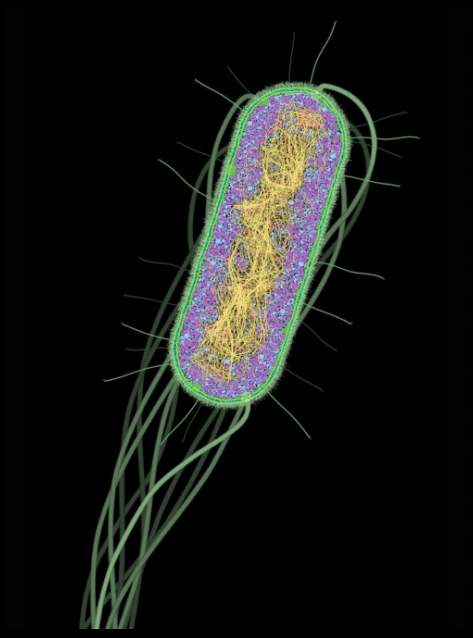
MOLECULAR SYSTEMS BIOLOGY IN A NUTSHELL

Studying cell biology quantitatively using models and experiments



Frank J. Bruggeman

Systems Bioinformatics, VU University, Amsterdam, The Netherlands

Contents

Preface

Since the 2000's, biology has rapidly evolved into an interdisciplinary science. This development was largely due to the introduction of key measurement technologies: genome sequencing, measurement of mRNA, protein and metabolite concentrations, and fluorescence microscopy. One of the revolutions was that these measurements gave genome-wide information, so of *all* the metabolites, mRNA and proteins in a cell. This in itself led to new questions in cell biology, regarding the entire cell and how its functioning emerges from the concerted behaviour of all its molecules. Together these molecules from intricate reaction networks, described by biochemistry.

These new technologies led to two new subdisciplines in biology that have an important role to play – bioinformatics and systems biology – and caused the maturation of another – biophysics. These three disciplines have had an enormous impact on biological research and thought. That is so, because they enable a molecular understanding of life that is not only of fundamental interest bu also required for biotechnological and medical applications.

This small book illustrates, at a basic level, how systems biologists use well-defined quantitative concepts, basic biochemistry, and simple mathematical models to study the molecular networks that underlie the phenomena studied in cell biology. This book is introductory and deliberately kept simple.

This text is part of the course *Introduction to Systems Biology* and is meant as an illustration of some of the approaches taken in systems biology. It is not a complete overview: it is limited to basic analysis of the kinetic properties of molecules, their reactions and how the properties of molecular networks can be understood in these terms. Analysis of data generated by large molecular networks, of which the structure and molecular players are poorly understood, is not considered. So, analysis of microarray, proteomics, sequencing or metabolomics data is beyond this text and logically follows after the focus of this text on the basics of molecular networks driving cell behaviour.

The *Introduction to Systems Biology* course also involves a mathematics refresher classes. Mathematics is used throughout this text to illustrate how this can be useful in addressing molecular and cell biology questions. In case you experience any problems with the mathematics in the text then discuss your

problem in the mathematics refresher class and ask for additional exercises.

This text is meant as an introduction to the field. So, you are not expected to be able to carry out all derivations yourself; just try to follow them step by step. At this stage, it is more important that you experience the approaches, concepts, and the way of thinking. Systems biology takes time to appreciate and getting used to.

I encourage you to read the text prior to the classes. Read the text carefully; it is okay if you do not understand everything immediately. In this way you are much better prepared for the lectures. Also, do not miss any of the classes! In the exercise classes, we make exercises together; make sure you participate in all those classes, because this is when you learn most, by doing it yourself!

Frank Bruggeman, July 2021

Chapter 1

What is systems biology?

1.1 Characterising systems biology

It is not so straightforward to define a scientific discipline, as scientific disciplines evolve, merge, and exchange methods. Especially in biology many disciplines exist such as genetics, molecular biology, biophysics, biochemistry, microbiology, cell biology, ecology, bioinformatics, and systems biology. Together they achieve understanding of biological systems, at all its level of organisation from molecules to ecosystems. Biological disciplines nonetheless differ, because they use different methods and ask different questions. They are also often reliant on one and the other, breakthroughs in one discipline influence developments in others.

Systems biology arose in the early 2000s, because many scientists started realising at the same time that understanding of the enormous molecular networks that living cells consist of was required. They knew about proteins, genes and cell behaviours, but how networks of interacting molecules gave rise to these cell behaviour was, and largely still is, an open problem. Examples of molecular networks are metabolic, signalling and transcription-factor-gene networks. A dedicated discipline was needed to study them.

In the early 2000s, and also earlier, several methods were introduced to cell biology that asked for a network perspective on biological data such as transcriptomics, proteomics, metabolic-flux analysis and genome sequencing. Nowadays we have even more of those high throughput experimental methods, which give rise to huge amounts of data that needs to be interpreted and even integrated with one another. Many systems biologists and bioinformaticians are thinking about ways to integrate such data sets. Clearly this requires appreciation of the biochemical mechanisms underlying the activity molecular networks, because this is how all those different molecules are interlinked in a cell. And, if we intend to change the behaviour of a cell, e.g. in medicine with a drug or with genetic engineering in biotechnology, mechanistic understanding is needed, merely looking at correlations without establishing with bioinformatics methods

is not enough.

Understanding networks of interacting molecules, from a biochemical mechanistic perspective, requires other approaches than those that are required for understanding of properties of isolated (macro)molecules. It turns out that approaches from other disciplines are highly useful for understanding of molecular networks. The main 'cross-feeding' disciplines are biochemistry, dynamical systems theory, computer science, control theory and physical chemistry.

Systems biology is therefore an interdisciplinary endeavour that is primarily, not exclusively, focussed on understanding cell behaviour in terms of the molecular networks of cells and borrows approaches from other disciplines – i.e. those cites above – to design and make sense of biological experiments. Systems biology therefore attracts scientists from different areas of research.

1.2 Approaches used in systems biology

The following approaches characterise systems biology research:

1. **Dynamics of molecular networks.** A characteristic feature of living systems is that they are dynamic. They sense their environment. Adapt to changing conditions, rewire their networks, etc.. A key aspect is therefore to understand how properties of molecules, which define with their interactions, give rise to network dynamics associated with cell functioning. Many cell-biological processes are dynamic too: such as signal perception, responses, cell-cycle, structure formation, etc.. We need to understand those dynamics to better understand the consequence of mutations on disease development and predict how genetic engineering can improve cells for biotechnological purposes.

2. **Quantitative measurement.** Qualitative measurements such as up- or down, responding or not-responding are not very informative when one aims to characterise the behaviour of cells in terms of its molecular networks or for comparing different of systems. Many systems biology efforts are therefore focussing on the development of quantitative measurement methods and the analysis of the resulting data.

3. **Understanding the design of molecular networks ('functions').** Since molecules interact very selectively with one another – biological networks are very sparse – it is of great interest to understand the function of particular network topological features such as feedbacks, cycles, cascades and pathways.

