^J + C_2^J + C_3^J &= 1 \\ C_1^J \epsilon_{x_1}^{v_1} + C_2^J \epsilon_{x_1}^{v_2} &= 0 \\ C_1^J \epsilon_{x_2}^{v_1} + C_2^J \epsilon_{x_2}^{v_2} + C_3^J \epsilon_{x_2}^{v_3} &= 0 \end{aligned}$$

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,

$$C_3^J = -\frac{\epsilon_{x_1}^{v_1} \epsilon_{x_2}^{v_2} - \epsilon_{x_2}^{v_1} \epsilon_{x_1}^{v_2}}{\epsilon_{x_2}^{v_1} \epsilon_{x_1}^{v_2} - \epsilon_{x_1}^{v_1} \epsilon_{x_2}^{v_2} + \epsilon_{x_1}^{v_1} \epsilon_{x_2}^{v_3} - \epsilon_{x_1}^{v_2} \epsilon_{x_2}^{v_3}} \quad (12.71)$$

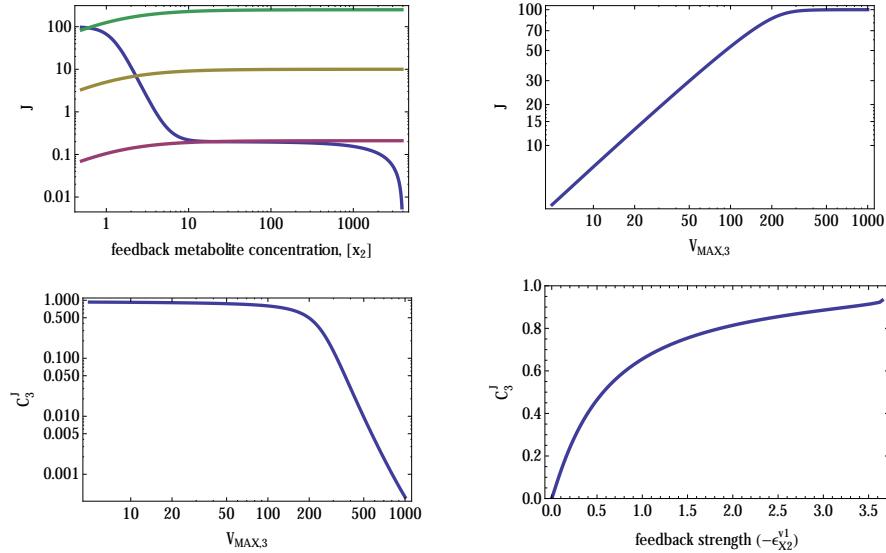
We will analyze the consequences of the feedback strength ( $-\epsilon_{x_2}^{v_1}$ ) in this section. This we do in a more transparent fashion when we consider the product inhibition of the  $x_1$  on enzyme 1 to be negligible,  $\epsilon_{x_1}^{v_1} \approx 0$ ; the flux control coefficient then equals,

$$C_3^J = \frac{\epsilon_{x_2}^{v_1}}{\epsilon_{x_2}^{v_1} - \epsilon_{x_2}^{v_3}} \quad (12.72)$$

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 equipped with negative feedback on the first reaction it is

---

<sup>8</sup>Do you understand why?



**Figure 12.7: Analysis of the consequence of negative feedback for the flux control of metabolic pathways (network C in figure 12.3).** 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.

known that the first reaction is often irreversible and has a low affinity for the product, a high  $K_M$  [20, 19]. 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 [40]. 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

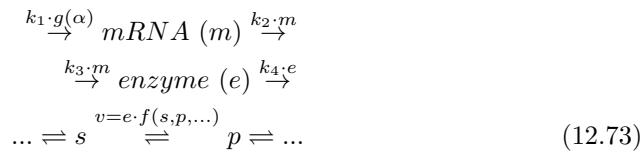
## 12.10. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY; A TRANSCRIPTION AND TRANSLATION CASCADE

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 12.7 where the metabolic pathway displayed in figure 12.3C 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 6.

### 12.10 The central dogma of molecular biology; a transcription and translation cascade

We consider the synthesis of mRNA by transcription, of the associated protein by translation and this protein acts as metabolic enzyme in metabolic pathway,



with  $\alpha$  as the concentration of a transcription factor and  $g(\alpha)$  as a promoter dependent kinetic function. What is the response of mRNA to the TF,  $R_\alpha^m$ ? This equals,

$$R_\alpha^m = \frac{\partial \ln m}{\partial \ln \alpha} = C_1^m \epsilon_\alpha^{v_1} \quad (12.74)$$

The concentration control coefficient of reaction 1 on the concentration of mRNA, i.e.  $C_1^m$ , we can obtain from the summation theorem and connectivity theorem defined for the subsystem,  $\xrightarrow{k_1 \cdot g(\alpha)} \text{mRNA } (m) \xrightarrow{k_2 \cdot m}$ ; since,  $e$ ,  $s$  and  $p$  do not feedback this is correct,

$$\begin{aligned} C_1^m + C_2^m &= 0 \\ C_1^m \epsilon_m^{v_1} + C_2^m \epsilon_m^{v_2} &= -1 \end{aligned} \quad (12.75)$$

Therefore,

$$C_1^m = \frac{1}{\epsilon_m^{v_1} - \epsilon_m^{v_2}} \quad (12.76)$$

Given the kinetics  $v_1 = k_1 g(\alpha)$  and  $v_2 = k_2 m$  this concentration control coefficient simplifies to,

$$C_1^m = 1 = \frac{\partial \ln m}{\partial \ln v_1} = \frac{\partial \ln m}{\partial \ln k_1} \quad (12.77)$$

Hence, the steady state concentration of  $m$  depends linearly on the activity of reaction 1. Indeed at steady state we find  $0 = k_1 g(\alpha) - k_2 m_s$  and thus  $m_s = k_1 g(\alpha)/k_2$ ; hence  $m_s$  is a linear function of  $k_1$ .

The second module,  $\xrightarrow{k_3 \cdot m}$  enzyme ( $e$ )  $\xrightarrow{k_4 \cdot e}$ , has a output  $e$  and as input  $m$ ; so the relevant response coefficient is,

$$R_m^e = \frac{\partial \ln e}{\partial \ln m} = C_3^e \epsilon_m^{v_3} = \frac{-1}{\epsilon_e^{v_3} - \epsilon_e^{v_4}} \epsilon_m^{v_3} = 1 \quad (12.78)$$

This equates to one with the kinetics  $v_3 = k_3 m$ , and  $v_4 = k_4 e$ .

Finally the response of the flux  $J = v = ef(s, p, \dots)$  equals,

$$R_e^J = \frac{\partial \ln J}{\partial \ln e} = C_v^J \epsilon_e^v = C_v^J \quad (12.79)$$

Taken all the results together, we find that the response of the flux to the transcription factor equals,

$$R_\alpha^J = R_e^J R_m^e R_\alpha^m = C_v^J \epsilon_\alpha^{v_1} = C_{\text{activity of metabolic enzyme}, v}^J \cdot \epsilon_\alpha^{v_{\text{transcription}}} \quad (12.80)$$

Thus this simple model predicts that the promoter design, setting  $\epsilon_\alpha^{v_1}$ , and the flux control of the enzyme,  $C_v^J$ , are the only determinants of the steady state flux response to a change in the transcription factor concentration,  $\alpha$ . Note that the last equation derives from the following chain of implicit dependencies  $J = J(e(m(\alpha)))$  and that therefore  $\frac{\partial J}{\partial \alpha} = \frac{\partial J}{\partial e} \frac{\partial e}{\partial m} \frac{\partial m}{\partial \alpha}$ .

## 12.11 Ultrasensitivity of signaling networks

### 12.11.1 A single covalent modification cycle

Figure 12.3D 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 6 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 12.1 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 12.8. 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

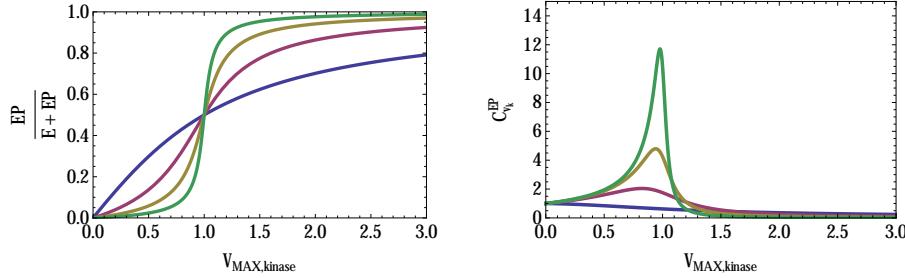


Figure 12.8: **Input-output relationship and the sensitivity of a signaling cycle composed out of a kinase and a phosphatase and another enzyme as substrate (network D in figure 12.3).** The kinetics was modelled with irreversible Michaelis-Menten kinetics ( $K_{M,kinase} = 1$ ,  $K_{M,phosphatase} = 1$ , 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  $c_{signal}^{v_{kinase}} > 1$ ; for instance, when the kinase is a cooperative enzyme.

with the steady state mass balance (we consider product independent kinetics of the kinase and phosphatase),

$$0 = v_k(V_k, e(V_k)) - v_p(ep(V_k)) \quad (12.81)$$

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  $V_k$ ,

$$0 = \frac{\partial v_k}{\partial V_k} + \frac{\partial v_k}{\partial e} \frac{\partial e}{\partial V_k} - \frac{\partial v_p}{\partial ep} \frac{\partial ep}{\partial V_k} \quad (12.82)$$

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

$$0 = \frac{\partial v_k}{\partial V_k} - \frac{\partial v_k}{\partial e} \frac{\partial ep}{\partial V_k} - \frac{\partial v_p}{\partial ep} \frac{\partial ep}{\partial V_k} \quad (12.83)$$

MCA requires the normalization of those derivatives,

$$0 = \frac{\partial \ln v_k}{\partial \ln V_k} - \frac{\partial \ln v_k}{\partial \ln e} \frac{ep}{e} \frac{\partial \ln ep}{\partial \ln V_k} - \frac{\partial \ln v_p}{\partial \ln ep} \frac{\partial \ln ep}{\partial \ln V_k} \quad (12.84)$$

In terms of MCA those normalized derivatives become,

$$0 = \epsilon_{V_k}^{v_k} - \epsilon_e^{v_k} \frac{ep}{e} C_{v_k}^{ep} - \epsilon_{ep}^{v_p} C_{v_k}^{ep} \quad (12.85)$$

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  $EP$ ,

$$C_{v_k}^{ep} = \frac{1}{\epsilon_e^{v_k} \frac{ep}{e} + \epsilon_{ep}^{v_p}} \quad (12.86)$$

This coefficient is the scaled slope  $\partial ep / \partial V_k \cdot V_k / ep$  of the left plot of figure 12.8. 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 12.8 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,

$$R_s^{ep} = C_{v_k}^{ep} \epsilon_s^{v_k} \quad (12.87)$$

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 [27, 14, 3].

### 12.11.2 A signaling cascade: sensitivity amplification

Equation 12.87 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 12.3E). In a cascade the kinase of a signaling cycle is the output, the phosphorylated enzyme, of the previous cycle. So for a cascade of length three we obtain,

$$\begin{aligned} R_s^{e_1 p} &= C_{v_{k,1}}^{e_1 p} \epsilon_s^{v_{k,1}} \\ R_{e_1 p}^{e_2 p} &= C_{v_{k,2}}^{e_2 p} \epsilon_{e_1 p}^{v_{k,2}} \\ R_{e_2 p}^{e_3 p} &= C_{v_{k,3}}^{e_3 p} \epsilon_{e_2 p}^{v_{k,3}} \end{aligned} \quad (12.88)$$

As the kinase of cycle 2 and 3 are the previous phosphorylated enzyme species, i.e.  $e_1 p$  and  $e_2 p$ , and the kinase activity depends linearly on the concentration of those species,  $\epsilon_{e_1 p}^{v_{k,2}} = \epsilon_{e_2 p}^{v_{k,3}} = 1$ . So the sensitivity of the last phosphorylated enzyme,  $e_3 p$ , to  $S$  is given by,

$$R_s^{e_3 p} = R_{e_2 p}^{e_3 p} R_{e_1 p}^{e_2 p} R_s^{e_1 p} \quad (12.89)$$

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, 26].

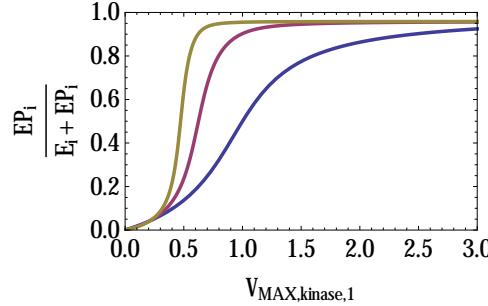


Figure 12.9: **Sensitivity amplification of a signaling cascade composed out of a linear chain of signaling cycles.** Each of the signaling was modeled in the same way, each has the blue input/output characteristic. The output of the second cycle,  $EP_2/(EP_2 + E_2)$  (red line), is more sensitive to 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).

### 12.11.3 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. In the last section, this was not necessary as we considered a cascade without feedback. We will now consider the network shown in figure 12.3F; 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,

$$\begin{aligned} e_{1p} &= e_{1p}(e_{3p}, s) \\ e_{2p} &= e_{2p}(e_{1p}) \\ e_{3p} &= e_{3p}(e_{2p}) \end{aligned} \quad (12.90)$$

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.

$$0 = V_{max,1} \cdot \frac{K_D}{K_D + e_{3p}} \cdot \frac{s}{K_s + s} \cdot \frac{e_{1,total} - e_{1p}}{e_{1,total} - e_{1p} + K_{k,1}} - V_{MAX,ph,1} \cdot \frac{e_{1p}}{K_{ph} + e_{1p}} \quad (12.91)$$

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,

$$\begin{aligned} de_{1p} &= \frac{\partial e_{1p}}{\partial e_{3p}} de_{3p} + \frac{\partial e_{1p}}{\partial s} ds \\ de_{2p} &= \frac{\partial e_{2p}}{\partial e_{1p}} de_{1p} \\ de_{3p} &= \frac{\partial e_{3p}}{\partial e_{2p}} de_{2p} \end{aligned} \quad (12.92)$$

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

$$\begin{aligned} \frac{d \ln e_{1p}}{d \ln s} &= \frac{\partial \ln e_{1p}}{\partial \ln e_{3p}} \frac{d \ln e_{3p}}{d \ln s} + \frac{\partial \ln e_{1p}}{\partial \ln s} \\ \frac{d \ln e_{2p}}{d \ln s} &= \frac{\partial \ln e_{2p}}{\partial \ln e_{1p}} \frac{d \ln e_{1p}}{d \ln s} \\ \frac{d \ln e_{3p}}{d \ln s} &= \frac{\partial \ln e_{3p}}{\partial \ln e_{2p}} \frac{d \ln e_{2p}}{d \ln s} \end{aligned} \quad (12.93)$$

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  $s$ ,

$$R_s^{e_3p} = \frac{r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_s^{e_1p}}{1 - r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_s^{e_1p}} \quad (12.94)$$

This equation illustrates that the response of the cascade without the feedback loop, i.e.  $r_{e_2 p}^{e_3 p} r_{e_1 p}^{e_2 p} r_s^{e_1 p}$ , is reduced by the feedback strength of the entire loop  $r_{e_2 p}^{e_3 p} r_{e_1 p}^{e_2 p} r_{e_3 p}^{e_1 p}$ . Note that  $r_{e_3 p}^{e_1 p} < 0$  for a negative feedback loop, the phosphorylation level of  $e_1 p$  is reduced upon a increase in  $e_3 p$ .

## 12.12 Using the summation and connectivity theorems to make deductions

### 12.12.1 Metabolic pathway with one enzyme without product inhibition

Consider a linear metabolic pathway with four enzymes,



Note that all enzymes have the usual substrate (activating) and product (inhibiting) sensitivities, except for enzyme 2, which is insensitive to its product. Enzyme 2 should therefore be irreversible and not have kinetic product inhibition, e.g. it has a rate equation such as  $v_2 = V_{max,2}x/(K_{2,x}+x)$ . The summation and connectivity theorems give a quick overview of the control properties of this network, because  $\epsilon_y^{v_2} = 0$  we obtain for the connectivity theorems with respect to  $Y$ ,

$$C_3^J \epsilon_y^{v_3} = 0 \Rightarrow C_3^J = 0 \quad (12.96)$$

$$C_3^X \epsilon_y^{v_3} = 0 \Rightarrow C_3^X = 0 \quad (12.97)$$

$$C_3^Y \epsilon_y^{v_3} = -1 \Rightarrow C_3^Y = \frac{-1}{\epsilon_y^{v_3}} \quad (12.98)$$

$$C_3^Z \epsilon_y^{v_3} = 0 \Rightarrow C_3^Z = 0 \quad (12.99)$$

The connectivity theorem for  $Z$  equals,

$$C_3^J \epsilon_Z^{v_3} + C_4^J \epsilon_Z^{v_4} = 0 \Rightarrow C_4^J = 0 \quad \text{because } C_3^J = 0 \quad (12.100)$$

As a consequence the summation theorem for flux control coefficients,  $C_1^J + C_2^J + C_3^J + C_4^J = 1$ , simplifies into,

$$C_1^J + C_2^J = 1 \quad (12.101)$$

Together with connectivity theorem for  $X$ ,

$$C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} = 0 \quad (12.102)$$

we are left with two equations to express the unknown, nonzero flux control coefficients,  $C_1^J$  and  $C_2^J$  in terms of the elasticity coefficients  $\epsilon_x^{v_1}$  and  $\epsilon_x^{v_2}$ . The solution becomes,

$$C_2^J = \frac{1}{1 - \frac{\epsilon_x^{v_2}}{\epsilon_x^{v_1}}} \quad \text{and} \quad C_1^J = 1 - C_2^J \quad (12.103)$$

Thus the flux control of this system is determined only by the properties of the first two enzymes; it is as if we were studying just the first two reactions of the linear pathway! This should also make sense when you realise that enzyme 3 and 4 cannot communicate with enzyme 1 and 2 because enzyme 1 and 2 do not sense any changes made to enzyme 3 and 4 because enzyme 2 is insensitive to  $y$ . Thus, enzyme 1 and 2 together form a "mass pump".

### 12.12.2 Metabolic pathway with one enzyme without product inhibition and one negative feedback

In this section, we study the same network as in the previous section but, in addition,  $Z$  inhibits the first reaction by way of negative feedback (indicated in bold font),



Again we have (as above),

$$C_3^J \epsilon_y^{v_3} = 0 \Rightarrow C_3^J = 0 \quad (12.105)$$

$$C_3^X \epsilon_y^{v_3} = 0 \Rightarrow C_3^X = 0 \quad (12.106)$$

$$C_3^Y \epsilon_y^{v_3} = -1 \Rightarrow C_3^Y = \frac{-1}{\epsilon_y^{v_3}} \quad (12.107)$$

$$C_3^Z \epsilon_y^{v_3} = 0 \Rightarrow C_3^Z = 0 \quad (12.108)$$

The connectivity theorem for  $Z$  and the flux control coefficients now has an additional term, in contrast to the previous section,

$$C_1^J \epsilon_Z^{v_1} + C_3^J \epsilon_Z^{v_3} + C_4^J \epsilon_Z^{v_4} = 0 \Rightarrow C_1^J \epsilon_Z^{v_1} + C_4^J \epsilon_Z^{v_4} = 0 \quad \text{because } C_3^J = 0 \quad (12.109)$$

Thus, the summation theorem now reduces two,

$$C_1^J + C_2^J + C_4^J = 0 \quad (12.110)$$

And the relevant connectivity theorems are,

$$\begin{aligned} C_1^J \epsilon_X^{v_1} + C_2^J \epsilon_X^{v_2} &= 0 \\ C_1^J \epsilon_Z^{v_1} + C_4^J \epsilon_X^{v_4} &= 0 \end{aligned} \quad (12.111)$$

Thus the flux control is distributed between enzyme 1, 2, and 4.

### 12.12.3 Metabolic pathway without product inhibition of the first reaction and one negative feedback

Next, we consider the previous scenario with two small changes: i. the first reaction is product insensitive but sensitive to a feedback by  $Z$  and ii.  $Y$  also inhibits enzyme 2,



Now we conclude,

$$C_2^J \epsilon_x^{v_2} = 0 \Rightarrow C_2^J = 0 \tag{12.113}$$

$$C_2^X \epsilon_x^{v_2} = -1 \Rightarrow C_2^X = \frac{-1}{\epsilon_x^{v_2}} \tag{12.114}$$

$$C_2^Y \epsilon_x^{v_2} = 0 \Rightarrow C_2^Y = 0 \tag{12.115}$$

$$C_2^Z \epsilon_x^{v_2} = 0 \Rightarrow C_2^Z = 0 \tag{12.116}$$

And because  $C_2^J = 0$  the connectivity theorems for  $Y$  and  $Z$  reduce to,

$$\begin{aligned}
 C_2^J \epsilon_y^{v_2} + C_3^J \epsilon_y^{v_3} &= C_3^J \epsilon_y^{v_3} = 0 \Rightarrow C_3^J = 0 \\
 C_1^J \epsilon_z^{v_1} + C_3^J \epsilon_z^{v_3} + C_4^J \epsilon_z^{v_4} &= C_1^J \epsilon_z^{v_1} + C_4^J \epsilon_z^{v_4} = 0
 \end{aligned} \tag{12.117}$$

Since,  $C_2^J = C_3^J = 0$  we obtain as the relevant theorems,

$$C_1^J + C_4^J = 1 \tag{12.118}$$

$$C_1^J \epsilon_z^{v_1} + C_4^J \epsilon_z^{v_4} = 0 \tag{12.119}$$

Thus,

$$C_4^J = \frac{1}{1 - \frac{\epsilon_z^{v_4}}{\epsilon_z^{v_1}}} \quad \text{and} \quad C_1^J = 1 - C_4^J \tag{12.120}$$

Thus, enzyme 2 and 3 have no flux control and the flux is therefore robust to changes in the activities or concentrations of enzyme 2 and 3. If the negative feedback is strong  $-\epsilon_z^{v_1} \gg \epsilon_z^{v_4}$  such that  $C_4^J \approx 1$ . Thus, strong negative feedback shifts the control to the reaction(-s) upstream of the feedback metabolite ( $Z$ ) in the pathway.

We have studied this system earlier in the book when we studied robustness in Chapter 3. Then we concluded that at steady state the concentration of  $Z$  could be determined from  $v_1(z) - v_4(z) = 0$ , which lead to the same conclusions as in this section.



# Chapter 13

## Flux maximization by optimal metabolic enzyme expression

### 13.1 Metabolic pathways, enzyme levels, and fitness

Unicellular organisms often compete for growth rate. This means that the genotype that gives rise to the highest growth rate leaves most offspring and overtakes the population. This is especially relevant when nutrients are in excess. Those conditions can be mimicked in the lab in so-called batch cultivations.

For instance, consider two organisms,  $A$  and  $B$ , with growth rates  $\mu_A$  and  $\mu_B$  at excess nutrient conditions such that the number of organisms  $A$  and  $B$  increases exponentially at a fixed rate,

$$\frac{dN_A}{dt} = \mu_A N_A \quad \text{and} \quad \frac{dN_B}{dt} = \mu_B N_B \quad (13.1)$$

Clearly, the bug with the highest  $\mu$  wins: consider for instance  $\frac{d}{dt} \ln \frac{N_A}{N_B} = \frac{d \ln N_A}{dt} - \frac{d \ln N_B}{dt} = \frac{1}{N_A} \frac{dN_A}{dt} - \frac{1}{N_B} \frac{dN_B}{dt} = \mu_A - \mu_B$  to see this. Here the growth rates are fixed because the nutrient is in excess. If one of the nutrients is limiting then the growth rate depends on the concentration of this nutrient as,  $\mu_A = \mu_A^{MAX} \frac{s}{s+K_s}$  with  $s$  as the nutrient concentration and  $K_s$  as the so-called Monod constant.

So, the fastest organism will win, but how can cells achieve a state of maximal growth rate? One approach to think about this phenomenon is to consider organisms that have limited resources at their disposal for enzyme synthesis and that try to optimise the enzyme concentrations to maximise their growth rate.

Why would this work in the first place? In other words, why is it that adjusting the enzyme concentrations appropriately leads to optimisation of the flux through a metabolic pathway (leading to new cells)? This derives from the general fact that enzyme rates depend linearly on the enzyme concentration, i.e.  $v = e \cdot f(\text{metabolite concentrations, kinetic parameters})$ , and that therefore also the maximal enzyme rate depends linearly on the enzyme, i.e.  $V_{max} = k_{cat} \cdot e$ . As a consequence, in a linear pathway where the total enzyme concentrations is bounded due to a resource limit, every change in an enzyme concentration is at the expense of other enzyme concentrations. So, increasing one enzyme concentration, say  $i$ , to increase its  $V_{max}$ , causes a reduction in or more other  $V_{max}$ 's in the pathway. Thus, those other enzymes loose capacity whereas enzyme  $i$  gains capacity. When you play this game long enough and aim at optimising the steady-state pathway flux, by only making increases and simultaneous reductions in particular enzyme concentrations such that the steady-state pathway flux always increases, then you end up in the state where for all enzymes an increase in the concentrations would cause a reduction in the flux, then you have arrived at the optimal state. We will approach this problem first with a mathematical model example and then using theory.

### 13.2 Numerical illustration of pathway flux optimisation

We consider a metabolic pathway that is composed out of three enzymes catalysing the following reactions,



Each enzymes follows reversible Michaelis-Menten kinetics:  $v(s, p) = \frac{V_M \frac{s}{K_s} \left(1 - \frac{p}{s K_{eq}}\right)}{1 + \frac{s}{K_s} + \frac{p}{K_p}}$  with  $s$  and  $p$  as the concentration of the substrate and the product for the enzyme, e.g. for enzyme 2 this would be  $x$  and  $y$ .

The enzyme concentrations set the maximal rate:  $V_M = k_{cat}e$ . This immediately leads to the following conclusion the steady state flux  $J$  equals  $J = v_1 = v_2 = v_3$  and each  $v_i$  is proportional to  $e_i$  and hence multiplication of all enzyme concentration with  $\alpha$  causes  $J$  to change and become equal to  $\alpha J$ :  $\alpha J(e_1, e_2, e_3) = J(\alpha e_1, \alpha e_2, \alpha e_3)$ . Thus, optimisation of the pathway flux by varying the enzyme concentrations should be done under an constraint of total enzyme, otherwise infinity would always be obtained. So, we consider  $e_T = e_1 + e_2 + e_3$ . Interestingly, regardless of the choice of the value of  $e_T$  the maximal value of  $J/e_T$  is not dependent on  $e_T$  because  $\frac{J}{e_T} = \frac{\alpha J}{\alpha e_T}$ . Therefore we take this as our objective function. This objective function gives the maximal pathway flux per total invested enzyme. In microbiology, this flux is called a

specific flux. Interestingly, the specific flux is also the correct perspective when one is interested in the optimisation of the growth rate of organisms.

The exponential growth of a population of cells is described by,

$$\frac{d\mathcal{X}}{dt} = \mu\mathcal{X} \Rightarrow \frac{1}{\mathcal{X}} \frac{d\mathcal{X}}{dt} = \mu : \text{Rate of biomass growth per unit biomass}$$
(13.3)

with  $\mathcal{X}$  as the amount of cells in grams and  $\mu$  equals the specific growth rate. So, if the metabolic pathway we consider makes protein then  $J/e_T$  equals the rate of protein synthesis per total protein; which is equal the specific growth rate. If this would not equal the specific growth rate then protein would either accumulate or deplete during growth. To see this consider  $\frac{de_T}{dt} = J - \mu e_T$ ; at steady state growth  $\mu = J/e_T$  with  $J$  as the protein synthesis rate and  $J/e_T$  as the *specific* protein synthesis rate. Thus, optimisation of  $J/e_T$  makes a lot of sense when fitness equals the specific growth rate of cells.

Let's return to the problem: optimisation of the specific flux of a three-enzyme metabolic pathway by varying the enzyme concentrations under the  $e_T$  constraint:  $e_1 + e_2 + e_3 = 1$ . In Figure 13.1, we show the result of such an optimisation. The important figure to understand is Figure 13.1B. We will start at the borders. We have three of them: along the vertical border  $e_1 = 0$ , along the horizontal  $e_2 = 0$ , and along the diagonal  $e_3 = 0$  as along this line  $e_1 + e_2 = 1$ . Thus, along the border always one enzyme concentration is zero and therefore the steady state flux,  $J$ , has to equal 0. Outside those border one or more enzyme concentrations are negative (given the constraint) and therefore this is not an infeasible area. The triangle is therefore the space of feasible flux solutions. The black dot marks the maximal flux value in this space. This is also indicated by Figure 13.1C & D which give a 2D view on the solution space. If we would have chosen different kinetic parameters for the enzymes or different values for  $S$  and  $P$  a different optimum value for  $J$  would have been found at a different combination of enzyme concentrations.

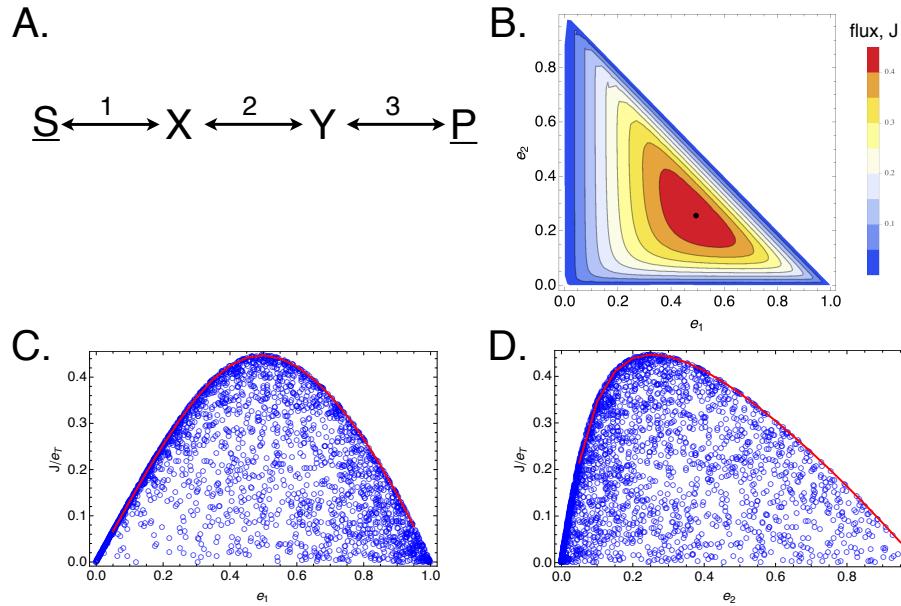
So, now we know that optimal enzyme concentrations exist that optimise the specific flux through the metabolic pathway. Whether gene networks can actually steer metabolic pathways to those states is at this stage still unclear and will be considered in the next chapter.

### 13.3 Growth rate reduction as a response to non-functional protein expression

In the last section, we concluded that for a linear pathway,

$$\alpha J(e_1, e_2, \dots, e_r) = J(\alpha e_1, \alpha e_2, \dots, \alpha e_r) \quad \text{and similarly } \alpha J(e_T) = J(\alpha e_T)$$
(13.4)

Suppose now that we make a protein  $\bar{E}$  that has no influence on growth but its expression is of course at the expense of the other growth-promoting proteins.



**Figure 13.1: Overview of optimisation results of three-enzyme metabolic pathway.** A. We consider a metabolic pathway with three enzymes. Each enzyme follows reversible Michaelis Menten kinetics. The concentrations of the pathway substrate,  $S$ , and the product,  $P$ , are held fixed to allow for the occurrence of a steady state. B. The enzyme levels are varied under the constrain  $e_1 + e_2 + e_3 = 1$  and the steady state flux  $J$  is calculated. Clearly, one set of enzyme concentrations maximises the steady state flux. Note that along the border of the triangle the flux has to equal zero because at those borders one enzyme concentration is alway zero. Outside the borders one of the enzyme concentration is always negative, which is not possible; this is due to the fact that the sum of the enzyme concentrations has to equal 1. C & D The dependency of the steady state flux is shown when the enzyme concentrations are sampled randomly under the constraint that  $e_1 + e_2 + e_3 = 1$ . 2500 samples were taken leading to the same number of models for which the steady state flux was calculated.

Thus, the protein available for growth equals  $e_T - \bar{e} = e_T(1 - \frac{\bar{e}}{e_T})$ . Thus, we obtain from the previous relations that,

$$\left(1 - \frac{\bar{e}}{e_T}\right) J(e_T) = J\left(\left(1 - \frac{\bar{e}}{e_T}\right) e_T\right) \quad (13.5)$$

Thus, the fractional change in the growth rate depends on the protein fraction of dummy protein as,

$$\frac{\Delta J}{J} = \frac{\left(1 - \frac{\bar{e}}{e_T}\right) J(e_T) - J(e_T)}{J(e_T)} = -\frac{\bar{e}}{e_T} \quad (13.6)$$

(Note that this result assumes that all other enzymes are expressed at the same fractions,  $e_i/e_T$ , before and after the synthesis of the dummy protein at level  $\bar{e}$ .) Thus, the fractional reduction in growth rate of pathway flux upon dummy protein expression equals  $\bar{e}/e_T$  indicating that the  $J$  reduces in a linear fashion with  $\bar{e}$ , i.e.  $J(\bar{e}) = J(\bar{e}=0)(1 - \frac{\bar{e}}{e_T})$ . Since, we require always the same protein expression fractions of the growth promoting enzymes then the relationship of this section immediately apply for the states of metabolic pathway at maximal specific flux.