4. **Mathematical modelling.** Simulation of network behaviours using mathematical models are indispensable to understand quantitative measurements of molecular-network dynamics and the role of network structure, Since network dynamics, the outcome of molecular interactions in networks, is generally too complicated to work out on paper and on the basis of static diagrams, mathematical models are extremely useful. They allow

for the simulation of network dynamics in response to perturbations. Not in all cases do those models have to be predictive or quantitatively precise. Often 'toy' models are used too, to sharpen the intuition. They help grasp the essential parameters and features of a particular molecular network. It is also often the case that this 'essentialisation' is what is most useful: negative feedback, for instance, often serves the same purpose in molecular networks, regardless of the exact network function or organism. Cells often exploit the same network principles. Identifying those common principles becomes problematic when all molecular details are considered. In other cases, for instance, when a drug target needs to be predicted, predictive models are useful that are parameterised on experimental data. Thus, the type of models that is required depends on the scientific question.

5. **Theory development.** In addition to concrete models, theories turn out to be useful guides for generating testable predictions and explaining experimental findings. Since cells of different biological species are confronted with similar problems during their struggles for existence, they are evolutionarily related, and carry out their functions in a multicellular context, they turn out to exploit similar network principles. Therefore, they obey similar regularities that can sometimes be phrased in terms of testable theories. For instance, theories exist about enzyme kinetics, metabolic pathway control, robustness of molecular networks, stochasticity of molecular processes, sensitivity of sensing and signalling networks, and maximisation of growth rate.

6. **Evolutionary aspects.** Unicellular organisms that live in large populations – with population sizes of millions of organisms – and grow fast, can evolve rapidly new properties. Fitter genotypes may then take over the population and fix. This mechanism shapes the properties of molecular networks such as molecular properties and network wiring according to an evolutionary objective that natural selection maximises. Evolutionary approaches are therefore a powerful method to understand network design and behaviour, and a new way of thinking about cells to our classical, more-molecular viewpoint from biochemistry, genetics, biophysics and molecular biology.

7. **Molecular-system constraints, limits and trade offs.** Molecular systems are physicochemical systems constraint by chemical and physical 'laws'. Likewise, network behaviours are constrained by the 'laws' of dynamical systems. This means that networks are limited in their behaviours. They can, for instance, display trade offs – they can be good at one function, at the expense of another. For instance, a microbe that aims to grow fast should spend all most of its protein into metabolism and translation and, for instance, not on stress proteins as this does not contribute to making new cells. Therefore, a fast growing cell is more stress sensitive than a slow growing cell. Cells have various solutions to this problem, one of them is to diversify the population into two subpopulations:

into one that is stress-sensitive, fast-growing and the other that is stress-resistant, slow-growing. Theory can then be developed to predict the optimal subpopulation fractions given the probabilities of growth-supporting and stress conditions. Here models, theories, quantitative experiments and biology all come together.

1.3 Systems biologists are not pure theoreticians, but "Jacks of all trades"

What is often thought is that systems biology are theoreticians. That is a misconception, they are often working in close collaboration with experimentalists, embedded in experimental labs, and often carry out quantitative experiments. They are driven by the biological question, and not so much by the mathematical and theoretical methods for which physicists are generally better suited.

A system biologist is more a "jack of all trades" and capable of integrative, interdisciplinary research involving different biological disciplines (such biochemists, oncologists, biotechnologists, medical doctors) and methods (mathematical modelling, computational data analysis, basic bioinformatics, genome sequencing analysis, cell biology in the lab, etc.). See figure 1.1 for an example of a "Jack of all trades".

In simple terms, the difference between systems biologists, computational biologists and bioinformaticians is that systems biologists want to understand to understand the biology, for instance how cells work, while computational biologists and bioinformaticians are more into the data-analysis methods.

1.4 Topics in systems biology

Despite its short existence, systems biology has permeated entire cell biology, including microbiology. Systems biologists are, for instance, working on metabolism, gene regulation (transcription) and signalling; how those are interrelated and together lead to cell growth and adaptation.

Systems biology turns out to be the long-sought-after 'glue' that can link different biological disciplines to one another and non-biological disciplines. It is therefore a highly vibrant and innovative field where there exists always a lot of excitement about new findings and approaches. It is also not the easiest field to get into – **so pay attention to the lectures and ask questions whenever you can**.

1.5 Exercises

1. Read the following two papers:

Figure 1.1: **A "Jack of all trades".** A "Jack of all trades" is someone who can solve many different problems since he knows many basic principles and tricks.

(a) Yuri Lazebnik. Can a biologist fix a radio?—or, what i learned while studying apoptosis. *Cancer cell*, 2(3):179–182, 2002

(b) Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002

Make a list of all the scientific terms in these papers (and in this syllabus) chapter, that you found confusing, or do no the meaning of, discuss them with fellow students or look up their definitions online. Try to formulate into your own words what the aim of systems biology are, why its approaches are different from classical approaches, and why it is needed. What problems can systems biology now solve that could not be solved before its inception? (This exercise does not have an answer in the back of this book. Discuss with colleagues and your exercise group teacher about this.)

2. Try do define in your own words what is meant with "molecular network", "dynamics", "dynamic model", "network structuree", "parameter", "variable", "natural selection", "fitness", and other terms you found confusing and unclear in the previous text. (This exercise does not have an answer in the back of this book. Discuss with colleagues and your exercise group teacher about this.)

Chapter 2

Cellular composition and time scales

2.1 Introduction

In this chapter we will take a "look" at bacterial and eukaryotic cells. We shall discuss in a nutshell the molecular composition of cells, the rate of movement of molecules inside them, and the rates of important molecular processes, such as replication, transcription, translation and the cell cycle. This will give you a useful perspective on cells: as molecular self-replicating and self-organising systems, packed with proteins, that display continuous dynamics. These insights provide a useful mental picture of cells when you think about and try to understand with modern-day experimental methods and theory such as fluorescence microscopy, proteomics or metabolomics.

A dynamic view on cells is hardly taken in standard biology textbooks. It is a quantitative view and you will see that basic calculations with quantitative data provide you with insightful information about cell biology. Biology textbooks often tend to describe cell biology in a more qualitative manner and emphasise the types of processes occurring in cells, rather than their quantitative description. Many useful numbers for cell biology can be found at `http://bionumbers.hms.harvard.edu` – in fact I took many of the numbers used in this chapter from this website.

Systems biology adds a quantitative perspective to cell biology, which allows you to do basic calculations to figure out aspects of the nature of cells (and eventually even make predictive models). You will see that this leads to a powerful approach to study how cell behaviour arises out of molecular interactions.