### 13.4 Theoretical illustration of pathway flux optimisation (advanced material)

We will optimise the flux,  $J$ , through a linear metabolic pathway at steady state by changing the enzyme concentrations under the constraint  $R = \sum_i \omega_i e_i$ . Here  $\omega_i$  is the resource cost per unit enzyme concentration  $e_i$  and  $R$  denotes the total resource available for enzyme synthesis. For constraint optimisation, the Lagrange multiplier method gives often valuable insights (see Wikipedia or a Calculus textbook). For this problem the Lagrangian  $\mathcal{L}$  is defined as,

$$\mathcal{L}(\mathbf{e}) = J(\mathbf{e}) + \lambda \left( R - \sum_i \omega_i e_i \right) \quad (13.7)$$

The vector  $\mathbf{e}$  contains all the enzyme concentrations;  $e_i$  is the  $i$ -th entry of  $\mathbf{e}$ . At the optimum state we have that for all  $i$ ,

$$\frac{\partial \mathcal{L}}{\partial e_i} = \frac{\partial J}{\partial e_i} + \lambda \omega_i = 0 \Rightarrow \frac{\partial J}{\partial e_i} = -\lambda \omega_i \quad (13.8)$$

We know that  $\sum_i \frac{\partial \ln J}{\partial \ln e_i} = 1$  from metabolic control analysis<sup>1</sup>. Therefore,

$$\sum_i \frac{e_i}{J} \frac{\partial J}{\partial e_i} = \sum_i -\lambda \omega_i \frac{e_i}{J} = -\lambda \frac{R}{J} = 1 \Rightarrow \lambda = -\frac{J}{R} \quad (13.9)$$

Thus,

$$\frac{\partial J}{\partial e_i} = \frac{J}{R} \omega_i \Rightarrow \frac{\partial \ln J}{\partial \ln e_i} = \frac{\omega_i e_i}{R} \quad (13.10)$$

---

<sup>1</sup>This derives from the following fact: the steady state flux is first-order homogeneous function of the enzyme concentrations:  $\alpha J(\mathbf{e}) = J(\alpha \mathbf{e})$  and from Euler's theorem for first-order homogeneous functions  $\sum_i e_i \frac{\partial J}{\partial e_i} = J$ .

in the optimal state. The relation  $\frac{\partial J}{\partial e_i} = \frac{J}{R}\omega_i$  tells us that the enzyme with largest resource cost  $\omega_i$  gives the largest absolute change in the flux, i.e.  $\delta J = \frac{\partial J}{\partial e_i}\delta e_i = \frac{J}{R}\omega_i\delta e_i$ , if the same  $\delta e_i$  is applied to all enzymes. If all the enzymes are equally costly, they all give the same absolute flux change upon an identical absolute concentration change. However, for the same fractional change, i.e. when for all enzymes,  $\frac{\delta e_i}{e_i} = \delta \ln e_i$  is the same, the enzyme with the highest concentration gives the highest fractional flux change, which immediately follows from the conclusion that we arrived at in the previous sentence – the higher the  $e_i$  the higher the  $\delta e_i$  required for the same fractional change  $\frac{\delta e_i}{e_i}$  and  $\delta J$  is proportional to  $\delta e_i$ .

## Chapter 14

# Optimal gene control systems

### 14.1 What's so amazing about molecular control systems?

We often take it for granted: cells adapting to changes in their environment by changes in gene expression and rewiring of metabolic, signalling, and gene networks. It is so intrinsically part of the mindset of a cell biologist that asking "How is this even possible?" would hardly-ever cross the mind of a cell biologist.

The astonishing thing is that engineers and physicists continue to struggle with designing inanimate systems that have such properties; yet, cell biologists do not even frown. Properties such as self-repair, reproduction, and stress-resistance are far from trivial to engineer. So, autonomous, adaptive control systems are still a big challenge in engineering while this is what cells do "out of the box", it is what defines them. I think that cell biologists do not understand these self-regulation properties much better than engineers and physicists do. Biologists understand much better "what" cells are made of than "how" they work and "why". "How" asks about the mechanism and "why" for the (evolutionary) benefit. Molecular cell biology and genetics is often not busy with such questions but addresses the "what cells are made of" question. So, we need others to think about how cells achieve self-regulation. Systems biologists have taken up this challenge.

In this chapter, we will play with some molecular systems that have quite surprising properties, all associated with self-regulation. What will be surprising thing is the small number of molecular components required for self-regulation.

## 14.2 Sensing and signal integration by enzymes, genes and networks

### 14.2.1 Sensing and integration

In order to respond cells have to sense extracellular signals, such as forces and concentrations of for instance nutrients, stresses, and growth factors. Often the cellular response with respect to one signal depends on the magnitude of other signals and the cell integrates both. This integration is done at the level proteins as well as in networks. We discuss examples of both these mechanisms.

### 14.2.2 Regulation of single enzymes, proteins, and genes

#### Regulation of metabolic enzymes by changes in enzyme, reactant and effector concentration

The rate of a metabolic enzyme can always be written as,

$$v = k_{cat}e \prod_j \frac{s_j}{K_j} \left( 1 - \frac{\prod_i p_i}{K_{eq} \prod_j s_j} \right) f(\mathbf{x}, \mathbf{p}) \quad (14.1)$$

Where  $\mathbf{x}$  is a vector containing the concentrations of reactants and effectors. The vector  $\mathbf{p}$  contains the kinetic parameters of the enzymes, such as the  $K_M$ 's. The reversible Michaelis-Menten is an example of such an equation,

$$v = \underbrace{k_{cat}^+ e}_{V_{max}^+} \frac{s}{K_s} \left( 1 - \frac{p}{s \cdot K_{eq}} \right) \underbrace{\frac{1}{1 + \frac{s}{K_s} + \frac{p}{K_p}}}_{f(\mathbf{x}, \mathbf{p})} \quad (14.2)$$

The sensitivity of a metabolic enzyme to the concentration of a reactant, itself, or effector is captured by its so-called elasticity coefficient,

$$\epsilon_x^v = \frac{x}{v} \frac{\partial v}{\partial x} = \frac{\partial \ln v}{\partial \ln x} \quad (14.3)$$

where  $x$  denotes the concentration of a reactant or effector. We find that (with  $y$  as the concentration of some effector),

$$\epsilon_e^v = 1 \quad (14.4)$$

$$\epsilon_y^v = \frac{\partial \ln f}{\partial \ln y} \quad (14.5)$$

$$\epsilon_{s_j}^v = \frac{1}{1 - \frac{\prod_i p_i}{\prod_j s_j \cdot K_{eq}}} + \frac{\partial \ln f}{\partial \ln s_j} \quad (14.6)$$

$$\epsilon_{p_i}^v = \frac{-\frac{\prod_i p_i}{\prod_j s_j \cdot K_{eq}}}{1 - \frac{\prod_i p_i}{\prod_j s_j \cdot K_{eq}}} + \frac{\partial \ln f}{\partial \ln p_i} \quad (14.7)$$

At thermodynamic equilibrium  $\frac{\prod_i p_i}{\prod_j s \cdot K_{eq}} = 1$ ; therefore the displacement from thermodynamic equilibrium is therefore captured by  $1 - \frac{\prod_i p_i}{\prod_j s \cdot K_{eq}}$ . This indicates that the sensitivity of an enzyme to its reactants depends on the enzyme's displacement from thermodynamic equilibrium as captured by  $\frac{\prod_i p_i}{K_{eq} \prod_j s_j}$ , whereas its sensitivity to its own concentration (i.e.  $e$ ) and that of an effector (i.e.  $y$ ) is independent of this term.

If an enzyme is sensitive to multiple effectors, such as for instance phosphofructokinase, then the outcome of one effector  $Y_2$  depends on the magnitude of another effector,  $Y_2$  if,

$$\frac{\partial^2 \ln v}{\partial \ln y_1 \partial \ln y_2} \neq 0 \quad (14.8)$$

because only in this case does the sensitivity of the enzyme of  $Y_1$  depend on the concentration of the second regulator.

### Signal integration by transcription factors

We consider a monomeric regulatory protein  $P$ , e.g. a transcription factor, that binds to two signals ( $X$  and  $Y$ ) that engage in an allosteric interaction. In addition, the protein exists in two conformations; one that is active  $R$  and the other  $T$  is in active. Following the usual methods for the derivation of equilibrium binding models, we obtain:

$$\begin{aligned} p_T &= r + rx + ry + rxy + t + tx + ty + txy \\ &= r \left( 1 + \frac{x}{K_{r,x}} + \frac{y}{K_{r,y}} + \frac{x \cdot y}{\alpha K_{r,x} K_{r,y}} + \frac{t}{r} \left( 1 + \frac{x}{K_{t,x}} + \frac{y}{K_{t,y}} + \frac{x \cdot y}{\alpha K_{t,x} K_{t,y}} \right) \right) \end{aligned} \quad (14.9)$$

The  $K$ 's are dissociation constants and  $\alpha$  is an allosteric interaction coefficient.<sup>4</sup> If we define  $RX$  as the active state then we obtain,

$$rx = p_T \frac{\frac{x}{K_{r,x}}}{1 + \frac{x}{K_{r,x}} + \frac{y}{K_{r,y}} + \frac{x \cdot y}{\alpha K_{r,x} K_{r,y}} + L \left( 1 + \frac{x}{K_{t,x}} + \frac{y}{K_{t,y}} + \frac{x \cdot y}{\alpha K_{t,x} K_{t,y}} \right)} \quad (14.11)$$

and  $Y$  acts as an inhibitor. Here we have defined  $L = \frac{t}{r}$ . So, the process activated by  $RX$ , for instance the transcription of a gene, is now influenced both by  $X$  and  $Y$ . And the sensitivity of  $rx$  with respect to  $x$  depends on  $y$ , i.e.

$$\frac{\partial \ln rx}{\partial \ln x} = \dots = f(x, y) \quad (14.12)$$

So, the influence of  $X$  on the regulatory proteins depends on the concentration of  $Y$ . Again a measure for this cross-regulation is  $\frac{\partial^2 \ln rx}{\partial \ln x \partial \ln y}$ .

### 14.2.3 Multiple transcription factors regulating the transcription rate of one gene

Describing how two transcription factor  $T_1$  and  $T_2$  influence the transcription rate  $v$  of a gene is quite straightforward,

$$v = v(t_1, t_2) \approx k \cdot f(t_1, t_2) \quad (14.13)$$

Say, they are both activators and can independently activate the gene then,

$$v = \frac{k_{t_1} \frac{t_1}{K_{t_1}} + k_{t_2} \frac{t_2}{K_{t_2}} + k_{t_1 t_2} \frac{t_1 \cdot t_2}{\alpha \cdot K_{t_1} \cdot K_{t_2}}}{1 + \frac{t_1}{K_{t_1}} + \frac{t_2}{K_{t_2}} + \frac{t_1 \cdot t_2}{\alpha \cdot K_{t_1} \cdot K_{t_2}}} \quad (14.14)$$

If  $T_2$  would be an inhibitor then we could for instance have,

$$v = \frac{k_{t_2} \frac{t_2}{K_{t_2}}}{1 + \frac{t_1}{K_{t_1}} + \frac{t_2}{K_{t_2}} + \frac{t_1 \cdot t_2}{\alpha \cdot K_{t_1} \cdot K_{t_2}}} \quad (14.15)$$

Again this all derives from equilibrium binding models of the two transcription factors to their regulatory sequences on the DNA and the two transcription factors engage in an allosteric interaction (quantified by  $\alpha$ ).

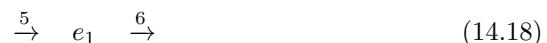
## 14.3 Regulation and networks

In the previous subsections we have shortly illustrated how signal binding to transcription factors and how transcription factors influence transcription rate can be captured in terms of kinetic models. The outcome of those events is eventually the changes in the expression of proteins, e.g. metabolic enzymes, which change the concentrations of the molecules that bind to the transcription factors, which changes transcription rates, which changes enzyme concentrations, etc, etc. So, an iterative process results until some new steady state is reached, assuming it can be reached. One would expect that in the new steady state the controlled system has a certain desired performance that enhances the fitness of the organism. Next, we will outline an approach for thinking about such problems.

## 14.4 Control of metabolism of by gene expression

### 14.4.1 A simple example

Consider the following system,



Reaction 1 and 2 are catalysed by enzymes and make up metabolism. Enzyme 2 is held fixed and enzyme 1's gene expression control is considered.  $m_1$  denotes the mRNA concentration and  $e_1$  that of the enzyme. At steady state we find for the enzyme concentration in the simplest scenario,

$$e_1 = \frac{k_5 m_1}{k_6} = \frac{k_5 k_3 f(x)}{k_6 k_4} \quad (14.19)$$

where  $f(x)$  is the gene regulatory function of gene encoding the mRNA  $m_1$ . We take,

$$f(x) = \frac{1}{1+x} = \frac{k_6 k_4}{k_5 k_3} e_1 \quad (14.20)$$

So, the synthesis of mRNA is inhibited by  $X$ .  $X$  is high when enzyme 2 starts to saturate with  $X$ , this occurs when enzyme 1 can function at higher activity than enzyme 2. Therefore, high  $x$  indicates excess of enzyme 1 and its synthesis should be therefore be reduced. At steady state, for the entire system  $x = x_s$  and we have,

$$v_1 = v_2 \Rightarrow k_1 e_1 \frac{s}{1+s+x_s} = V_2 \frac{x_s}{1+x_s} \quad (14.21)$$

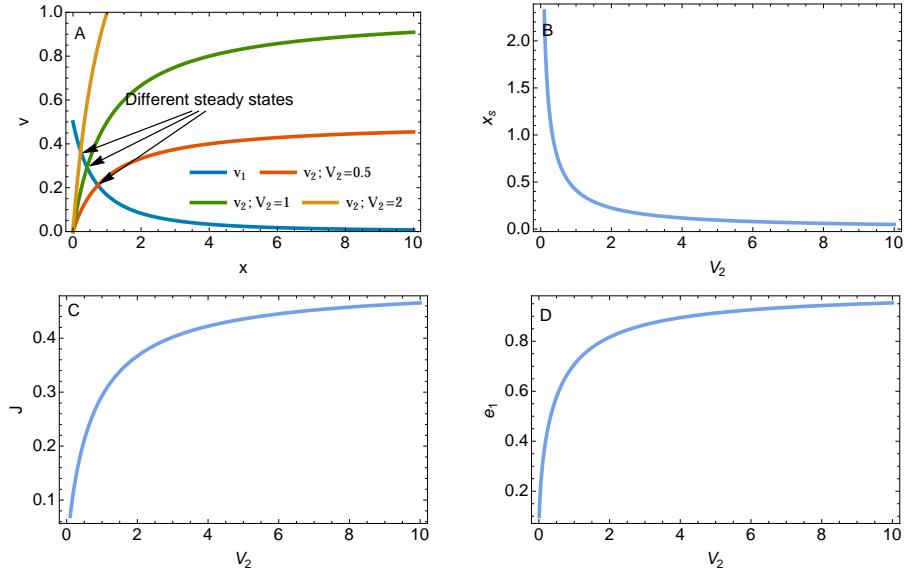
$$\Rightarrow k_1 \frac{1}{1+x} \frac{k_5 k_3}{k_6 k_4} \frac{s}{1+s+x_s} = V_2 \frac{x_s}{1+x_s} \quad (14.22)$$

Which we simplify to (this does not effect any of the conclusions of this section),

$$\overbrace{\frac{\alpha}{1+x_s}}^{e_1} \frac{s}{1+s+x_s} = \beta \frac{x_s}{1+x_s} \quad (14.23)$$

with  $\alpha = \frac{k_5 k_3}{k_6 k_4}$  and  $\beta = \frac{V_2}{k_1}$ . So,  $\frac{\alpha}{1+x_s}$  is our measure for the enzyme concentration,  $e_1$ . Clearly, when  $V_2$  changes, the concentration of  $x_s$  changes to return to steady state, and therefore  $e_1 = \frac{\alpha}{1+x_s}$  as well. Hence, gene expression adjusts! The last equation shows that if  $\beta$  goes up,  $x_s$  down and therefore more enzyme  $e_1$  is synthesised. This is shown in Figure 14.1. We interpret an increase in  $\beta$  as an increase in the demand of the product of the metabolic pathway with reaction 1 and 2 at its entry point. Thus, if the cells require more pathway activity somehow it increases  $\beta$ , e.g. via allosteric activation of enzyme 2 by some metabolite that influences  $V_2$ .

So, this very simple molecular circuitry behaves in a rather "smart" fashion: enzyme 1 synthesized only when the demand for the pathway activity is high to an extend that does not lead to major overcapacity of enzyme 1 - because then  $x$  accumulated and it inhibits enzyme 1 synthesis. This control system is therefore useful from an economical point of view: the enzyme is roughly tuned to its demand. I say roughly because this is not the optimal response of the pathway. This would require that at all values of  $V_2$ , the value  $\frac{J}{e_1+e_2}$  is maximised (assuming equal cost of enzyme 1 and 2), which is what we shall study next.



**Figure 14.1: Regulation of metabolic flux by gene expression.** The rate of degradation and synthesis of  $X$  is plotted as function of  $x$  in plot **A**. Three different scenario's are plotted for the degradation rate. For higher values of  $V_2$ , the flux increases (**C**) and the steady state concentration of  $X$  decreases (**B**). When the demand for metabolic activity increases, i.e. when  $V_2$  goes up, the gene control system responds by making more enzyme  $e_1$  (**D**).

#### 14.4.2 Optimal gene expression

We consider the same system as in the previous subsection,



but now our aim is to identify a gene networks that maximises as function of  $s$  the following objective,

$$\frac{\text{steady-state pathway flux}}{\text{enzyme investment}} = \frac{J}{e_1 + e_2} \quad (14.25)$$

(Note that at steady state:  $J = v_1 = v_2$ ). The rate equations for the two enzymes are chosen as,

$$v_1 = k_1 e_1 \frac{s}{1+s+x} \equiv e_1 f_1(x) \quad (14.26)$$

$$v_2 = k_2 e_2 \frac{x}{1+x} \equiv e_2 f_2(x) \quad (14.27)$$

We rewrite the objective,

$$\max \frac{J}{e_1 + e_2} = \max \frac{1}{\frac{e_1}{v_1} + \frac{e_2}{v_2}} = \max \frac{1}{\frac{1}{f_1} + \frac{1}{f_2}} \quad (14.28)$$

$$\Rightarrow \min \left( \frac{1}{f_1} + \frac{1}{f_2} \right) \quad (14.29)$$

At the minimum,

$$\frac{\partial}{\partial x} \left( \frac{1}{f_1} + \frac{1}{f_2} \right) = 0 \quad (14.30)$$

Which equals for our system (when we set  $k_1 = 1$  and  $k_2 = 1$ ),

$$\frac{1}{s} - \frac{1}{x^2} = 0 \Rightarrow x = \sqrt{s} \quad (14.31)$$

So, in all optima  $x = \sqrt{s} = x_{opt}$ . At optimal steady state we have,

$$e_1 \frac{s}{1+s+x} = e_2 \frac{x}{1+x} \quad \text{with: } x = \sqrt{s} \quad (14.32)$$

Say, we have the constraint  $1 = e_1 + e_2$  then,

$$e_1 = \frac{1 + \sqrt{s} + s}{1 + 2\sqrt{s} + 2s} \quad \text{and } e_2 = 1 - e_1 \quad (14.33)$$

The maximal value of the objective becomes,

$$\frac{J}{e_1 + e_2} = \frac{s}{1 + 2\sqrt{s} + 2s} \quad (14.34)$$

and ranges between 0 and  $\frac{1}{2}$ . See Figure 14.2 were the enzyme concentrations and the objective are plotted as function of  $s$ .

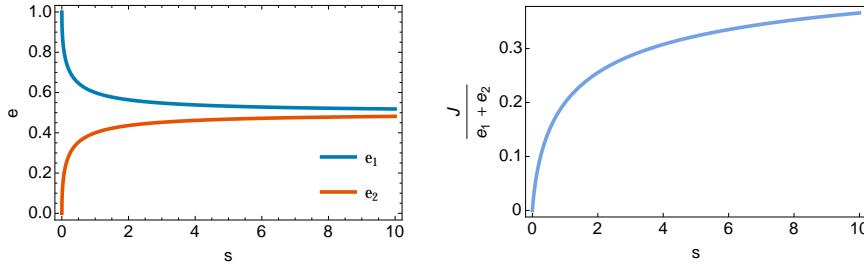


Figure 14.2: **Optimal control of metabolic pathway by gene expression.**  
A. Shows the optimal enzyme concentrations as function of  $s$ , i.e. equations 14.33. B. Shows the dependency of the maximal objective value, equation 14.34, as function  $s$ .

At steady state we have for the enzyme concentrations the following balance relations,

$$0 = k_m^+ g(x) - k_m^- m \quad (14.35)$$

$$0 = k_e^+ m - k_e^- e \quad (14.36)$$

with the transcription rate constant  $k_m^+$ , mRNA degradation constant  $k_m^-$ , translation rate constant  $k_e^+$ , protein degradation rate constant  $k_e^-$  and the gene regulatory function  $g(x)$ . When all parameters are set to 1 then the steady state concentration of the enzyme equals  $e = g(x)$ . So, this must mean that the gene regulatory function for enzyme 1 equals  $\frac{1+x+x^2}{1+2x+2x^2}$  and for  $e_2 = \frac{x+x^2}{1+2x+2x^2}$  (equation 14.33 with  $x = \sqrt{s}$  as required in the optimum). One can now verify that the following set of ode's describe the optimally regulated system,

$$\frac{dx}{dt} = e_1 \frac{s}{1+s+x} - e_2 \frac{x}{1+x} \quad (14.37)$$

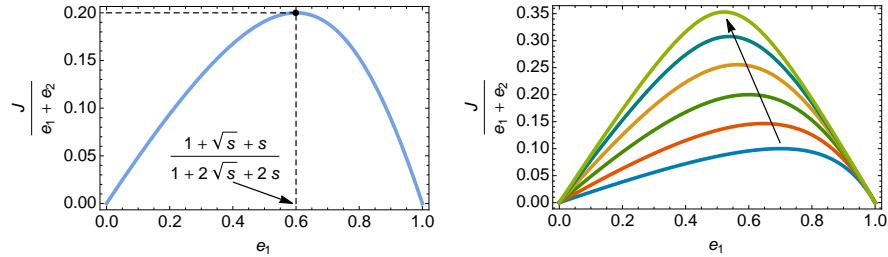
$$\frac{dm_1}{dt} = \frac{1+x+x^2}{1+2x+2x^2} - m_1 \quad (14.38)$$

$$\frac{de_1}{dt} = m_1 - e_1 \quad (14.39)$$

$$\frac{dm_2}{dt} = \frac{x+x^2}{1+2x+2x^2} - m_2 \quad (14.40)$$

$$\frac{de_2}{dt} = m_2 - e_2 \quad (14.41)$$

For all values of  $s$  this system will at a steady state with the maximal value for  $\frac{J}{e_1 + e_2}$  because it always has  $x = \sqrt{s}$  as steady state value! This is further illustrated in Figure 14.3.



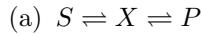
**Figure 14.3: Illustration of optimality of optimal control system. A.** We solve for the steady state concentration of  $x$  as function of  $e_1$  by solving  $e_1 \frac{s}{1+s+x} - (1 - e_1) \frac{x}{1+x} = 0$  and calculate  $J/e_T$ . Then the maximum occurs when  $x = \sqrt{s}$  and  $e_1 = \frac{1+\sqrt{s}+s}{1+2\sqrt{s}+2s}$ . Hence, the system is optimal. **B.** The "fitness landscape" for different values of  $s$ .  $s$  increases in the direction of the arrow; its values are 0.25, 0.5, 1, 2, 4, and 8.

# Chapter 15

## Answers to exercises

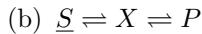
### 15.1 Answers to Chapter 2 exercises

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.



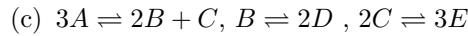
Answer:

$$\begin{aligned}\frac{ds}{dt} &= -v_1 \\ \frac{dx}{dt} &= v_1 - v_2 \\ \frac{dp}{dt} &= v_2 \\ v_1 &= k_1^+ s - k_1^- x \\ v_2 &= k_2^+ x - k_2^- p\end{aligned}$$



Answer:

$$\begin{aligned}\frac{dx}{dt} &= v_1 - v_2 \\ \frac{dp}{dt} &= v_2 \\ v_1 &= k_1^+ s - k_1^- x \\ v_2 &= k_2^+ x - k_2^- p\end{aligned}$$



Answer:

$$\begin{aligned}
 \frac{da}{dt} &= -3v_1 \\
 \frac{db}{dt} &= 2v_1 - v_2 \\
 \frac{dc}{dt} &= v_1 - 2v_3 \\
 \frac{dd}{dt} &= 2v_2 \\
 \frac{de}{dt} &= 3v_3 \\
 v_1 &= k_1^+ a^3 - k_1^- b^2 \cdot c \\
 v_2 &= k_2^+ b - k_2^- d^2 \\
 v_3 &= k_3^+ c^2 - k_3^- e^3
 \end{aligned} \tag{15.1}$$

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

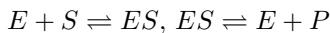
$$\begin{aligned}
 \frac{dxy}{dt} &= -v_1 \\
 \frac{dz}{dt} &= -v_1 + v_3 \\
 \frac{dxyz}{dt} &= v_1 - v_2 \\
 \frac{dx}{dt} &= v_2 \\
 \frac{dyz}{dt} &= v_2 - v_3 \\
 \frac{dy}{dt} &= v_3 \\
 v_1 &= k_1^+ xy \cdot z - k_1^- xyz \\
 v_2 &= k_2^+ xyz - k_2^- x \cdot yz \\
 v_3 &= k_3^+ yz - k_3^- y \cdot z
 \end{aligned} \tag{15.2}$$

Note that  $Z$  has functioned as a catalyst and that  $XY$  has been split into  $X$  and  $Y$  by  $Z$ !

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

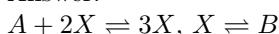
(a)  $\frac{de}{dt} = -k_1^+ e \cdot s + k_1^- es + k_2^+ es - k_2^- e \cdot p$ ,  $\frac{des}{dt} = k_1^+ e \cdot s - k_1^- es - k_2^+ es + k_2^- e \cdot p$ ,  $\frac{ds}{dt} = -k_1^+ e \cdot s + k_1^- es$ ,  $\frac{dp}{dt} = k_2^+ es - k_2^- e \cdot p$

Answer:



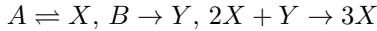
(b)  $\frac{dx}{dt} = k_1^+ a \cdot x^2 - k_1^- x^3 - k_2^+ x + k_2^- b$

Answer:



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

Answer:



- (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-phosphate and  $Z$  as ATP. What is should be the substrate of reaction 1 and the product of reaction 3?

Answer:

Glucose and pyruvate.

## Exercises

1. Determine the mass-action kinetics rate equations for all the reaction in Figure 2.8B,

Answer:

$$\begin{aligned} v_7 &= k_7^+ \cdot \bar{r_2 s_2} \cdot \alpha \beta \gamma GDP - k_7^- \cdot \bar{r_2 s_2} \alpha \beta \gamma GDP \\ v_8 &= k_8 \cdot \bar{r_2 s_2} \alpha \beta \gamma GDP \\ v_9 &= k_9^+ \cdot \bar{r_2 s_2} \alpha \beta \gamma GTP - k_9^- \cdot \bar{r_2 s_2} \cdot \beta \gamma \cdot \alpha GAP \\ v_{10} &= k_{10}^+ \cdot \alpha GTP \cdot GAP - k_{10}^- \cdot \alpha GTPGAP \\ v_{11} &= k_{11} \cdot \alpha GTPGAP \\ v_{12} &= k_{12}^+ \cdot \alpha GDP \cdot \beta \gamma - k_{12}^- \cdot \alpha \beta \gamma GDP \end{aligned}$$

2. Determine the mass balances for all the variable concentrations in Figure 2.8B,

Answer:

$$\begin{aligned} \frac{d\bar{r_2 s_2}}{dt} &= -v_7 + v_9 \\ \frac{d\alpha \beta \gamma GDP}{dt} &= -v_7 + v_{12} \\ \frac{d\bar{r_2 s_2} \alpha \beta \gamma GDP}{dt} &= v_7 - v_8 \\ \frac{d\bar{r_2 s_2} \alpha \beta \gamma GTP}{dt} &= v_8 - v_9 \\ \frac{d\beta \gamma}{dt} &= v_9 - v_{12} \\ \frac{d\alpha GTP}{dt} &= v_9 - v_{10} \\ \frac{d\alpha GTPGAP}{dt} &= v_{10} - v_{11} \\ \frac{d\alpha GDP}{dt} &= v_{11} - v_{12} \end{aligned}$$

3. Determine the total concentrations of signalling proteins that remain fixed over time in Figure 2.8B.

$$\begin{aligned}\text{conservation of } r_2 s_2 &= \bar{r}_2 \bar{s}_2 + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GDP + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GTP \\ \text{conservation of } \alpha &= \alpha \beta \gamma GDP + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GDP + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GTP + \alpha GTP \\ &\quad + \alpha GTPGAP + \alpha GDP \\ \text{conservation of } \beta \gamma &= \beta \gamma + \alpha \beta \gamma GDP + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GDP + \bar{r}_2 \bar{s}_2 \alpha \beta \gamma GTP\end{aligned}$$

Here we assumed that  $GAP$ ,  $GTP$ ,  $GDP$  and  $Pi$  can be considered fixed.

### 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 2.11 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  $s \rightleftharpoons x$ . Show that the time to reach half the steady-state concentration is halved when all rate constants are doubled in value.

Answer:

See the mathematica file: *equilibriumrelaxation.nb*. You do not have to understand the Mathematica code yet. You can evaluate a line in Mathematica by pressing SHIFT-ENTER. Just check whether you can follow the logic of this file. You will be trained in this much more often during the course.

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?

Answer:

The lines will intersect at  $x = 1$  (check by inspection of the equations). The concentration will increase to reach  $x = 1$  if initially below 1 because then  $v_1 > v_2$ . In other case, the concentration will reduce until  $x = 1$  is reached because then  $v_1 < v_2$ .

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  $V_2$ ?

Answer:

Clearly at steady state when  $v_1 = v_2$  we have:  $1/(1+x_s) = V_2 x_s / (1+x_s) \Rightarrow x_s = 1/V_2$ ! So a higher  $V_2$  reduces the steady state concentration of  $X$ .  $V_2$  correspond to the maximal rate that reaction 2 can achieve (because for very very large values of  $X$ :  $v_2 \approx V_2$ ).  $X$  inhibits the first

process and lower values of  $x$  therefore stimulate the first process, while lower values of  $x$  reduce  $v_2$ . So an increase in  $V_2$  leads to a reduction of  $x$  to enhance the rate of process 1 and slightly reduce the rate of  $v_2$  (with the new value of  $V_2$ ) such that they balance. If you find this hard to understand plot  $v_1$  and  $v_2$  as function of  $x$  in Excel for different values of  $V_2$  and then you will see what I mean.

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.

Answer:

If the system should reach equilibrium then all the rates should equal zero:  $v_1 = 0$ ,  $v_2 = 0$ , and  $v_3 = 0$ . Thus,  $b = k_1^+ a / k_1^-$  and then  $c = k_2^+ b / k_2^- = k_1^+ k_2^+ a / (k_1^- k_2^-)$  and  $d = k_3^+ c / k_3^- = k_3^+ k_2^+ k_1^+ a / (k_1^- k_2^- k_3^-)$ . Therefore, if the concentration of  $d/a$  is chosen as,

$$\frac{d}{a} = \frac{k_1^+ k_2^+ k_3^+}{k_1^- k_2^- k_3^-} \quad (15.3)$$

All the rates will be zero in the state where the concentration are constant in time.

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

Answer:

If the system should reach equilibrium then all the rates should equal zero:  $v_1 = 0$ ,  $v_2 = 0$ , and  $v_3 = 0$ . Thus,  $b = k_1^+ a / k_1^-$  and then  $c = k_2^+ b / k_2^- = k_1^+ k_2^+ a / (k_1^- k_2^-)$  and  $d = k_3^+ b / k_3^- = k_3^+ k_1^+ a / (k_1^- k_3^-)$ . Therefore, if the concentration of  $d/a$  is chosen as,

$$\frac{d}{a} = \frac{k_1^+ k_3^+}{k_1^- k_3^-} \quad (15.4)$$

All the rates will be zero in the state where the concentration are constant in time. The concentration of  $C$  is then equal to  $k_1^+ k_2^+ a / (k_1^- k_2^-)$ .

## 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?

Answer:

You will a hyperbolic relationship with:  $ab = 1/2a_T$  when  $b = K_D$ . And all  $a$  is in the complex  $ab$  when  $b >> K_D$ . So  $1/K_D$  is a natural measure for the affinity of  $A$  for  $B$  and vice versa.  $1/K_D$  is the association constant of the reaction. We can write,

$$ab = a_T \frac{b/K_D}{1 + b/K_D} \quad (15.5)$$

as

$$\frac{b}{K_D} = \frac{ab}{a_T - ab} = \frac{ab/a_T}{1 - ab/a_T} \quad (15.6)$$

If  $ab/a_T = 0.1$  (10% of  $a$  in complex) then  $b/K_D = 0.11$  and for 0.9 we find  $b/K_D = 9$ .

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%?

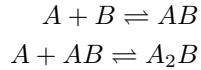
Answer:

This is the kind of question as above.

$$\frac{TF}{K_D} = \frac{TFDNA}{TF_T - TFDNA} = \frac{TFDNA/TF_T}{1 - TFDNA/TFDNA_T} \quad (15.7)$$

$\frac{TF}{K_D}$  equals 0.11, 1, and 9 when  $TFDNA/TF_T$  equals 0.1, 0.5, and 0.9, respectively. Since,  $K_D = 1$  nM the TF concentrations are 0.11, 1, and 9 nM. These are realistic concentrations for a bacterium and this approximately less than 10 molecules of TF per cell (one molecule/cell for *E. coli* is 1 nM)!

3. Consider the following reactions:



Define a  $K_D$  for the first reaction and the second reaction. Do you understand that those can indeed be different?

Answer:

They can for instance be different when the  $A$  that binds to  $AB$  also interacts with the  $A$  already in the complex besides its interaction with  $B$ . 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  $K_{D2}$ .

Answer:

We have,

$$\begin{aligned} b_T &= b + 2ab + a_2b = b + 2\frac{a \cdot b}{K_{D1}} + \frac{a \cdot ab}{K_{D2}} = b + 2\frac{a \cdot b}{K_{D1}} + \frac{a^2 \cdot b}{K_{D1}K_{D2}} \\ \Rightarrow b &= \frac{b_T}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{K_{D1}K_{D2}}} \end{aligned} \quad (15.8)$$

Here the 2 comes from the fact that  $B$  has two binding sites for  $A$  and two kinds of complexes of  $AB$  can exist and each needs to be counted. Since,  $a_2b = a^2 \cdot b / (K_{D1}K_{D2})$  we get,

$$a_2b = b_T \frac{\frac{a^2}{K_{D1}K_{D2}}}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{K_{D1}K_{D2}}} \quad (15.9)$$

And if  $A$  also interacts with  $A$  in  $AB$  besides its interaction with  $B$  then

$$a_2b = b_T \frac{\frac{a^2}{\alpha K_{D1} K_{D2}}}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{\alpha K_{D1} K_{D2}}} \quad (15.10)$$

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



Answer:

The total amount of  $b$  equals  $b_T = b + 3ab + 3a_2b + a_3b$ . The "3" derive from the fact that  $B$  has three binding states for  $A$  and three forms of  $AB$  and  $A_2B$  can then exist: i.e. if we mark the occupied sites of  $B$  with a "+" and an empty one with a "=" you can have  $\{+==, ==+, ==+\}$  and  $\{++=, +=+, =++\}$ . This means that we can write (following the logic of the last exercise),

$$b = \frac{b_T}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.12)$$

And we obtain

$$ab = b_T \frac{3\frac{a}{K_{D1}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.13)$$

$$a_2b = b_T \frac{3\frac{a^2}{\alpha K_{D1} K_{D2}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.14)$$

$$a_3b = b_T \frac{\frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.15)$$

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

Answer:

This requires solving

$$\frac{a_3b}{b_T} = \frac{\frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} = 0.5 \quad (15.16)$$

for a. This is a nightmare of course and we would let Mathematica do this for us.