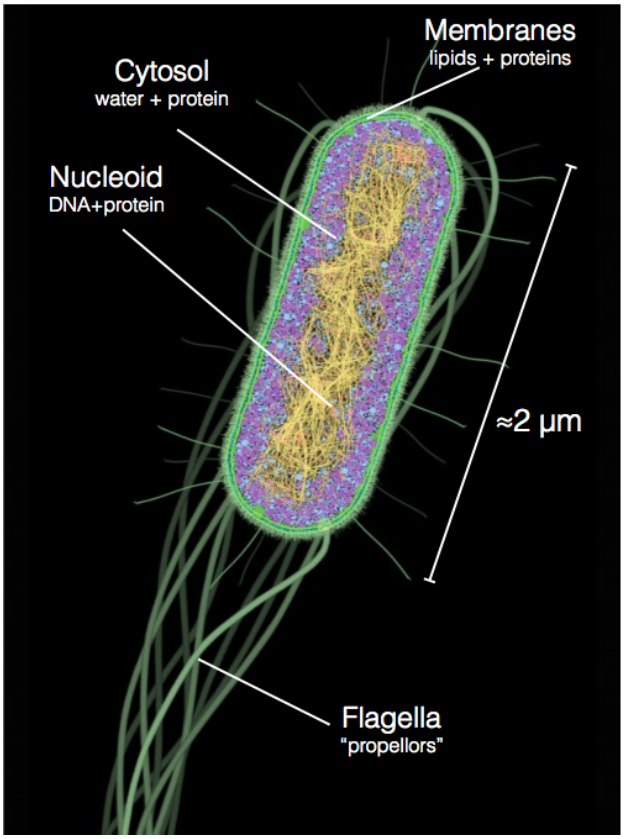


Figure 2.1: **Artist impression of *Escherichia coli* with realistic relative dimensions.** David Goodsell has drawn several well-known molecular systems in cell biology to help biologists envision what such systems are composed of (visit `http://mgl.scripps.edu/people/goodsell/`). If you are unfamiliar with cell biology it is instructive to read an introductory textbook chapter about cell organisation.

2.2 Cell size and composition

Cells are very small: bacteria range from $0.2 - 3$ μm in radius, eukaryotic microorganisms, such as Baker's yeast, are about 5 times bigger and mammalian cells are again 10 times bigger. Since, the volume (V) is proportional[1] to (de-

[1]The variable y is proportional to x is $y = a \times x$.

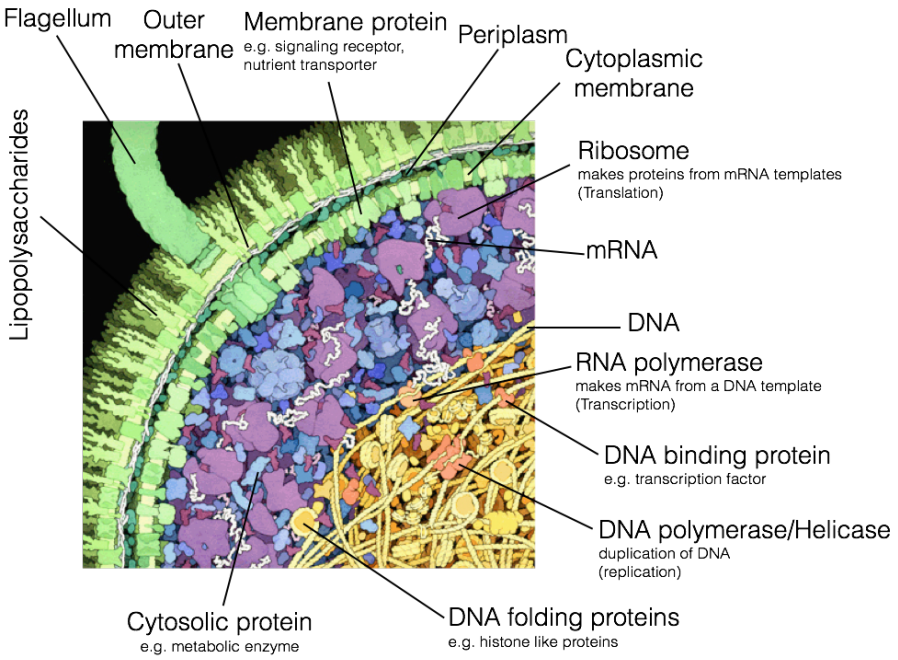


Figure 2.2: **A more detailed view of *Escherichia coli* by Goodsell.** Also in this picture all relative dimensions are realistic. What becomes clear from this picture is the high protein concentration in cells; the average distance between proteins is roughly their diameter (10 nm). So, cells are highly packed and proteins diffuse not in a watery cytosol but in a concentrated protein solution.

noted by $\propto$) the third power of the radius (r), i.e. $V \propto r^3$, the volumes of prokaryotes and eukaryotes vary orders of magnitude[2].

A representative protein radius is 5 $nm = 5 \times 10^{-9} \frac{\text{m}}{\text{protein}}$. So, if a bacterium is 2 $\mu m = 2 \times 10^{-6} m$ long then $\frac{2 \times 10^{-6}\ \text{m}}{5 \times 10^{-9}\ \frac{\text{m}}{\text{protein}}} = 400$ proteins can be aligned to cover the distance from its left to its right cell pole (Figure 2.1).

The cytosol of a bacterium is packed with protein and it is estimated that on average, proteins are 1-protein-diameter separated in distance; the protein-protein distance is therefore about 10 nm (Figure 2.2). A consequence of this 'macromolecular crowding' [4] is that, inside cells, diffusion and reactions can be quite different from how they operate in dilute, watery environments [28]. So, proteins move much slower inside cells than in pure water and protein complexes form more easily inside the cell.

Since a bacterial genome can easily be hundreds of time longer than the cell length (table 2.1), the genome needs to be highly packed and will take up a

[2]One order of magnitude means a factor of 10.

Table 2.1: **Dimensions** (numbers for *Escherichia coli*, indicative for other bacteria)

(Body) Length	2 μm
Volume	1 fl
Genome size	$4.6 \cdot 10^6$ $bps/genome$
Genes	4288 $genes/genome$
Operons	2584 $operons/genome$
Intergenic distance	118 bps
Protein radius	5 nm
Chemotactic speed	$10 - 20$ $body\ lengths/s$
base pair (bp) length	3.4 $angstrom/bp$
genome length	1.6 $mm = 800$ $body\ lengths$

considerable fraction of cell volume (Figure 2.1).

Proteins are also found in the cellular membrane – where they function as signal sensors and transporters – and some are even excreted by the cell to perform extracellular functions. A single cell contains millions of proteins.