- (b) At what concentration of  $A$  is 50% of  $B$  in the  $A_2B$  complex? This requires solving

$$\frac{a_2b}{b_T} = \frac{\frac{a^2}{\alpha K_{D1} K_{D2}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1} K_{D2}} + \frac{a^3}{\alpha \beta K_{D1} K_{D2} K_{D3}}} = 0.5 \quad (15.17)$$

for  $a$ ; let call this value  $a^*$ . This is a nightmare of course and we would let Mathematica do this for us.

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

$$\frac{ab}{b_T} = \frac{3\frac{a^*}{K_{D1}}}{1 + 3\frac{a^*}{K_{D1}} + 3\frac{a^{*2}}{\alpha K_{D1} K_{D2}} + \frac{a^{*3}}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.18)$$

$$\frac{a_3b}{b_T} = \frac{\frac{a^{*3}}{\alpha \beta K_{D1} K_{D2} K_{D3}}}{1 + 3\frac{a^*}{K_{D1}} + 3\frac{a^{*2}}{\alpha K_{D1} K_{D2}} + \frac{a^{*3}}{\alpha \beta K_{D1} K_{D2} K_{D3}}} \quad (15.19)$$

### 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  $fL$ ).

Answer:

Let's first define the dissociation constant.

$$k^- ab = k^+ a \cdot b \Rightarrow ab = \frac{k^+ a \cdot b}{k^-} = \frac{a \cdot b}{\frac{k^-}{k^+}} = \frac{a \cdot b}{K_D} \quad (15.20)$$

Thus  $K_D = k^-/k^+$ . We know that  $K_D = 1 \text{ nM}$  and  $k^+ = 1 \text{ nM}^{-1}s^{-1}$  and therefore  $k^- = 1 \text{ s}^{-1}$ . So, the complex lives  $1/k^- = 1 \text{ s}$  on average. All the other information was not required to obtain this answer. :)

### Exercise

Consider the previous section.

1. How are the dissociation constants defined in terms of rate constants?

Answer:

$$k^- ab = k^+ a \cdot b \Rightarrow ab = \frac{k^+ a \cdot b}{k^-} = \frac{a \cdot b}{\frac{k^-}{k^+}} = \frac{a \cdot b}{K_D} \quad (15.21)$$

Thus  $K_D = k^-/k^+$ .

2. Why do we need to introduce the  $\alpha$  when the regulatory sites are different or when the transcription factors can interact on the DNA?

Answer:

Transcription factor binding then depends on the state of the DNA, whether other transcription factors are already bound or not. Thus the  $K_D$  then depends on whether transcription factors are bound or not.

3. Make the derivation for  $tfptf$  (equation 2.49) yourself.

Answer:

See the exercises above for examples of those derivations; you should be able to do this by now. If you have problems then talk to one of the teachers of this course.

4. Plot the concentration of  $tfptf$  as function of the transcription factor concentration. Investigate the influence of  $K_1$ ,  $K_2$ , and  $\alpha$ .

Answer:

Use Excel or Mathematica for this. A lower  $K_1$  and  $K_2$  shifts the curve to the left and lower  $\alpha$  achieves the same outcome. This is the case because the binding sites are now given a higher affinity for the transcription factor. Check this with Excel! Increasing those parameters reduces affinity of DNA for transcription factors.

5. What does  $K_1 < K_2$  indicate?

Answer:

Site 1 has a higher affinity than site 2.

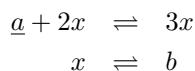
6. What does  $\alpha < 1$  indicate?

Answer:

Interactions between the bound TF and TF that is in the process of binding enhance the affinity of DNA for the TF that is in the process of binding.

## Exercises

1. Consider the following reactions,



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.

Answer: see the mathematica file *schloglmodel.nb*.

## 15.2 Answers to Chapter 3 exercises

### Exercises

1. Show for the models in figure 3.4 that they each show perfect adaptation; check for perfect adaptation of concentrations and reaction rates.

Answer:

The first network displays perfect adaptation in variable  $Y$  with respect to  $L$  because the steady state concentration of  $Y$  does not depend on  $l$ ,

$$y_s = \frac{k_3 s_2 l}{k_4 x_s} = \frac{k_3 s_2 l}{k_4 \frac{k_1 s_1 l}{k_2}} = \frac{k_2 k_3 s_2}{k_4 k_1 s_1}$$

In addition, you may not have spotted it yet; this network is a feedforward network.

The next network (upper right figure) shows robustness of  $x_s$  with respect to any of the kinetic parameters occurring in the rate equations  $v_1(x, y)$  and  $v_2(x)$  as from the steady state balance of  $Y$  we can determine  $x_s$  without making any reference to properties of reactions 1 and 2, because  $dy/dt = 0 = v_3(x_s) - v_4$ .

In the final network, all fluxes will be equal at steady state:  $J = v_1(z_s) = v_2(x_s, y_s) = v_3(y_s, z_s) = v_4(z_s)$  and, hence, from  $v_1(z_s) = v_4(z_s)$  the steady state concentration of  $Z$  can be expressed in terms of kinetic properties of enzyme 1 and 4 without making reference to the properties of reaction 2 and 3. Thus  $z_s$  is robust to all the kinetic properties of reaction 2 and 3. Since, the steady state flux  $J$  only depends on  $Z$ , i.e.  $J = v_4(z_s)$  also the flux is robust to all the properties of reaction 2 and 3. Note that these robustness conclusions do not depend on whether the feedback of  $Z$  on reaction 1 is present or not. The only requirement is that reaction 1 is insensitive to its product.

### Exercises

1. Will two fold change detection systems in series give rise to fold change detection of the output of the second system with respect to the signal of the first?

Answer:

If the first network displays fold change detection its output also shows perfect adaptation. So, the second network, which is also a fold change detector, has a transient input only and not a step input as the first network does. Hence, now it is a matter of time-scales how the second system is going to respond. If you are interested in working this out then make a model in Mathematica and see what comes out and study the influence of time-scale separation between the first and the second network.

### Exercises

1. Show that with two irreversible reactions the error rate becomes:

$$\Phi = \frac{b}{a} \left( \frac{K_A}{K_B} \right)^3$$

Answer:

When a student sends me the answer I will add it to the exercise section.  
It follows exactly the same reasoning as above, nothing new.

2. And with  $n$  irreversible reactions,

$$\Phi = \frac{b}{a} \left( \frac{K_A}{K_B} \right)^{n+1}$$

Answer:

When a student sends me the answer I will add it to the exercise section.

3. What is the negative consequence of kinetic proofreading? (Hint: i. a delay is introduced with every new irreversible reaction and ii.  $A$  also dissociates during this delay time. What is the net outcome of both of these phenomena?)

Answer:

The system gets a very low rate of product formation when it is very precise.

### Exercises

1. Is the equilibrium binding assumption for  $S$  and  $L$  required for the robustness of  $RP$  with respect to the total sensor and response regulator level?

Answer:

Yes, otherwise the derivation does not lead to robustness; the equilibrium assumption was explicitly used in the derivation. I invite you to study this more carefully and the derivation from perfect robustness depends on the deviation from thermodynamic equilibrium of the binding reaction. Is this dependency strong or mild?

### 15.3 Answers to Chapter 4 exercises

#### Exercise

1. Perform the following conversions:

(a)  $pM$  to  $nM$  ( $pM = \text{picoliter} = 10^{-12}$ ).

$$\text{Answer: } 1 \text{ } pM = 10^{-12} M = 10^{-3} 10^{-9} M = 10^{-3} nM.$$

(b)  $m^3$  to  $\mu\text{l}$ .

Answer: We have  $1 \text{ } m^3 = (10\text{dm})^3 = 1000 \text{ l}$  and  $1 \text{ } \mu\text{l} = 10^{-6} \text{ l}$ ; so,  $1 \text{ l} = 10^6 \text{ } \mu\text{l}$ . Thus,  $1000 \text{ l} = 1000 \cdot 10^6 \text{ } \mu\text{l} = 10^9 \text{ } \mu\text{l}$ .

(c)  $pM$  to  $\text{molecules} \cdot \text{l}^{-1}$ .

$$\text{Answer: } 1 \text{ } pM = 10^{-12} M = 10^{-12} \frac{\text{mol}}{\text{l}} = 10^{-12} \cdot 6 \cdot 10^{23} \frac{\text{molecules}}{\text{mol}} \frac{\text{mol}}{\text{l}} = 6 \cdot 10^{11} \frac{\text{molecules}}{\text{l}}$$

(d)  $\text{dm}^2/\text{min}$  to  $\mu\text{m}^2/\text{s}$ .

$$\text{Answer: } \frac{\mu\text{m}^2}{\text{s}} = \frac{\mu\text{m}^2}{\text{dm}^2} \frac{\text{min}}{\text{s}} \frac{\text{dm}^2}{\text{min}} = \left( \frac{\mu\text{m}}{\text{dm}} \right)^2 \frac{1}{60} \frac{\text{dm}^2}{\text{min}} = \left( \frac{10^{-6} \text{ m}}{10^{-1} \text{ m}} \right)^2 \frac{1}{60} \frac{\text{dm}^2}{\text{min}} = (10^{-5})^2 \frac{1}{60} \frac{\text{dm}^2}{\text{min}} = 10^{-10} \frac{1}{60} \frac{\text{dm}^2}{\text{min}}$$

(e) 1  $\text{molecules}/\text{cell}$  to  $nM$  if the cell has a volume of 1  $\text{fl}$  ( $\text{fl} = \text{femtoliter} = 10^{-15} \text{ l}$ ;  
is approximately the volume of *E. coli*).

$$\text{Answer: } \frac{\text{mol}}{\text{l}} = \frac{\text{mol}}{\text{molecules}} \frac{\text{cell}}{\text{l}} \frac{\text{molecule}}{\text{cell}} = \frac{1}{6 \cdot 10^{23}} 10^{15} \frac{\text{molecule}}{\text{cell}} = \frac{1}{6} 10^{-8} \frac{\text{molecule}}{\text{cell}}$$

$$\text{Because } 1M = 10^9 nM, 1 \text{ } nM = 10^9 \frac{1}{6} 10^{-8} \frac{\text{molecule}}{\text{cell}} = 10/6 \frac{\text{molecule}}{\text{cell}}$$

2. The volume of *E. coli* is approximately  $1 \text{ } \mu\text{m}^3$ . Assume that *E. coli* is a sphere.

1. Calculate the concentration of one molecule per cell in  $nM$ .

Answer: The concentration of 1 molecule per *E. coli* cell is,

$$\frac{1 \text{ molecule}}{1 \text{ } \mu\text{m}^3 6 \cdot 10^{23} \text{ molecules/mol}} = 1/(6 \cdot 10^{23}) \text{ mol } \mu\text{m}^{-3} = 1.67 \text{ } nM$$

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

Answer: The radius of *E. coli* is,

$$r = \left( \frac{V}{4/3\pi} \right)^{1/3} = 0.6 \text{ } \mu\text{m}$$

The area is,

$$A = 4\pi r^2 = 4.5 \text{ } \mu\text{m}^2$$

The number of receptors is,

$$\frac{\text{area membrane}}{\text{area receptor}} = \frac{4.5 \cdot 10^{-12} \text{ m}^2}{4\pi(5 \cdot 10^{-9})^2 \text{ m}^2} = 14324 \text{ receptors}$$

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

Answer: The number of proteins in the cytoplasm that have the same size as the receptor,

$$\frac{\text{volume cytoplasm}}{\text{volume receptor}} = \frac{1 \cdot 10^{-18} \text{ m}^3}{4/3 \cdot \pi (5 \cdot 10^{-9})^3 \text{ m}^3} = 1.9 \cdot 10^6 \text{ receptors/cytoplasm}$$

So about 1/4 million in *E. coli* if each macromolecule is at 1 macromolecule distance

### Exercise

1. Equation 4.2 gives the probability distribution for the particles at some location along a line and the distribution broadens over time. Often, it is more convenient to picture the number of molecules than probability. Equation 4.2 can also be written as,

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

With  $N$  as the total number of diffusing particles and  $n(\Delta, t)$  as the number of particles at distance  $\Delta$  and time  $t$ . This means that the probability  $p(\Delta, t)$  equals  $\frac{n(\Delta, t)}{N}$ .

- (a) Explain why this probability definition indeed makes intuitive sense.

Answer: The probability equals the fraction of molecules in this manner and this makes sense. The number of molecules to observe then equal the probability times the total number of molecules. If more molecules are at a certain distance from the origins then is more likely to find molecules there. This is the frequentist interpretation of probability.

- (b) If  $D = 20 \mu\text{m}^2/\text{s}$  how much time does it take before  $2/3$  of the particles are more than  $5 \text{ cm}$  away from  $\Delta = 0$ ? Those would be realistic times for an ion diffusing through an axon by diffusion.

Answer: Let's first convert perform a unit conversion  $5 \text{ cm} = 5 \cdot 10^4 \mu\text{m}$ . Then we would like to at what the probability for observing particles further away than  $5 \text{ cm}$  equals  $2/3$ . So, we are looking for the time,  $\tau$ , such that,

$$p(|\Delta| > 5 \text{ cm}, \tau) = \int_{-\infty}^{-5} \frac{n(\Delta, \tau)}{N} d\Delta + \int_5^{\infty} \frac{n(\Delta, \tau)}{N} d\Delta = \text{Erfc}\left(\frac{5}{\sqrt{4D\tau}}\right) = \frac{2}{3}$$

The integral was evaluated Mathematica and the Erfc means complementary error function. The next question for what time does this relation hold? This is plotted in the figure 15.1. An estimate from the figure shows after about 3.25 seconds the probability to observe

particles at a distance further than 5 cm equals  $2/3$ . Why doesn't your brain function like this?

Answer: A diffusion limited brain would be much too slow in making computations.

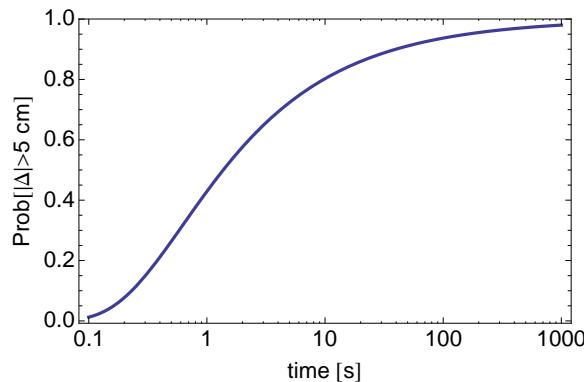


Figure 15.1: Probability to observe molecule further away than 5 cm from the origin as function of time with a diffusion coefficient of  $20 \mu\text{m}^2/\text{s}$ .

## Exercise

### 1. Diffusion of GFP

- (a) Plot the diffusion coefficient as function of the radius of a spherical particle. Take  $30^\circ\text{C}$  and express this temperature in units Kelvin.

Answer: Temperature is 303.15 K and for the viscosity we take  $10^{-3} \text{ Pas}$  and the Boltzmann constant equals  $1.38 \text{ m}^2\text{kgs}^{-2}\text{K}^{-1}$ . Blue line in Figure 15.2.

- (b) Double the temperature and make the same plot.

Answer: Red line in Figure 15.2. Thus the temperature is not so important.

### 2. What do you think $kT$ means? (Check its units).

Answer:  $kT$  is the amount of energy one molecule has to spend to diffuse.

### 3. What is the unit of $f$ ? What do you think this quantity means?

Answer:  $6\pi\eta a$  has as unit  $\text{kg} * \mu\text{m}^{-1}\text{s}^{-1}\mu\text{m} = \text{kg/s}$ .  $6\pi\eta a$  is the friction or drag coefficient it is related to the force a particle experiences when moving particular speed. In order to overcome this force the particle requires energy and it has  $kT$  to spend. If it loses more energy in overcoming the drag force it will move slower; its diffusion coefficient will be slower

4. The dimension of an average macromolecule is 5 nm in diameter. Calculate the diffusion coefficient using  $\eta = 10^{-3} \text{ Pa s}$  and  $kT = 4 * 10^{-21} \text{ J}$ . Is this a realistic value?

Answer:  $1 \text{ Pa/s} = 1 \text{ kg m}^{-1} \text{ s}^{-1}$

$$D = \frac{kT}{6\pi\eta a} = \frac{4 * 10^{-21} \text{ kg m}^2 \text{ s}^{-2} (= J)}{6 \pi 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1} 5 * 10^{-9} \text{ m}} = 4.2 * 10^{-11} \text{ m}^2/\text{s} = 42 \mu\text{m}^2/\text{s}$$

And this fits nicely with GFP in mammalian cytoplasm ( $\approx 25 \mu\text{m}^2/\text{s}$ ).

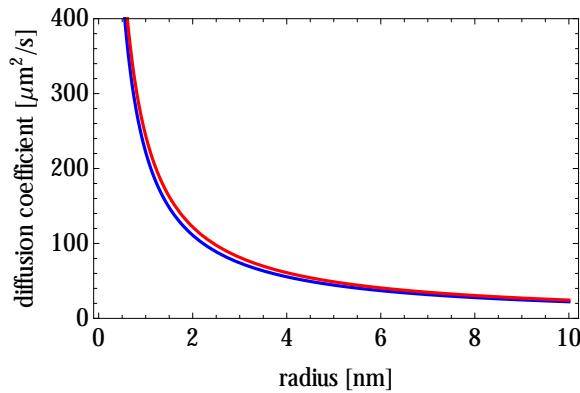


Figure 15.2: Diffusion coefficient as function of the radius of a molecule.

## Exercises

- Can a second-order rate constant be higher than the diffusion limit?  
Answer: no, the diffusion limit is the fastest limit.
- Use the diffusion-limited second order rate constant relationship to estimate the binding rate of the lac repressor to its DNA target site in  $M^{-1}s^{-1}$ . The diffusion coefficient of the receptor is  $5 * 10^{-7} \text{ cm}^2/\text{s}$ . The average radius of the receptor is  $40 * 10^{-10} \text{ m}$  and that of the DNA site  $10 * 10^{-10} \text{ m}$ .

Answer: The diffusion-limited (second-order) rate constant can be calculated from,

$$k_a = 4\pi(D_{\text{repressor}} + D_{\text{DNA}})(r_{\text{repressor}} + r_{\text{DNA}})$$

Taking:  $D_{\text{repressor}} = 5 \mu\text{m}^2/\text{s}$ ,  $D_{\text{DNA}} = 0$ , and  $r_{\text{repressor}} = r_{\text{DNA}} = 2.5 \text{ nm}$ , we calculate for  $k_a$ ,

$$k_a = 4\pi 5 \cdot 5 \cdot 10^{-3} = 0.314 \text{ molecule}^{-1} \mu\text{m}^3 \text{ s}^{-1} \quad (15.22)$$

Multiplied by Avogadro's number makes the units  $\text{mol}^{-1} \mu\text{m}^{-3} \text{ s}^{-1}$  and divided by  $10^{15} \mu\text{m}^3/\text{l}$  gives  $\text{M}/\text{s}$ , which yields  $0.2 * 10^9 M^{-1} s^{-1} = 0.2 nM^{-1} s^{-1}$

### Exercises

- Study Figure 4.5 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 much longer do these processes take for a bigger cell with 1000 times the radius of *E. coli*.

Answer: With  $R_{cell} = 1 \mu m$ ,  $R_r = 5 nm$ ,  $V_{cell} = 4.2 \mu m^3$ , and  $D = 5 \mu m^2/s$ ; the time to hit the membrane is, to find a molecule in the membrane, to travel the radius of the cell, and to find another molecule in the cytoplasm is,

$$\begin{aligned}\tau &= \frac{R_{cell}^2}{15D} = \frac{0.6^2}{15 \cdot 5} = 0.005 s \\ \tau &= \frac{V_{cell}}{4DR_r} = \frac{4.2}{1 \cdot 5 \cdot 10 \cdot 10^{-3}} = 13 s \\ \tau &= \frac{R_{cell}^2}{6D} = \frac{0.6}{6 \cdot 5} = 0.02 s \\ \tau &= \frac{V_{cell}}{4\pi R_r^2 D} = \frac{1}{4\pi 10 \cdot 10^{-3} \cdot 2 \cdot 5} = 7 s\end{aligned}$$

The times that depend linearly on the radius will increase by a factor of 1000 and those that depend linearly on volume increase by a factor of a  $10^9$ , as  $V = 4/3\pi r^3$ .

- 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  $80 \mu m$ . 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?

Answer: We can calculate the volume from  $V = 4/3\pi R^3 = 4/3\pi 80^3 = 2.14 * 10^6 \mu m^3$ . The time for a single cytosolic protein to find a membrane protein is  $\tau = V/(4DR_r) = 2.14 * 10^6 / (4 * 5 * 10 * 10^{-3}) = 10 s$ . Here we took a  $D$  of  $5 \mu m^2/s$  and a protein radius of  $5 nm$ . The association rate equals  $v = k^+ * [M] * [C]$  with  $[M]$  as the membrane receptor and  $[C]$  as the cytosolic protein concentrations. The rate constant  $k^+$  equals  $(1/\tau \cdot V_{cell}/molecules)$ . The multiplication with  $1/concentration = V_{cell}/molecules$  makes sure this rate constant has as unit  $[concentration^{-1} \cdot time^{-1}]$ . The association rate equals now  $v = 1/\tau [s^{-1}] * V_{cell}/molecules * 300 molecules/V_{cell} * 1000 molecules/V_{cell} = 0.013 molecules/(V_{cell}s)$ . If a single complex lives for 30 seconds on average, the dissociation rate constant is  $k^- = 1/30s$ . At equilibrium  $k^+ * [M] * [C] = k^- * [MC]$ , the equilibrium constant  $K$  is now defined as  $[MC]/([M][C]) = k^-/k^+ =$

$2.56 V_{cell}/molecules = K$ . Since,  $M_T = [M] + [MC]$  and  $C_T = [C] + [MC]$  give the conservation of C and M, we have with  $[MC]/([M][C]) = K$  three equations with three unknown equilibrium concentrations  $[M]$ ,  $[C]$ , and  $[MC]$ . Solving this set of equation gives for their values at equilibrium:  $MC = 299.83$ ,  $C = 0.17$ , and  $M = 700.167$ . You can obtain this result in Mathematica by evaluating  $Solve[\{300 == C + MC, 1000 == M + MC, MC/(M*C) == 2.56\}, \{M, C, MC\}]$  or by solving this system by hand.

3. Can an average reaction time be shorter than an average diffusion time?  
Answer: No, reaction time is always equal or longer than a diffusion time.
4. A mammalian cell has typically a dimension of  $10000 \mu m^3$ . 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?

Answer: The radius of the cell is  $R = (4/3\pi V)^{1/3} = 13.4 \mu m$ . The concentration of 25000 receptors in a cell of  $1000 \mu m^3$  equals  $(25000/(6 * 10^{23}) mol)/1000 \mu m^3 = 4.1 * 10^{-24} mol/\mu m^3 = 4nM$ . The time to diffuse a distance of  $13.4 \mu m$  is  $t = d^2/(6D) = 13.4^2/(6 * 2) = 15 s$ . The radius of the receptors is  $10 nm$ , their volume is therefore  $4/3\pi R^3 = 524 nm^3 = 524 * 10^{-9} \mu m^3$ . So  $10000/(524 * 10^{-9}) = 1.9 * 10^{10}$  receptors fit inside a eukaryotic cell. The time for  $N_A$  androgen receptors to  $N_T$  transporters is  $0.9 * V_{cell}/(4\pi R_r(D_A + D_T)N_A N_T) = 7.2 * 10^{-5} s$ . The rate constant is  $(4\pi R_r(D_A + D_T))$ ; this has as units  $concentration^{-1}s^{-1}$  where concentration is molecule per cell volume. A single androgen receptor finds one of the 500 nuclear targets in  $V_{nucleus}/(4\pi R_r D_A N_{targets}) = 3.97 s$ . The dissociation rate constant  $k^-$  equals  $1/50 s^{-1}$ . The association rate constant  $k^+$  is  $4\pi R_r D_A N_{targets} = 251 s^{-1}$ . The dissociation constant is  $AT/(A * T) = k^+ / k^- = K_D = 251/(1/50) = 12550 V_{nucleus}/molecules$ .

## 15.4 Answers to Chapter 6 exercises

### Exercise

1. Determine the change in the substrate concentration when the enzyme rate changes from 10% to 90% of the maximal value.

Answer: We have the following relationship  $\frac{v}{V_{max}} = \frac{\frac{s}{K_M}}{1 + \frac{s}{K_M}}$  and we would like to know  $\frac{s}{K_M}$  when  $\frac{v}{V_{max}}$  equals 0.1 and 0.9. Rewriting the Michaelis-Menten relationship gives  $\frac{s}{K_M} = \frac{\frac{v}{V_{max}}}{1 - \frac{v}{V_{max}}}$ , which gives for 0.11 and 9 for  $\frac{s}{K_M}$ ; this means that the concentration of  $s$  should by a factor of about 90 to change the rate by a factor of 9! This shows why Michaelis-Menten are such poor enzymes for regulatory purposes. Cooperative enzymes do a much better job as we shall see later.

2. Describe the quasi-steady state assumption in your own words.

Answer: The quasi-steady state assumption states that the enzyme species remain fixed in time during a relatively long period while the substrate concentration drops and the product concentration rises; the enzyme is continuously busy with converting substrates into product. This can be achieved by having a much higher total concentration of substrate than enzyme.

### Exercise

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

Answer: The equilibrium binding assumption as it assumed every binding reaction to the enzyme to be in equilibrium whereas the quasi-steady state assumption only assumed more substrate than enzyme.

### Exercises

1. Derive the kinetics in the case that  $X$  cannot bind to  $ES$  but only to  $E$ .

Answer: then we start from  $e_T = e + ex + es = e\left(1 + \frac{x}{K_1}\right) + es$ , which is the case treated in the theory section if  $K_2 \rightarrow \infty$ ; which means that  $X$  cannot bind to  $ES$ . Hence, the  $V_{MAX}$  is not influenced by the inhibitor and the rate equation can be written as,

$$v = V_{MAX} \frac{s}{s + K_S \left(1 + \frac{x}{K_1}\right)} = V_{MAX} \frac{\frac{s}{K_S}}{1 + \frac{s}{K_S} + \frac{x}{K_1}} \quad (15.23)$$

This equation shows that in this case the inhibitor increases the  $K_M$  of the enzyme for the substrate and more substrate is required to attain the same rate.

2. Derive the kinetics in the case that  $X$  cannot bind to  $E$  but only to  $ES$ .

Answer: then we start from  $e_T = e + es + esx = e + es \left(1 + \frac{x}{K_2}\right)$ , which is the case treated in the theory section if  $K_1 \rightarrow \infty$ . Hence, only the  $V_{MAX}$  is influenced by the inhibitor and the rate equation can be written as,

$$v = V_{MAX} \frac{\frac{s}{K_S}}{1 + \frac{s}{K_S} \left(1 + \frac{x}{K_2}\right)} = V_{MAX} \frac{\frac{s}{K_S}}{1 + \frac{s}{K_S} + \frac{s \cdot x}{K_S K_2}} \quad (15.24)$$

This equation shows that the inhibitor reduces the rate of the enzyme and this can be overcome by adding more substrate provided that  $\frac{s}{K_s} \gg \frac{s \cdot x}{K_S K_2}$ .

3. Compare the two equations that you have derived in the previous two exercises. One of these mechanisms is called competitive inhibition. Which one do you think and why?

Answer: When  $X$  can bind to  $E$  only competitive inhibition occurs because then  $X$  can no longer bind to the enzyme, so  $X$  must bind to the same site as  $S$ ; hence, they compete for binding.

## Exercises

1. Consider equation 6.42 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  $S$ ?

Answer: By setting the product concentration to zero equation 3.12 simplifies to:

$$v = \frac{V_{MAX}^+ * \frac{s}{K_{MS}}}{1 + \frac{s}{K_{MS}}}$$

A given reaction with rate  $v$  is the number of molecules of product formed (or substrate consumed) per second per amount of the enzyme. The reaction rate increases with increasing substrate concentration  $[s]$ , asymptotically approaching the maximum rate  $V_{MAX}$ . There is therefore no clearly-defined substrate concentration at which the enzyme can be said to be saturated with substrate. A more appropriate measure to characterize an enzyme is the substrate concentration at which the reaction rate reaches half of its maximum value ( $\frac{V_{MAX}}{2}$ ). The rate of this enzyme is most sensitive when the concentration of  $s$  is much smaller than the  $K_{MS}$ , i.e when  $s$  is approaching zero:  $\lim_{s \rightarrow 0}$

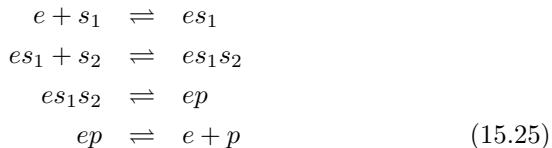
2. Plot the rate of an enzyme modelled with equation 6.42 as function of  $S$  for constant values of  $P$  (0.25, 0.75, 1.5, 7.5) take a  $V_{MAX}^+$  of 10 mM/min,  $K_S$  of 0.1 mM,  $K_P$  of 0.75 mM, and an equilibrium constant of 1000. Find the concentration of  $S$  where the enzyme is in thermodynamic equilibrium and check equation 6.42.

Answer: Equilibrium concentrations for  $s$  are: 0.00025, 0.00075, 0.0015 and 0.0075, e.g.  $s = \frac{p}{K_{eq}}$ . See mathematica file: equilibrium plot.nb.

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.

Answer: See mathematica file: QSS assumption enzyme kinetics.nb

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,



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?

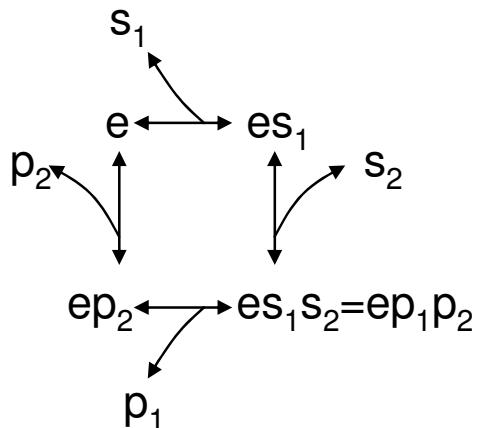
Answer: See mathematica file: ordered bi uni enzyme kinetics.nb

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

Answer: See Figure 15.4.

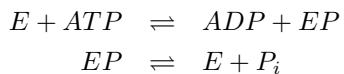
## Exercises

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.  
Answer: See mathematica file: equilibrium enzyme and control.nb
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.

Answer: The reactions catalyzed by the kinase and phosphatase are:



The overall (driving) reaction is thus:  $ATP \rightleftharpoons ADP + Pi$ . When the ATP concentration is not fixed it has the following differential equation:  $\frac{dATP}{dt} = -V_{kinase}$ , from this it follows that ATP can only be at steady state when the rate of the kinase is zero (when the system is in thermodynamic equilibrium). At thermodynamic equilibrium the concentration of the substrate and product are solely dependent on the equilibrium constant, there is no influence from the other kinetics parameters.

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?

Answer: The equilibrium constant is dimensionless because products and

substrates have the same units and therefore they cancel-out.

$$adp_{eq} = 4.12$$

$$amp_{eq} = 1.94$$

$$atp_{eq} = 3.94$$

The ATP-ADP ratio is not dependent on the initial amount of adenosine and phosphate. See mathematica file: AK-enzyme.nb

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  $K_m$ 's to 1 and  $V_{max}$ 's to 10. In the first model, you set the equilibrium constant of the second enzyme 1 and determine the  $V_{max}$  of this enzyme to have it operate 10% from thermodynamic equilibrium at steady state by judging  $\Gamma/K_{eq}$ . In the second model, set the equilibrium constant to a 100 and determine again the value for the  $V_{max}$  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  $V_{max}$  is perturbed?

Answer: See mathematica file: equilibrium enzyme and control.nb

### Exercises

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

Answer: Uncompetitive inhibition is a more potent mechanism. This is because uncompetitive inhibition cannot be overcome by adding more reactant concentration, which is the case for competitive 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  $K_m$ 's to 1, all  $V_{max}^+$  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?

Answer: See mathematica file: metabolic pathway with product independent first reaction.nb

### Exercises

In figure 6.5, 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,

$$v = V_{max} \frac{\frac{s_o}{K_m} - \frac{s_i}{K_m}}{1 + \frac{s_o}{K_m} + \frac{s_o}{K_m} + K_i \frac{s_i s_o}{K_m^2}} \quad (15.26)$$

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.

Answer: See mathematica file: "transporter kinetics.nb"

## 15.5 Answers to Chapter 7 exercises

### Exercises

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

$$v = V_{MAX} \frac{\text{pep} \cdot \text{adp} \left( \frac{\text{pep}}{K_{pep}} + 1 \right)^n}{K_{pep} \left( L \left( \frac{1 + \frac{\text{atp}}{K_{atp}}}{\frac{\text{fdp}}{K_{fdp}} + \frac{\text{amp}}{K_{amp}} + 1} \right)^n + \left( \frac{\text{pep}}{K_{pep}} + 1 \right)^n \right) (\text{adp} + K_{adp})} \quad (15.27)$$

The kinetic parameters are:  $K_{pep} = 0.31 \text{ mM}$ ,  $K_{adp} = 0.26 \text{ mM}$ ,  $K_{amp} = 0.2 \text{ mM}$ ,  $K_{fdp} = 0.19 \text{ mM}$ ,  $K_{atp} = 22.5 \text{ mM}$ ,  $L = 1000$ , 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 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.

Answer: See mathematica file "pyruvate kinase MWC mechanism.nb"

- The intricate regulation of glycolysis and occurrence of the cooperative enzyme, phosphofructokinase, pyruvate decarboxylase, and puryvate kinase, inspired the analysis of kinetic models of glycolysis. A glycolysis model was developed by Goldbeter and Lefever [15]. 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:  $\text{fructose - 6 - phosphate} + \text{ATP} \rightleftharpoons \text{ADP} + \text{fructose - 1, 6 - bisphosphate}$ . In figure 7.4 the network diagram of this model is shown. Here  $\gamma$  indicates ADP and  $\alpha$ , fructose-6-phosphate. PFK is an allosteric enzyme, modelled with a MWC mechanism,

$$v_2 = \sigma_M \frac{\alpha e (1 + \alpha e)^{n-1} (1 + \gamma)^n + L \theta \alpha e' (1 + \alpha e')^{n-1}}{L (1 + \alpha e')^n + (1 + \gamma)^n (1 + \alpha e)^n} \quad (15.28)$$

Where  $e = (1 + \epsilon)^{-1}$  and  $e' = (1 + \epsilon')^{-1}$  with  $\epsilon$  and  $\epsilon'$  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? Answer: See mathematica file "goldbeter model.nb".

## 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,  $p$ ,

$$v = \frac{\frac{V_f s}{s_{0.5}} \left(1 - \frac{s}{x_1 K_{eq}}\right) \left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^{n-1}}{\left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^n + \frac{1 + \left(\frac{p}{p_{0.5}}\right)^n}{1 + \alpha \left(\frac{p}{p_{0.5}}\right)^n}} \quad (15.29)$$

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 [19, 20].