2.3 Exercises

1. In this chapter, we will be doing calculations with units. To practise, answer the following exercises. Realise that this way of thinking is correct: $2\ m^3 = 2 \times (10\ dm)^3 = 2000\ dm^3 = 2000\ l$ (l is liter).
   (a) $1\ m = ...\ cm$, $1\ m^2 = ...\ cm^2$, and $1\ m^3 = ...\ cm^3$?
   (b) $1\ dm^3 = ...\ l$, $1\ cm^3 = ...\ l$, and $1\ nm^3 = ...\ l$?
   (c) $1\ fl = ...\mu l$?, $1\ ml = ...\ l$, and $1\ \mu l = ...\ fl$?
   (d) $1\ mol = ...\ molecules$?, $1\ \mu mol = ...\ molecules$?, and $1\ pmol = ...\ molecules$?
2. Assume that the average protein contains 300 amino acids.
   (a) How many amino acids should a cell make during its cell cycle when it contains 2×10^6 proteins at birth?
   (b) Calculate how many nitrogen atoms occur on average in an amino acid (use the first table on this Wikipedia page `https://en.wikipedia.org/wiki/Amino_acid`).
   (c) Given the previous two answers, how many ammonium molecules (NH_4^+) should a cell consume during its cell cycle?

(d) How much does that amount of ammonium weigh in femtograms?

3. The length of the DNA of *E. coli*, a bacterium, is 1.5 mm long, consists of 4558953 bps, and its DNA polymerase runs at a speed of 800 bps/sec. How much time does it take for this enzyme to have replicated *E. coli*'s DNA by 50%?

Exercises tips
When answering those exercises you may have realised that working organised and neatly on a piece of paper works best. In my experience many of you make a mess of the calculations – they are not organised. Do not do that. Given the question, write down what information was supplied in the question and what needs to be calculated. Then formulate a strategy on how you are going to solve the problem. Only when you find you organise your thoughts, by keeping track of them on a piece of paper, you work logically and trace back errors that you may have made. Realise what you have learned! Take your time! Always try a question first before you ask a fellow student and then ask a teacher. First process the problem yourself.

2.4 Concentrations of molecules in cells

The volume of a bacterial cell is approximately 1 $\mu m^3 = (10^{-6}m)^3 = (10^{-6} \times 10 \times dm)^3 = 10^{-15}dm^3 = 1fl$ (with "f" meaning femto, 10^{-15}). The volume of a protein is approximately $(5\ nm)^3 = (125 \times 10^{-9} \times 10\ dm)^3 = 1.25 \times 10^{-22}l$. So, if the cytoplasm is completely filled with protein – which it is not, only for about 40% – then the number of protein molecules per cell would be equal to $\frac{10^{-15}\frac{l}{cell}}{1.25\times10^{-22}\frac{l}{protein}} = 8\times10^6\ \frac{proteins}{cell}$. This we have to multiply with the percentage of protein volume, say 40%, to obtain $3 \times 10^6\ \frac{proteins}{cell}$. This is a good estimate when compared to actual experimental data.

Different molecules occur at different concentrations in an *E. coli* cell. A molecule that occurs as a single copy per cell has the following concentration: $\frac{1\ \frac{molecule}{cell}}{10^{-15}\frac{l}{cell}}\frac{1}{6\times10^{23}}\frac{mol}{molecules} = \frac{1}{6\times10^8} = \frac{1}{6\times10^8}\frac{mol}{l} \approx 1\ nM$. Thus, 10^3 molecules per cell equals 1 μM, and 10^6 equals 1 mM. The concentration ranges for classes of molecules in *E. coli* are shown in figure 2.3.

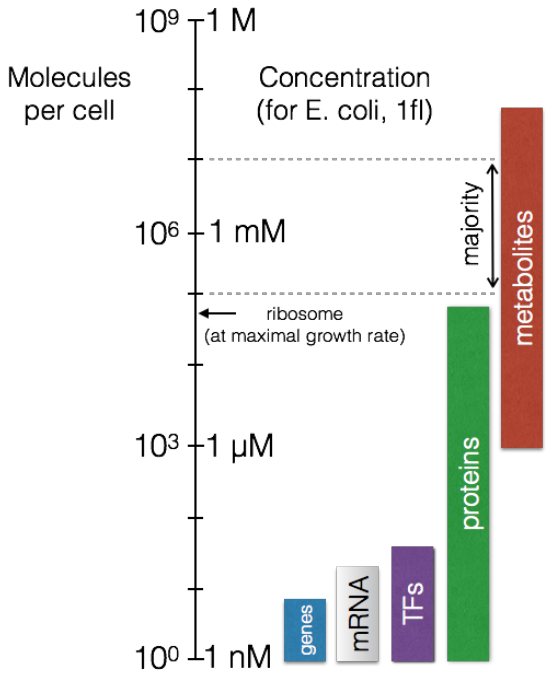


Figure 2.3: **Concentrations and number of molecules per cell (*E. coli* numbers).** Overview of the range of molecule copy number per cell and their concentrations given the volume of *E. coli* of 1 fl. Metabolites are low-molecular weight molecules that are the reactants of the enzymes in metabolism, such as glucose, pyruvate, amino acids, et; thus, the intermediates of metabolic pathways.

2.5 Major components of cells and their synthesis by metabolism

The basic principle of metabolism and cell growth is that the cell doubles its volume, mass and its entire molecular content during one doubling time (sometimes called the generation time; a cell cycle duration) and then splits into two identical daughter cells. The doubling time is defined as the time that it takes for a daughter cell, which has just emerged after a cell division event, to grow into a mother cell that is about to divide (figure 6.1).

After the doubling time τ_d we therefore have twice the number of cells (as one mother gives rise to two daughters):

$$N(t+\tau_d) = 2^1 N(t) = 2^{\frac{\tau_D}{\tau_D}} N(t) \tag{2.1}$$

with $N(t)$ as the number of cells at time t; thus, $N(t+\tau_d)$ equals the number of cells τ_d time later.

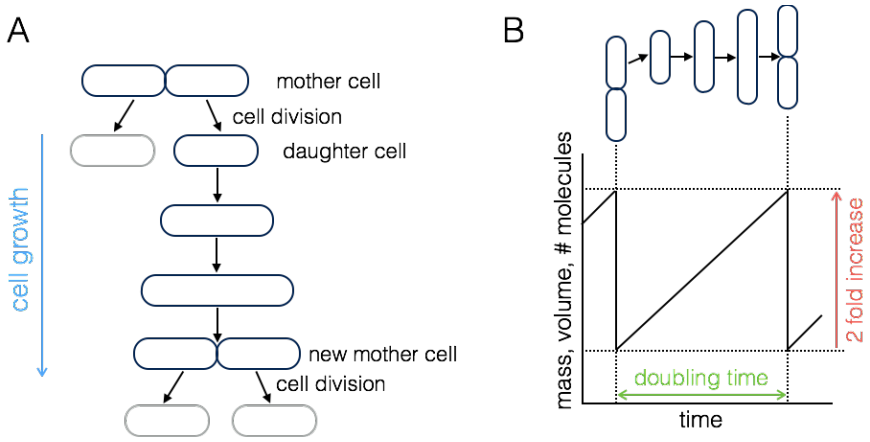


Figure 2.4: **Overview of cell growth of symmetrically dividing bacterium, such as *E. coli*.** After the division of a mother cell two daughter cells emerge. Each matures, grows, and eventually doubles its volume, mass, and molecular content at its doubling time at which they become mother cells and divide themselves.