Answer: See mathematica file "negative feedback metabolic pathway.nb".

## 15.6 Answers to Chapter 8 exercises

### 15.6.1 Exercise

1. What is the advantage of a dimeric sensor over a monomeric sensor?

Answer: A dimeric sensor is more sensitive to the signal, i.e. the fully bound state depend in a quadratic fashion on the signal concentration.

2. What does the exponent of '2' signify in equation 8.2?

Answer: the number of subunits.

3. Derive the Hill equation from equation 8.7.

Answer: We start from,

$$r_2 s_2 = e_{2T} \frac{\frac{s^2}{K_R^2}}{\left(1 + \frac{s}{K_R}\right)^2 + L \left(1 + \frac{s}{K_T}\right)^2} \quad (15.30)$$

And make  $L = 0$ ,

$$r_2 s_2 = e_{2T} \frac{\frac{s^2}{K_R^2}}{\left(1 + \frac{s}{K_R}\right)^2} \quad (15.31)$$

We have to add allosteric interactions now between the sites such that

$$r_2 s_2 = e_{2T} \frac{\frac{s^2}{\alpha K_R^2}}{1 + 2\frac{s}{K_R} + \frac{s^2}{\alpha K_R^2}} \quad (15.32)$$

And then assume  $\alpha$  to be really small; a very strong allosteric interaction,

$$r_2 s_2 \approx e_{2T} \frac{\frac{s^2}{\alpha K_R^2}}{1 + \frac{s^2}{\alpha K_R^2}} = e_{2T} \frac{\frac{s^2}{K^2}}{1 + \frac{s^2}{K^2}} \quad (15.33)$$

Where the last equation is Hill-type equation.

## 15.7 Answers to Chapter 9 exercises

### 15.7.1 Exercise

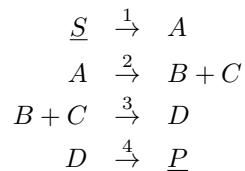
Determine the stoichiometric matrices for the reaction systems described in the exercises of section 2.2.2.

Answer:

$$\begin{aligned}
 S &= X = P \\
 \begin{pmatrix} & v_1 & v_2 \\ \dot{s} & -1 & 0 \\ \dot{x} & 1 & -1 \\ \dot{p} & 0 & 1 \end{pmatrix} \\
 \underline{S} &= \underline{X} = \underline{P} \\
 \begin{pmatrix} & v_1 & v_2 \\ \dot{x} & 1 & -1 \\ \dot{p} & 0 & 1 \end{pmatrix} \\
 3A &\rightleftharpoons 2B + C \\
 B &\rightleftharpoons 2D \\
 2C &\rightleftharpoons 3E \\
 \begin{pmatrix} & v_1 & v_2 & v_3 \\ \dot{a} & -3 & 0 & 0 \\ \dot{b} & 2 & -1 & 0 \\ \dot{c} & 1 & 0 & -2 \\ \dot{d} & 0 & 2 & 0 \\ \dot{e} & 0 & 0 & 3 \end{pmatrix}
 \end{aligned} \tag{15.34}$$

### Exercise

- To see whether you have understood the previous section, you can try to determine the  $\mathbf{L}$  and the  $\mathbf{K}$  matrix for the following "strange" metabolic network composed out the reactions,



$S$  and  $P$  are underlined to denote that they are fixed.

Answer: When you have drawn the network you will quickly conclude that at steady state all the four fluxes are equal – only then all the metabolite concentrations are at steady state. Thus we conclude that there is one independent flux. Since, we have four reactions in total we must have

three independent intermediates because the number of fluxes minus the rank of the stoichiometric matrix equals the number of independent fluxes, i.e.  $4-3=1$ . But we have four variable metabolites – not 3 – and thus we must have one conserved moiety. Indeed the mass balances of  $B$  and  $C$  are linearly dependent:  $dB/dt = dC/dt = v_2 - v_3$ . This means that  $C - B$  is conserved; this makes sense because during the dynamics of the system the amount  $B$  and  $C$  produced and consumed will always be equal and, therefore, their difference will remain constant. Thus we have one conserved moiety, three independent concentrations, three dependent fluxes, and one independent fluxes.

- Determine the  $\mathbf{L}$  and the  $\mathbf{K}$  matrix for the following network involved in signal transduction,



ERK is a signalling protein in the so-called MAPK pathway of mammalian cells that has kinase activity when it is doubly phosphorylated. In this state, it translocated to the nucleus to activate nuclear transcription factors.  $ATP$ ,  $ADP$ , and  $Pi$  were considered fixed here and not shown.

Answer:  $ERK$  is consumed by reaction 1 and produced by 4; thus at steady state  $v_1 = v_4$  and  $ERKPP$  is produced by  $v_2$  and consumed by  $v_3$  and therefore  $v_2 = v_3$  at steady state. Thus we have two independent fluxes, out of the four fluxes that occur. Because of this and because we have three variable concentrations we must have one moiety because the number of fluxes (4) minus the number of independent variables (the unknown) gives the number of independent fluxes (2). Indeed there is one conserved moiety as  $ERK$  is only converted into different forms and no net synthesis nor degradation occurs of  $ERK$ . Thus the conserved moiety is  $erk_T = erk + erkP + erkPP$ .

- How many independent flux occur in the two-component signalling network shown in Figure 3.8? Given this knowledge can there be more conserved moieties in addition to the total sensor and response regulator concentration?

Answer: Inspection of the network shows that at steady state all the fluxes are equals and that we therefore have one independent flux. We only need to know this flux value to determine all the flux values. We have 6 fluxes in total and one independent fluxes, so we must have 5 independent intermediates. We have 7 variable concentrations in total, so we must have 2 conserved moieties; those are the total sensor concentration,  $sensor_T = s + sl + sp + spr + rps$  and the total response regulator level,  $regulator_T = spr + rp + r + rps$ . So we cannot have more moieties than those mentioned.

### Exercise

- Identify the flux modes of the network displayed in Figure 9.5. Construct a new valid kernel matrix  $\mathbf{K}$  that has only positive entries by taking a linear combination of the columns of the kernel matrix  $\mathbf{K}$  given in Equation 9.35. Answer: Their exist two flux modes; one involves the linear path from  $S$  to  $P_1$  via reactions: 1, 2, 3, and 4 and the other from  $S$  to  $P_2$  via reaction 1, 2, 5, 6, and 7. Alternatively, one can define a flux route from the linear combination of the previous two routes giving rise to a path from  $P_2$  to  $P_1$  and reaction 5, 6, and 7 then run in their negative direction and 3 and 4 run in their positive direction; this path corresponds to the second column of the  $\mathbf{K}$  in equation 9.35. A new and more logical flux mode can be obtained by the addition of the columns of  $\mathbf{K}$  to give rise to  $(1, 1, 1, 1, 0, 0, 0)$  which corresponds to the path from  $S$  to  $P_1$ .
- In Figure 9.6 the TCA cycle is shown, which is an important pathway in the central metabolism of cells as it generates precursors for growth and electron carriers for oxidative phosphorylation, which makes ATP. The kernel matrix of this network is given by,

$$\mathbf{K} = \begin{pmatrix} & \text{PEPCL} & \text{KDH} & \text{KGout} \\ \text{PEPCL} & 1 & 0 & 0 \\ \text{KDH} & 0 & 1 & 0 \\ \text{KGout} & 0 & 0 & 1 \\ \text{PK} & 0 & 1 & 1 \\ \text{PDH} & 0 & 1 & 1 \\ \text{CS} & 0 & 1 & 1 \\ \text{OAAout} & 1 & 0 & -1 \\ \text{ACNB} & 0 & 1 & 1 \\ \text{IDH} & 0 & 1 & 1 \\ \text{SCS} & 0 & 1 & 0 \\ \text{SDH} & 0 & 1 & 0 \\ \text{FUMA} & 0 & 1 & 0 \\ \text{MDH} & 0 & 1 & 0 \end{pmatrix} \quad (15.36)$$

This matrix indicates that if  $PEPCL$ ,  $KDH$ , and  $KGout$  are chosen as independent fluxes all flux values can be determined from their flux values. Show that each of the columns of the  $\mathbf{K}$  matrix represents a valid steady state route of the network by drawing them in Figure 9.6. This is the basic interpretation of the nullspace of the stoichiometric matrix. However, because the nullspace is not unique, the pathways obtained directly from MATLAB or Mathematica may look counterintuitive and a linear combination of them would yield more intuitive valid steady-state flux distributions as pathway routes.

Answer: See Figure 15.3.

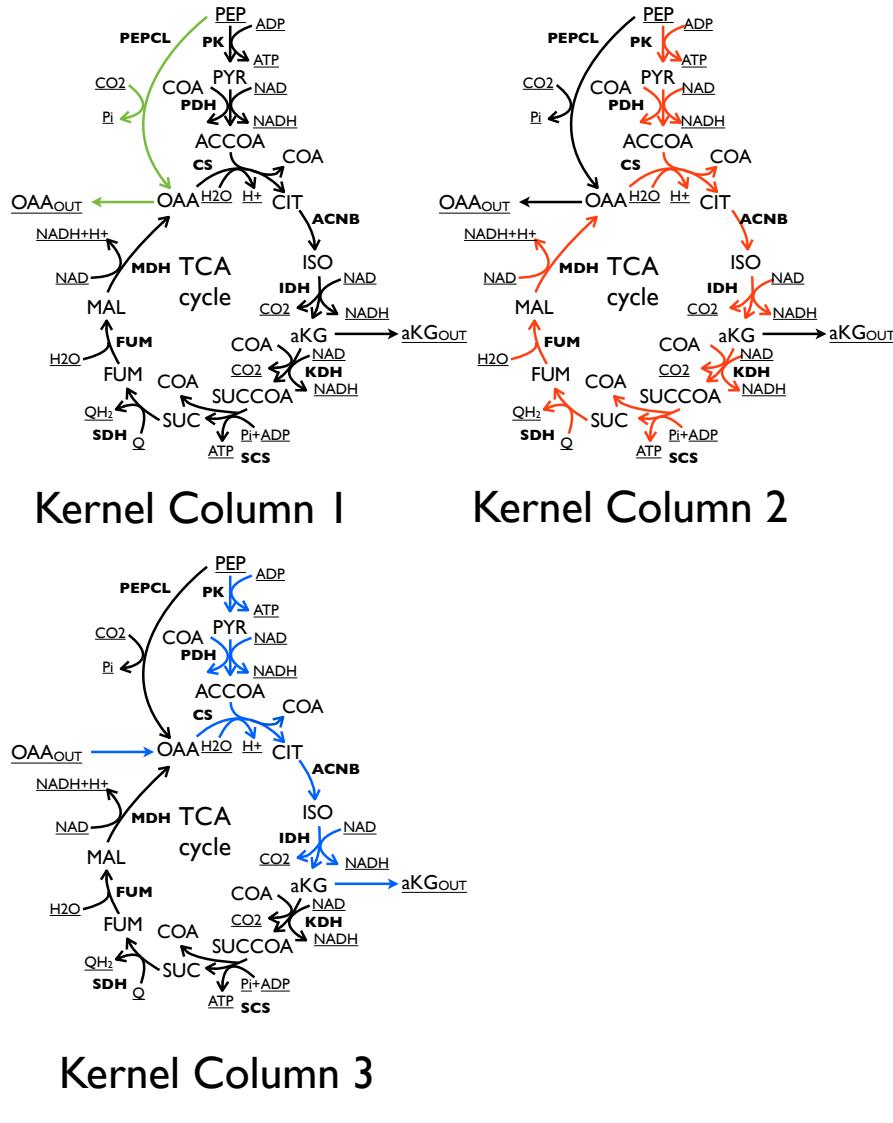


Figure 15.3: The kernel columns visualized for the TCA cycle.

### Exercises

- (a) Use Mathematica to derive the kernel matrix  $\mathbf{K}$  for the network displayed in Figure 9.9.  
Answer: See mathematica file: *trypKmatrix.nb*.
- (b) Determine the  $\mathbf{N}$ ,  $\mathbf{K}$ , and  $\mathbf{L}$  matrix for the four metabolic networks

displayed in Figure 9.8.

Answer:

Network A (15.37)

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 \\ \dot{x} & 1 & -1 & 0 \\ \dot{a} & -1 & 0 & 1 \\ \dot{b} & 1 & 0 & -1 \\ \dot{y} & 0 & 1 & -1 \end{pmatrix} \quad (15.38)$$

$$\mathbf{L} = \begin{pmatrix} & \dot{x} & \dot{a} \\ \dot{x} & 1 & 0 \\ \dot{a} & 0 & 1 \\ \dot{b} & 0 & -1 \\ \dot{y} & -1 & -1 \end{pmatrix} \quad (15.39)$$

$$\mathbf{K} = \begin{pmatrix} & v_1 \\ v_1 & 1 \\ v_2 & 1 \\ v_3 & 1 \end{pmatrix} \quad (15.40)$$

Network B (15.41)

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 \\ \dot{x} & 1 & -1 & 0 & 0 \\ \dot{y} & 0 & 1 & -1 & 0 \\ \dot{a} & -1 & 0 & 1 & 1 \\ \dot{b} & 1 & 0 & -1 & -1 \end{pmatrix} \quad (15.42)$$

$$\mathbf{L} = \begin{pmatrix} & \dot{x} & \dot{y} & \dot{a} \\ \dot{x} & 1 & 0 & 0 \\ \dot{y} & 0 & 1 & 0 \\ \dot{a} & 0 & 0 & 1 \\ \dot{b} & 0 & 0 & -1 \end{pmatrix} \quad (15.43)$$

$$\mathbf{K} = \begin{pmatrix} & v_1 \\ v_1 & 1 \\ v_2 & 1 \\ v_3 & 1 \\ v_4 & 0 \end{pmatrix} \quad (15.44)$$

Network C (15.45)

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ \dot{x} & 1 & -1 & -1 & 0 & 0 \\ \dot{y} & 0 & 1 & 0 & -1 & 0 \\ \dot{a} & -1 & 0 & 0 & 1 & 1 \\ \dot{b} & 1 & 0 & 0 & -1 & -1 \\ \dot{z} & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \quad (15.46)$$

$$\mathbf{L} = \begin{pmatrix} & \dot{x} & \dot{y} & \dot{a} \\ \dot{x} & 1 & 0 & 0 \\ \dot{y} & 0 & 1 & 0 \\ \dot{a} & 0 & 0 & 1 \\ \dot{b} & 0 & 0 & -1 \\ \dot{z} & -1 & -1 & -1 \end{pmatrix} \quad (15.47)$$

$$\mathbf{K} = \begin{pmatrix} & v_3 & v_2 \\ v_1 & 1 & 0 \\ v_2 & 0 & 1 \\ v_3 & 1 & 1 \\ v_4 & 0 & 1 \\ v_5 & 1 & 0 \end{pmatrix} \quad (15.48)$$

Network D (15.49)

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ \dot{x} & 1 & -1 & -1 & 0 & 0 \\ \dot{y} & 0 & 1 & 0 & -1 & 0 \\ \dot{a} & -1 & 0 & 0 & 0 & 1 \\ \dot{z} & 0 & 0 & 1 & 0 & -1 \\ \dot{b} & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \quad (15.50)$$

$$\mathbf{L} = \begin{pmatrix} & \dot{x} & \dot{y} & \dot{a} & \dot{z} \\ \dot{x} & 1 & 0 & 0 & 0 \\ \dot{y} & 0 & 1 & 0 & 0 \\ \dot{a} & 0 & 0 & 1 & 0 \\ \dot{z} & 0 & 0 & 0 & 1 \\ \dot{b} & 0 & 0 & -1 & 0 \end{pmatrix} \quad (15.51)$$

$$\mathbf{K} = \begin{pmatrix} & v_1 \\ v_1 & 1 \\ v_2 & 0 \\ v_3 & 1 \\ v_4 & 0 \\ v_5 & 1 \end{pmatrix} \quad (15.52)$$

### Exercises

The metabolic network displayed in Figure 9.9 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.

Answer:

$$\begin{aligned}
 \mathbf{N} &= \left( \begin{array}{ccccccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \\ \text{hex6p} & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dot{fbp} & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dot{gap} & 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ \dot{gabp} & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ \dot{dhap} & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ \dot{nad} & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ \dot{glp} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ \dot{adp} & 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \\ \dot{nad} & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ \dot{amp} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ \dot{atp} & -1 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \end{array} \right) \\
 \mathbf{L} &= \left( \begin{array}{ccccccccc} & \text{hex6p} & \dot{fbp} & \dot{gap} & \dot{gabp} & \dot{dhap} & \dot{nad} & \dot{glp} & \dot{adp} \\ \text{hex6p} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dot{fbp} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dot{gap} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \dot{gabp} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \dot{dhap} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \dot{nad} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \dot{glp} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \dot{adp} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \dot{nad} & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ \dot{amp} & 1 & 2 & 1 & 1 & 1 & 0 & 1 & 1 \\ \dot{atp} & 1 & 2 & 1 & 1 & 1 & 0 & 1 & 1 \end{array} \right) \\
 \mathbf{K} &= \left( \begin{array}{c} v_1 \\ v_1 \\ v_2 \\ v_2 \\ v_3 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_6 \\ v_7 \\ v_7 \\ v_8 \\ v_8 \\ v_9 \\ v_9 \end{array} \right)
 \end{aligned} \tag{15.53}$$

The matrix  $\mathbf{N}_R$  is given by the first 8 rows of  $\mathbf{N}$ . The moieties are:

$$\text{constant} = \text{nad} + \text{nad}\text{h}$$

$$\text{constant} = \text{adp} + \text{amp} + \text{atp}$$

$$\text{constant} = \text{adp} + 2\text{atp} + \text{dhap} + 2\text{fbp} + \text{gabp} + \text{gap} + \text{glp} + \text{hex6p}$$

## 15.8 Answers to Chapter 10 exercises

1. Consider the metabolic network given in Figure 10.1. Maximize  $Z = 3J_1 + 2J_2$  subject to  $0 \leq J_1 \leq 4$ ,  $0 \leq J_2 \leq 6$ , and  $3J_1 + 2J_2 \leq 18$ .
  - Draw the solution space in 2D similar to Figure 10.6
  - Is the solution unique? Why (not)?