Accordingly, after t time we have had $\frac{t}{\tau_d}$ doublings; so,

$$N(t) = 2^{\frac{t}{\tau_D}} N(0) \tag{2.2}$$

In biology, the "per-capita" or "specific" growth rate of a cell culture is defined as $\mu = \frac{\ln 2}{\tau_d}$ (such that $\tau_d = \frac{\ln 2}{\mu}$) and, therefore,

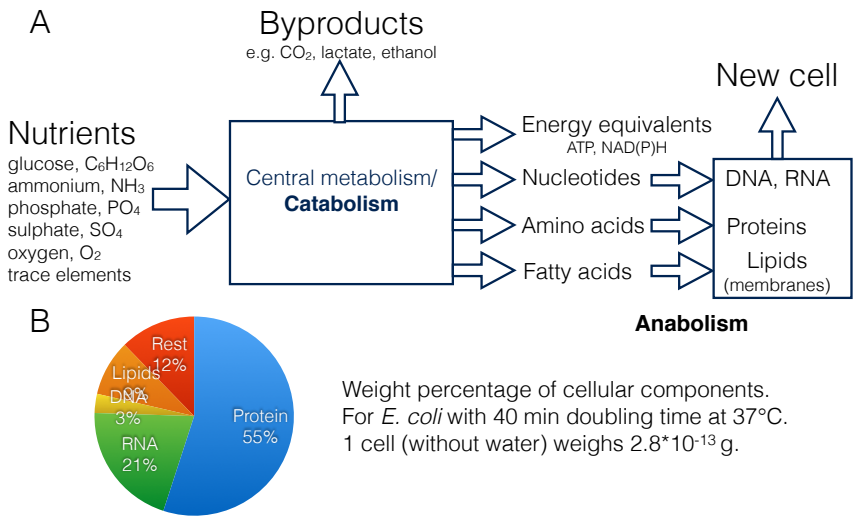
$$N(t) = 2^{\frac{t}{\tau_D}} N(0) = 2^{\frac{\mu}{\ln 2} t} N(0) = e^{\mu t} N(0) \tag{2.3}$$

(The last relation is easy to derive; the statement is $2^{\frac{1}{\ln 2}} = e$. This is correct because $\ln 2^{\frac{1}{\ln 2}} = \frac{1}{\ln 2} \ln 2 = 1 = \ln e$; so, $2^{\frac{1}{\ln 2}} = e$. In the mathematics class, these tricks will be explained and trained.)

In other words, equation 2.3 corresponds to 'exponential growth': the number of cells increases exponentially with time. Now we can derive an equation, which is perhaps more familiar,

$$\begin{aligned} \frac{d}{dt} N(t) &= \frac{d}{dt} e^{\mu t} N(0) \quad \Rightarrow \\ \frac{d}{dt} N(t) &= e^{\mu t} N(0) \mu = \mu N(t) \quad \Rightarrow \\ \frac{d}{dt} N(t) &= \mu N(t) \quad \text{(Exponential growth differential equation)} \end{aligned} \tag{2.4}$$

In other words, doubling of cell volume and mass corresponds to exponential growth.



C $0.24\ C_6H_{12}O_6 + 0.42\ O_2 + 0.26\ NH_4^+ + 0.023\ PO_4^{3-} + 0.006\ SO_4^{2-} \rightarrow CH_{1.59}O_{0.374}N_{0.263}P_{0.023}S_{0.006} + 0.47\ HCO_3^- + 0.64\ H_3O^+$

Figure 2.5: **The essence of cellular metabolism.** Cells are composed of carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and some others minor components (trace elements), such as metals (Fe, Mg, Co). **A.** So, the construction of a new cell by a mother cells requires uptake of nutrients that supply those components. In addition, the mother cells needs to maintain herself, which also requires such nutrients. From nutrients, the building blocks and energy are made in catabolism. Those are then used by anabolism to synthesise the cell's macromolecules: RNA, DNA, proteins, and membranes. Eventually, when the mother cell has doubles its volume and entire molecular content, it divides in two cells to give rise to two daughter cells that each mature into new mother cells after the so-called doubling or generation time. **B.** The percentage contribution of cellular macromolecules to the weight of a single cell. **C.** A very coarse grained view on cell growth. Left of the arrow ($\rightarrow$) the nutrients are shown for cell synthesis and right of the arrow the products, including the cell itself ($CH_{1.59}O_{0.374}N_{0.263}P_{0.023}S_{0.006}$).

2.6 Metabolism supplies the material for cell growth

Doubling of the cell mass and volume means that the entire cellular content doubles in amount. A cell therefore has to synthesise its own components to double them in amount; only then it can divide into two identical daughter cells (on average). While taking up nutrients and making new molecules out of them the volume of the cell grows automatically, because molecules take up volume themselves. Clearly, the largest molecules contribute most to the cell volume and those are the proteins.

Figure 2.5 indicates the basic principles of cellular component synthesis

Table 2.2: **Molecule counts** (numbers for *Escherichia coli*, indicative for other bacteria)

Elemental composition [2]	$CH_{1.59}O_{0.374}N_{0.263}P_{0.023}S_{0.006}$
Weight of a cell	$9.5 \cdot 10^{-13}\ g/cell$
Dry weight of a cell	$2.8 \cdot 10^{-13}\ g/cell$
Water content of a cell (70%)	$6.7 \cdot 10^{-13}\ g/cell$
Protein content of a cell	$156 \cdot 10^{-15}\ g/cell$
RNA	$58 \cdot 10^{-15}\ g/cell(82\%rRNA)$
DNA	$8.8 \cdot 10^{-15}\ g/cell$
Lipid	$25.9 \cdot 10^{-15}\ g/cell$
Lipopolysaccharide	$9.7 \cdot 10^{-15}\ g/cell$
Number of proteins/cell	$2.35 \cdot 10^{6}\ molecules/cell$
1 ml $OD_{600} = 1$ contains	10^9 cells

via metabolism. Metabolism can be roughly divided into two processes: i) catabolism, which converts nutrients into building blocks and energy, and ii) anabolism, which converts the products of catabolism into cellular macromolecules: DNA, RNA, lipids (membranes) and protein. DNA and RNA are composed out of nucleotides, lipids make up membranes, and proteins are composed of amino acids.

Since, cells are composed out of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulphur (S), phosphorus (P), and trace elements (metals, vitamins, etc) they need to take those up and metabolise them into cellular components.

A single *E. coli* cell has a very, very small weight of about $9.5 \times 10^{-13}\ g/cell$ (table 2.2). Protein contributes for 55%, RNA for 20.5%, DNA for 3.1%, lipids for 9.1% and the rest for 12.3 % to the weight of an *E. coli* cell (figure 2.5).

More than 95% of energy is used for protein synthesis; so protein synthesis is central to understanding growth and metabolism of bacteria.

2.7 Exercises

1. We will consider exponential growth of cells and how this can become limited by food supply.

   (a) The meaning of $\frac{dN(t)}{dt}$ is that it equals the "rate of change" of $N(t)$ as function of time t. When you think about this for a moment you will probably realise that this equals the slope of a plot of $N(t)$ as function of t. Since it is the slope you can think of $\frac{dN(t)}{dt}$ as being equal to $\frac{N(t+dt)-N(t)}{t+dt-t}$ – which is after all the equation for the slope of

the $N(t)$-vs-t plot. Given that $\frac{dN(t)}{dt} = \mu N(t)$ (with μ as the growth rate and positive), what happens to the slope in the $N(t)$-vs-t plot when t increases? (Does $N(t)$ increase with time? Why?)

(b) When a population of cells grows according to $\frac{dN(t)}{dt} = \mu N(t)$ new cells are all the time being made by the existing cells. This requires nutrients. What happens to the nutrient concentration, denoted by $s(t)$, over time? How would you mathematically describe the slope of $s(t)$-vs-t? Would it have to depend on $s(t)$, $N(t)$ or both?

(c) What is the problem with the following description $\frac{ds(t)}{dt} = -kN(t)$ with k as a nutrient consumption rate per cell? Why is it impossible?

(d) Eventually, when the nutrients run out, the growth of the population of cells stops. This can be described by the following differential equation $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t)$ with K as the so-called carrying capacity (or yield) of the environment.

i. What changes in the equation when we double the amount of nutrients?

ii. How would you characterise the growth when $N(t)$ is small relative to K?

iii. What is growth rate of the culture when $N(t) = K$?

iv. Sketch the dependency of $N(t)$ on t. (If you want actual numbers set $N(t) = 1$, $K = 1000$ and $\mu = 1\ hr^{-1}$).

v. What has happened when $N(t) = K$?

(e) Calculate the generation time of a cell in minutes that grows at a rate of 1 hr^{-1}.

2. "Numerically solving a differential equation such as $\frac{dN(t)}{dt} = \mu N(t)$" means that the values of $N(t)$ are calculated at different times given a starting value of N at time zero, denoted by $N(0)$. From the definition of $\frac{dN(t)}{dt}$ it becomes clear how this can be done.

(a) Write down the definition of $\frac{dN(t)}{dt}$ as the slope of a $N(t)$-vs-t curve.

(b) Express the value of N at time $t + dt$ in terms of N at t, $\frac{dN}{dt}$ at t and dt. (We assume that dt has a constant value.)

(c) Calculate $N(dt)$ when $\mu = 1\ hr^{-1}$, $dt = 0.1$ and $N(0) = 1$.

(d) Calculate $N(2dt)$ given $N(dt)$.

(e) Now use Excel to calculate $N(t)$ as function of time t from 0 to 10. If you are not familiar with Excel do this with a fellow student or consult `http://bmi.bmt.tue.nl/sysbio/Education/Excel_Euler_simulation.pdf`.

(f) Check that the resulting curve obeys $N(t) = N(0)e^{\mu t}$.

(g) Solve also $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t)$ numerically, using Excel, when $\mu = 1\ hr^{-1}$, $K = 10$ and $N(0) = 1$. Why is this curve and the previous one you have made the same for small times?

Numerically solving a differential equation
In the last exercise you have used Excel to solve a differential equation. This is a rather clumsy way, but one that makes it quite clear how differential equations work and what they mean. Numerically solving of differential equation is typically done with more suitable software such as "R", "Python", "Matlab", "Mathematica" or other packages and then it is all more automated. Those options are not required for this course, but you can use them if you want to.

2.8 Speed of molecule movement inside cells

Molecules move spontaneously through the cell by a process called diffusion – only at the temperature 0 Kelvin do molecules not move. The diffusion coefficient captures the mobility of a molecule in the cell and be used to calculate transport times.

1. **The interpretation of the diffusion coefficient** D. Albert Einstein discovered that the diffusion coefficient of a spherical molecule with radius a equals

$$D = \frac{kT}{6\pi\eta a} \tag{2.5}$$

at temperate T in K, and medium viscosity η in $\frac{kg}{m\times s}$ with $k = 1.38 \times 10^{-23}\ J/K$ (Boltzmann's constant; `https://en.wikipedia.org/wiki/Boltzmann_constant`; $1\ J = 1\ \frac{kg\times m^2}{s^2}$), $\pi = 3.14...$ and the radius a in m. Let's figure out the unit of the diffusion coefficient,

$$D = \frac{kT}{6\pi\eta a} = \frac{J/K \times K}{\frac{kg}{m\times s}\times m} = \frac{J}{\frac{kg}{s}} = \frac{J\times s}{kg} = \frac{\frac{kg\times m^2}{s^2}\times s}{kg} = \frac{m^2}{s} \tag{2.6}$$

Tip
Some of you may have found this derivation hard to follow. This means that you have to do it yourself on paper and check whether you agree. Some of you will then have a problem with realising that $\frac{a}{\frac{b}{c}\times d} = \frac{a\times c}{b\times d}$. Since numbers are symbols too, as letters are, you might as well substitute numbers (the rules are for symbols, not numbers). So is $\frac{2}{\frac{3}{4}\times 6}$ equal to $\frac{2\times 4}{3\times 6}$?

It equals $\frac{2}{\frac{3}{4}\times 6} = \frac{2}{\frac{18}{4}} = \frac{2}{4.5} = \frac{4}{9}$. According to $\frac{a}{\frac{b}{c}\times d} = \frac{a\times c}{b\times d}$ we could also have written $\frac{2}{\frac{3}{4}\times 6} = \frac{2\times 4}{3\times 6} = \frac{8}{19} = \frac{4}{9}$. Thus the rule is correct! Many of you have forgotten about these "high school tricks".

Consider the equation of the diffusion coefficient again

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

it agrees with intuition: 1. A molecule in a more viscous fluid, e.g. maple syrup vs water, moves slower: $\eta\uparrow \quad \Rightarrow D\downarrow$., 2. A bigger molecule moves slower (i.e. it experiences more friction with the fluid): $a\uparrow \quad \Rightarrow D\downarrow$., and 3. Molecules move faster at higher temperatures: $T\uparrow \quad \Rightarrow D\uparrow$. Statements you will intuitively agree with – I expect.

Thus, proteins move slower than amino acids, ribosomes bound to mRNA slower than mRNA and ribosomes, etc. A typical diffusion coefficient of a protein in the cytosol of the cell equals $1-10\ \mu m^2/s$ [26].

2. **Time to travel the length of a cell in distance.** The time to move a distance d for a molecule equals

$$\tau = \frac{d^2}{6D}. \tag{2.7}$$

The length of *E. coli* is 2 μm and a representative diffusion coefficient is 5 $\frac{\mu m^2}{s}$ for a protein. So the time to travel this length in distance by a protein equals $\tau = \frac{(2\ \mu m)^2}{6\times 5\ \frac{\mu m^2}{s}} = \frac{4}{30}\ s = 0.13\ s$.