Answer:

We start with making a graph of the feasible solution space, which is shown in figure 15.4A. Here we have eliminated  $J_3$  from the problem; since, we are studying steady states where  $J_3$  can be obtained from  $J_1$  and  $J_2$  because of the steady state constraint:  $\dot{x} = J_1 - J_2 - J_3 = 0$ . Since, the constraints and the objective refer only to  $J_1$  and  $J_2$ , it is indeed the simplest to eliminate  $J_3$  and not  $J_1$  or  $J_2$ . The constraint determines the solution space as shown in Figure 15.4A. In Figure 15.4B, the objective is shown for different values of  $Z$  and we would like to know the solution to the system when  $Z$  is maximal. The lines indicate that this occurs when  $Z = 18$ . (This is not a surprise for the students how immediately spotted that the objective and one of the constraints are the same.) The intriguing aspect of this problem is that not a single solution is found but the whole line  $3J_1 + 2J_2 = 18$  for  $J_1 \leq 4$  and  $J_2 \leq 6$ ; denoted in more mathematical language by the set  $\{3J_1 + 2J_2 = 18 | J_1 \leq 4, J_2 \leq 6\}$ . This example indicates that not always a single optimal flux vector is found as a solution of a FBA calculation. If we would have changed the objective such that its slope is not equals to the constraint line  $3J_1 + 2J_2 \leq 18$  then we have found as a solution one of the end points of the line segment defined by  $3J_1 + 2J_2 = 18$  for  $J_1 \leq 4$  and  $J_2 \leq 6$ , i.e. (2, 6) or (4, 3).

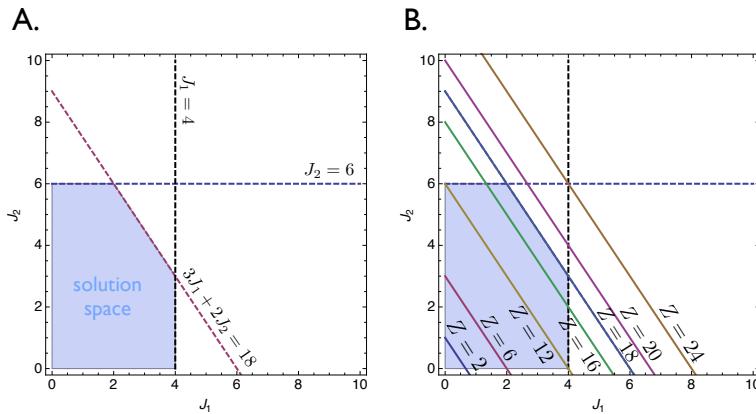


Figure 15.4: Solution space (A) and solution space with different objective values (b)

2. Draw examples of different network topologies where at least five fluxes can be estimated from only two measurements. Explore with what structure external flux measurements do not suffice to solve the set of fluxes.  
 Answer: Every network that you draw that agrees with the following requirement will be a correct answer to this question; any network with 7 fluxes will have 2 independent fluxes (as required by the question) in the rank of the stoichiometry matrix is 5. Thus, any network with 7 fluxes and with a rank of its stoichiometry matrix of 5 leads to a correct network. Think of any network with one branch and 7 reactions and 5 intermediates without moiety conservation as the simplest one.
3. In Figure 10.5 a metabolic network is shown. Assume all reactions are irreversible.

- Construct a kernel matrix  $\mathbf{K}$  by inspection and by linear algebra (section 9.9.2) and draw the flux modes in the network.

Answer:

This network has 7 fluxes and 5 metabolites and conserved moieties do not occur; hence, we need to know two fluxes to be able to determine all 7 fluxes. Thus, we have two independent fluxes and 5 dependent fluxes. If we take the by-product formation flux,  $J_5$ , and the biomass formation flux  $J_7$  as independent then we can determine all fluxes. (Note that not all choices of two fluxes are possible for the independent fluxes but this combination works.) A suitable kernel matrix is,

$$\mathbf{K} = \begin{pmatrix} J_5 & 1 & 0 \\ J_7 & 0 & 1 \\ J_1 & 1 & 2 \\ J_2 & 1 & 2 \\ J_3 & 1 & 1 \\ J_4 & 0 & 1 \\ J_6 & 0 & 1 \end{pmatrix} \quad (15.54)$$

This can be easily verified with Gaussian elimination or by rewriting the steady state mass balances of the network. This will not be done here.

- Draw the solution space in 3D in Mathematica.  
 Answer: This can be done in Mathematica for instance by the following code: `Plot3D[{j5 + 2 j7}, {j5, 1, 20}, {j7, 1, 20}, Mesh → None, Boxed → False]`. Note that we have used the steady-state flux relationship  $J_1 = J_5 + 2J_7$  to visualise the solution space.
- Try to maximise byproduct formation using FBA. Do you get a solution?  
 Answer:

A little thought will lead to the conclusion that  $J_5$ , the by-product formation flux, can become infinite because no upper bound exists for  $J_1$ ! So, no solution would be obtained for  $J_5$ . The absence of a  $J_1$  bound is of course unrealistic for biological applications because the nutrient transporters will always have a  $V_{max}$ .

- Constrain the network such that it becomes bounded.

Answer:

This can be done setting  $J_1 = 1$ .

- Calculate maximal byproduct and biomass production under such constraints

Answer:

Given  $J_1 = 1 = J_5 + 2J_7$ ,  $J_5$  can become maximally 1 and  $J_7$  becomes maximally 0.5.

- Do FVA for each optimization

Answer:

In both cases, no variability is obtained as all the input is diverted to by-product or biomass formation.

- A flux of  $J_1$  of  $10 \pm 2 \text{ mmol h}^{-1} \text{ gDW}^{-1}$  was measured, and a biomass production of  $4 \text{ h}^{-1}$ . Calculate the predicted span of the byproduct formation flux.

Answer:

The mean  $J_1$  equals 10 and then we have  $J_1 = 10 = J_5 + 2J_7 = J_5 + 8$  and therefore  $J_5$  will be 2. The variability in the product formation flux arises from the measurement error in  $J_1$  when  $J_1 = 8$ ,  $J_5$  equals zero and when  $J_1 = 12$ ,  $J_5$  equals 4. So the span of  $J_5$  equals  $\{0, 4\}$ .

4. In Figure 10.12 a metabolic network is shown with parallel pathways. Assume all reactions are irreversible.

- The input flux was measured to be  $10 \text{ mmol h}^{-1} \text{ gDW}^{-1}$ . Give the span of each reaction in the network.

Answer:

$J_2: \{0, 10\}$

$J_3: \{0, 10\}$

$J_4: \{0, 10\}$

$J_5: \{0, 10\}$

$J_6: \{10, 10\}$

Clearly,  $J_6$  can only be 10. And  $J_2 = J_3$  and if they are 10 then  $J_4 = J_5 = 0$  and vice versa.

- 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.

Answer:

$J_2: \{-\infty, \infty\}$

$J_3: \{-\infty, \infty\}$

$$J_4: \{-\infty, \infty\}$$

$$J_5: \{-\infty, \infty\}$$

$$J_6: \{10, 10\}$$

The infinities are unrealistic in this case and equally well for the previous example.

## 15.9 Answers to Chapter 12 exercises

### Exercise

1. Determine the elasticity coefficient of a reversible Michaelis-Menten rate equation for its product.

Answer:

$$\begin{aligned} v &= \frac{V_M^+ \frac{s}{K_s} - V_M^- \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} = \frac{V_M^+ \frac{s}{K_s} \left(1 - \frac{p}{s \cdot K_{eq}}\right)}{1 + \frac{s}{K_s} + \frac{p}{K_p}} = \frac{f(s, p)}{g(s, p)} \\ f(s, p) &= V_M^+ \frac{s}{K_s} - V_M^- \frac{p}{K_p} \\ g(s, p) &= 1 + \frac{s}{K_s} + \frac{p}{K_p} \end{aligned} \quad (15.55)$$

And,

$$\begin{aligned} \epsilon_p^v &= \frac{\partial \ln v}{\partial \ln p} = \frac{\partial \ln \frac{f}{g}}{\partial \ln p} = \frac{\partial \ln f}{\partial \ln p} - \frac{\partial \ln g}{\partial \ln p} = \frac{\partial f/p}{\partial p/f} - \frac{\partial g/p}{\partial p/g} \\ &= \frac{V_M^-}{K_p} \frac{p}{V_M^+ \frac{s}{K_s} + V_M^- \frac{p}{K_p}} - \frac{1}{K_p} \frac{p}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \\ &= \underbrace{\frac{-\frac{p}{s \cdot K_{eq}}}{1 - \frac{p}{s \cdot K_{eq}}}}_{\text{thermodynamic part}} - \underbrace{\frac{\frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}}}_{\text{kinetic part}} \end{aligned} \quad (15.56)$$

2. Determine the elasticity coefficient of an irreversible Michaelis-Menten rate equation for its substrate.

Answer:

$$\begin{aligned} v &= \frac{V_M^+ s}{K_s + s} = \frac{f(s)}{g(s)} \\ f(s, p) &= V_M^+ s \\ g(s, p) &= K_s + s \end{aligned} \quad (15.57)$$

And,

$$\epsilon_s^v = \frac{\partial f}{\partial s} \frac{s}{f} - \frac{\partial g}{\partial s} \frac{s}{g} = 1 - \frac{s}{K_s + s} = \frac{K_s}{K_s + s} \quad (15.58)$$

### 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?

Answer: An elasticity coefficient for a product is negative, which is due to

product inhibition  $\epsilon_p^v = \frac{\partial v}{\partial p} \frac{p}{v}$ ; hence,  $\frac{\partial v}{\partial p} < 0$ . For a substrate and allosteric activator the elasticity coefficient is positive and a competitive inhibitor it will be negative.

2. Make the rate characteristic for the 2-enzyme pathway described in the legend to Figure 12.6, set  $s$  to  $2 \text{ mM}$ . 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 12.6. 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?

Answer: In Figure 15.5 the rate characteristic is shown and the steady state that it predicts agrees with Figure 12.6.

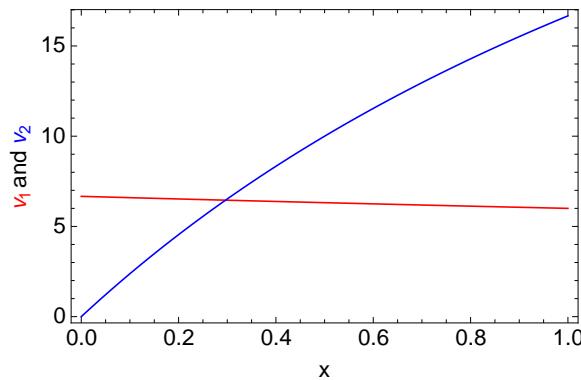


Figure 15.5: Rate characteristics for Figure 12.6 with  $s$  set to  $2 \text{ mM}$ .

The values for the elasticity coefficients are in this steady state:  $\epsilon_s^{v_1} = -0.03$ ,  $\epsilon_s^{v_2} = 0.87$  and  $\epsilon_x^{v_1} = 0.35$ . Enzyme 2 is most sensitive to  $x$ .  $C_1^x = \frac{\epsilon_x^{v_1} - 1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} = \frac{-1}{-0.03 - 0.87} = 1.11$ ; which means that  $x$  increase with 1.11% if enzyme 1 is increases in activity by 1%.

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.

Answer: The summation theorem for concentration control tells us that  $C_1^x + C_2^x = 0$  and that therefore  $C_2^x = -1.11$ .

## Exercises

1. Determine the elasticity coefficient for the reversible Michaelis-Menten mechanism (equation 6.42) with respect to its substrate and product. Write it as a difference between a term that contains the mass-action ratio,  $\Gamma/K_{eq}$  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.

Answer: for the derivation see equation 12.21. Close to thermodynamic equilibrium  $p/s \approx K_{eq}$  and the thermodynamic part then becomes much larger than the kinetic part which ranges between 0 and 1.

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?

Answer: If  $\epsilon_x^{v_1} = 0$  then the connectivity theorem for flux control tells us that  $C_1^J \underbrace{\epsilon_x^{v_1}}_{=0} + C_2^J \epsilon_x^{v_2} = C_2^J \epsilon_x^{v_2} = 0$ ; thus,  $C_2^J = 0$  and due to the summation theorem  $C_1^J + C_2^J = 1$ ,  $C_1^J$  becomes 1.

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

$$\begin{aligned} C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} &= 0 \\ C_1^X \epsilon_x^{v_1} + C_2^X \epsilon_x^{v_2} &= -1 \end{aligned} \quad (15.59)$$

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

Answer: this is obvious and immediately follows the definition.

4. Use the connectivity theorems (equations 15.59) to show that the enzyme, which is sensitive to a metabolite, that only influences this enzyme and no other, 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.

Answer: Let's call the enzyme  $i$  and the metabolite  $x$ . Then,  $C_i^J \epsilon_x^{v_i} = 0$ , and  $C_i^J \epsilon_x^{v_i} = -1$ , due to the connectivity theorems; hence, no flux control by this enzyme only concentration control.

## Exercises

1. Determine the flux control coefficients of a 3-enzyme linear pathway without feedback. Do the same for the pathway displayed in 12.3C. What is the effect of the feedback on the flux control coefficients?

Answer: This is quite a task, you can use Mathematica to solve the equations. You will see that strong feedback, as discussed in the MCA chapter leads to: i. homeostasis of the feedback metabolite and ii. the flux control shifts to the last reaction.

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

Answer: Then the connectivity theorem for  $Y$  becomes:  $C_3^J \epsilon_y^{v_3} = 0$  and therefore  $C_3^J = 0$  and thus all flux control is divided over enzyme 1 and 2.

3. A transient time,  $\tau$ , 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.

Answer:  $\tau = \frac{x}{J}$  thus we have for the control coefficient  $C_i^\tau = \frac{\partial \ln \tau}{\partial \ln v_i} = \frac{\partial \ln \frac{x}{J}}{\partial \ln v_i} = \frac{\partial \ln x}{\partial \ln v_i} - \frac{\partial \ln J}{\partial \ln v_i} = C_i^x - C_i^J$ ; The summation theorem now becomes  $\sum_i C_i^\tau = \sum_i C_i^x - \sum_i C_i^J = 0 - 1 = -1$ .

### 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  $J_3 \ll J_2$ . Under what conditions becomes  $C_2^{J_3} < -1$ ? This phenomenon is called branch-point ultrasensitivity.

Answer: All the answer to this question can be obtained from the main text; however, in the text enzyme 2 was considered as the small branch and not enzyme 3. So, if you view enzyme 2 in the main text as enzyme 3 then you obtain the answers.

# 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] A Cornish-Bowden, J H Hofmeyr, and M L Cárdenas. Stoichiometric analysis in studies of metabolism. *Biochem Soc Trans*, 30(2):43–46, Apr 2002.
- [9] O Decroly and A Goldbeter. Bhythmicity, 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.
- [10] T Drengstig, X Y Ni, K Thorsen, I W Jolma, and P Ruoff. Robust adaptation and homeostasis by autocatalysis. *J Phys Chem B*, 116(18):5355–5363, May 2012.

- [11] 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.
- [12] 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.
- [13] L Goentoro, O Shoval, M W Kirschner, and U Alon. The incoherent feed-forward loop can provide fold-change detection in gene regulation. *Mol Cell*, 36(5):894–899, Dec 2009.
- [14] 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.
- [15] A Goldbeter and R Lefever. Dissipative structures for an allosteric model. application to glycolytic oscillations. *Biophys J*, 12(10):1302–1315, Oct 1972.
- [16] 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.
- [17] I V Grigoriev, Y A Makhnovskii, A M Berezhkovskii, and V Y Zitserman. Kinetics of escape through a hole. *Journal of Chemical Physics*, 116:9574, 2002.
- [18] 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.
- [19] 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.
- [20] J S Hofmeyr and A Cornish-Bowden. Regulating the cellular economy of supply and demand. *FEBS Lett*, 476(1-2):47–51, Jun 2000.
- [21] J J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A*, 71(10):4135–4139, Oct 1974.
- [22] 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.
- [23] 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.

- [24] 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.
- [25] 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.
- [26] 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.
- [27] D E Koshland, A Goldbeter, and J B Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, 217(4556):220–225, Jul 1982.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] S V Popova and E E Sel'kov. Generalization of the model by monod, wyman and changeux for the case of a reversible monosubstrate reaction sr, tp. *FEBS Lett*, 53(3):269–273, May 1975.
- [32] 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.
- [33] 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.
- [34] I H Segel. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. Wiley-Interscience, 1993.
- [35] 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.
- [36] G Shinar, R Milo, M R Martínez, and U Alon. Input output robustness in simple bacterial signaling systems. *Proc Natl Acad Sci U S A*, 104(50):19931–19935, Dec 2007.

- [37] A Singer, Z Schuss, D Holcman, and R S Eisenberg. Narrow escape. part i. *Journal of statistical physics*, 122:437, 2006.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] T M Yi, Y Huang, M I Simon, and J Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, 97(9):4649–4653, Apr 2000.