3. **Time for two proteins A and B to find each other.** This time is given by,

$$\tau = \frac{V_{cell}}{4\pi(D_A+D_B)(r_A+r_B)} \tag{2.8}$$

with V_{cell} as the cell volume, D_A and D_B the diffusion coefficients of the proteins, and r_A and r_B the (reaction) radii of the proteins. For an *E. coli* cell, we arrive at,

$$\tau = \frac{10^{-15}(10^{-1}m)^3}{4\times\pi\times 2\times 5\times\frac{(10^{-6}m)^2}{s}\times 2\times 5\times 10^{-9}m} = \frac{10^{-18}m^3}{400\times\pi\times 10^{-21}\frac{m^3}{s}} \approx 1\ s \tag{2.9}$$

So, it takes a protein 8 times longer to find another protein than that it takes for this protein to travel the length of *E. coli*! In other words, both proteins have crossed *E. coli* 8 times before they find each other. (This also means that when you calculate the time for two proteins to find each other and you do not take their initial position into account that you make at most an error of $\frac{1}{8}\times 100\% = 12.5\%$.)

2.9 Exercises

Diffusion lies at the basis of life. Without it, no movement of molecules would occur and reactions cannot take place. It also limits life, since molecules cannot move faster than by diffusion. The speed of reactions involving two (or more) substrates is therefore limited by diffusion rates.

1. Read the abstract of Klumpp et al. [13]. What limits growth rate in *E. coli* according to them?
2. The time for two molecules to find each other in a cell volume is given by equation 2.8. What do you expect happens to this time – the time when the first collision occurs – when you have N copies of each molecule instead of 1 of each? Why does the time decrease when one of the molecules is bigger?
3. Why is the collision time of two molecules (equation 2.8) lower to the waiting time for them to form a complex?
4. Calculate how many minutes that it takes for one copy of a transcription factor with radius 5 nm to find a promoter of a gene. The diffusion coefficient of the transcription factor is 5 $\mu m^2/s$. Assume a spherical cell with a radius of 1.5 μm. Why can you assume that the promoter does not move?
5. How much slower does a molecular complex move than any of its components? (With which factor is the diffusion coefficient reduced?)
6. Why is it advantageous for a cell to construct large complexes on DNA, rather than forming them first in the cytoplasm and after that having them bind to the DNA?

2.10 Rates of molecular reactions in cells

Different kinds of reactions occur in cells,

1. **Spontaneous degradation or conversion:** This means that a molecule or protein changes its state, structure, or conformation and that a life time for this molecule can be defined that characterises the time before this change occurs. If the starting state if called A and the end state B, we have the following reaction,

$$A \rightarrow B \tag{2.10}$$

   A could be a protein in an active conformation and B in an inactive

conformation. If the life time of A is $\tau = 30\ s$ then the rate, k, of A-to-B state changes is 2 per minute: thus, $\frac{1}{\tau} = 2\ \frac{1}{min} = k$. The 'k' constant is called a rate constant and it characterises the rate at which the conversion $A \rightarrow B$ occurs, per molecule of A. So, we loose A at rate,

$$\frac{d}{dt}a = -k \times a \tag{2.11}$$

Here a denotes the concentration of molecule A. We gain molecule B at rate,

$$\frac{d}{dt}b = k \times a = -\frac{d}{dt}a \tag{2.12}$$

Since the unit of $\frac{d}{dt}b$ is $\frac{concentration}{time}$, the unit of k should be $\frac{1}{time}$.

We can solve equation 2.11 for $a(t)$,

$$\begin{aligned} \frac{d}{dt}a(t) = -k \times a(t) \quad &\Rightarrow \quad \int_{a(0)}^{a(t)} \frac{1}{a(t)} da(t) = -k \int_0^t dt \\ \ln a(t) - \ln a(0) = -k \times t \quad &\Rightarrow \quad a(t) = a(0)e^{-k \times t} \end{aligned} \tag{2.13}$$

This result is called exponential decay.

How much time does it take to have half of the starting amount of A converted into B? In other words, we ask for the time $t_{1/2}$, which is defined as $\frac{1}{2} = e^{-k \times t_{1/2}}$, and $t_{1/2}$ equals $\frac{\ln 2}{k}$. As expected, a higher rate constant indeed shortens the life time of A molecules.

Life times of proteins can vary from minutes to hours and mRNAs typically live shorter, several minutes to a few tens of minutes.

2. **Spontaneous complex formation:** Proteins often bind to other molecules, such as other proteins or DNA. This corresponds to the following association reaction,

$$A + B \rightarrow AB \tag{2.14}$$

The rate at which AB is formed depends on: (i) the concentrations of A and B, denoted by a and b, (ii) the likelihood that they find each other by chance, and (iii) their tendency to form a complex when they have collided – their 'stickyness'. All of this can be captured in a single rate constant, which we denote by k_a. The rate v_a at which A and B for a complex is the same for both as they form a complex together,

$$v_a = k_a \times a \times b \tag{2.15}$$

The unit of k_a is $\frac{1}{concentration \times time}$.

We can also have dissociation,

$$AB \rightarrow A + B, \tag{2.16}$$

which occurs at rate,

$$v_d = k_d \times ab \tag{2.17}$$

Taken together, the change in the concentrations equals,

$$v = v_a - v_b \tag{2.18}$$

$$\frac{d}{dt}a = \frac{d}{dt}b = -v = -\frac{d}{dt}ab \tag{2.19}$$

In Chapter 5, we will show that k_a can be related to the diffusion coefficient – a 'mobility parameter' – of the molecules A and B. It takes seconds for single proteins to find each other or a DNA site in a single bacterial cell. So, such molecular search processes occur quite fast.

3. **Catalysed reactions:** Reactions that occur in signal transduction and metabolism are generally catalysed by enzymes (which are proteins). The rate of an enzyme-catalysed reaction depends on the reactant concentrations and the kinetic parameters of the enzyme. For instance, for the reaction,

$$S \rightleftharpoons P \tag{2.20}$$

the enzyme-catalysed rate of the reaction can be shown to be equal to,

$$v = V_{max} \frac{\frac{s}{K_s}\left(1 - \frac{p}{s \cdot K_{eq}}\right)}{1 + \frac{s}{K}_s + \frac{p}{K}_p} \tag{2.21}$$

This equation is called the reversible Michaelis-Menten equation and we will return to it later. K_s, K_p and K_{eq} are constants, the maximal enzyme rate V_{max} is also a constant and proportional to the enzyme concentration, e, and the intrinsic activity of the enzyme, k_{cat} (the catalytic rate constant),

$$V_{max} = k_{cat} \times e \tag{2.22}$$

Enzymes can have a k_{cat} of thousands of molecules per second, indicating that within one second they can catalyse thousands of reactions.

With these values we can calculate the maximal rate of an enzyme in *E. coli* that occurs at a concentration of 1 μM and has a k_{cat} of 1000 $\frac{1}{s}$. The number of product molecules produced per second in an *E. coli* cell equals $1000 \frac{product\ molecules}{enzyme \times s} 10^3 \frac{enzyme}{cell} = 10^6 \frac{products}{s \times cell}$. When we calculate this in terms of the concentration of product molecules then we get: $10^6 \frac{product\ molecules}{s \times cell} \frac{1}{6 \times 10^{23}} \frac{mol}{product\ molecules} \frac{1}{10^{-15}} \frac{cell}{liter} 60 \frac{s}{min} = \frac{60 \times 10^{-2}}{6} \frac{M}{min} = 100\ \frac{mM}{min}$ as the maximal rate of this enzyme. Considering that this enzyme will be inhibited by the product concentration p and not always have excess substrate $s \approx K_s$ such that the maximal rate cannot be attained; a more realistic estimate would be 30 $\frac{mM}{min}$, which is a realistic (but high) enzyme rate value in central metabolism.

Table 2.3: **Times and rates** (numbers for *Escherichia coli*, indicative for other bacteria)

genome replication time	40 min
RNA polymerase elongation rate	50 nts/s
ribosome elongation rate	18 aa/s
diffusion coefficient of a protein	1-10 $\mu m^2/s$
doubling time	from 40 min to several hours

2.11 Exercises

1. Consider the following reaction

$$F + P \rightleftharpoons FP$$

with F denoting a transcription factor and P the promoter of a gene. We will assume that the concentration of the transcription factor greatly exceeds that of the promoter, which occurs at a single copy per cell. Accordingly, we assume that the concentration of the unbound, "free" transcription factor is constant. We denote it by f_T. The total concentration of the promoter equals the sum of the concentrations of the free and occupied promoter, i.e. $p_T = p + fp$. We also assume that p_T is constant. All assumptions made until now are realistic.

   (a) Show that the rate of the association (complex formation) reaction can be written in terms of only one unknown promoter concentration, choose fp.

   (b) Show that the rate of the dissociation reaction can written in terms of only one unknown promoter concentration, choose fp.

   (c) Which rate difference equals $\frac{dfp}{dt}$?

   (d) Sketch the association rate and the dissociation rate as function of the concentration of fp. (Make a plot of the two rates as function of fp, since you do not know the parameter values you have think carefully about this.)

   (e) Can those two rates become equal? What happens then?

   (f) Calculate the concentration of fp when those two rates are equal. Sketch this concentration as function of f_T. Does the outcome make sense? Is it what you would expect?

Tip: how to sketch functions
You are often asked to sketch a function; for instance, sketch v as function of s,

$$v(s) = V_{max}\frac{s}{s+K_s}$$

without knowing the numerical values of the parameters V_{max} and K_s. You may have forgotten how to do this. You have to think about how the function, which you have to sketch, "works". You do this be figuring out its value when its argument – the variable along the x-axis of the plot, which you vary – equals 0 and infinity (denoted by ∞). Clearly, $v(0) = 0$. But what is $v(\infty)$? When $s = \infty$ its value greatly exceeds that of K_s, so $\lim_{s\to\infty}\frac{s}{s+K_s} = 1$. Thus, $v(\infty) = V_{max}$. If the function to be sketched was a line – when $v(s) = as + b$ – we would now be finished, as only two points fully specify a line. That is now not the case. We need to know more points than the two we have found so far. Let's check the function again. One obvious next point is when $s = K_s$ then $v(K_s) = \frac{1}{2}V_{max}$. Now we have three points, and since the function always rises as function of s, that is enough. Now you can sketch it! Knowing how to sketch functions is a basic skill in any quantitative science.

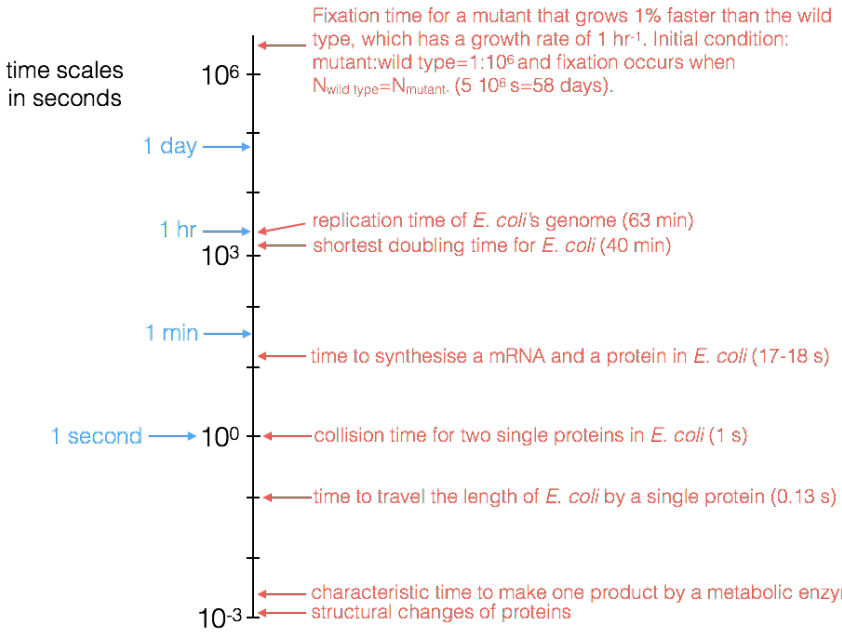


Figure 2.6: **Overview of some key time scales for bacteria, such as *E. coli*.**

2.12 Using the numbers: Cellular growth and division

1. **What is the time required for DNA replication?** The genome of *E. coli* has 4.6×10^6 base pairs and replication occurs at 600 $\frac{bp}{s}$ starting in two direction from the origin of replication – a site on the genome – this means that the time to replicate an *E. coli* genome amounts to $\frac{4.6\times10^6 \frac{\text{bps}}{\text{genome}}}{600\frac{\text{bp}}{\text{s}}\times 2} = 3833 \frac{\text{s}}{\text{genome}}$, which is $\approx 63\ min$.

2. **How much time does it take to make one mRNA by gene transcription?** The average amino-acid length of a protein is 300 amino acids. Since, each amino acid is encoded by three nucleotides, both on its mRNA and its gene, the mRNA and gene lengths are on average 900 nucleotides. RNA polymerases have an elongation rate of 50 $\frac{nucleotides}{s\times polymerase}$. So, it takes $\frac{900}{50} = 18\ s$ to transcribe a gene and make a transcript.

   The number of RNA polymerases that a single gene contains depends on the initiation frequency of transcription. If transcription initiation occurs more often than once per 18 s then more than one RNA polymerase can be expected on this gene.

   Interestingly, translation and transcription are coupled in many bacteria. This means that while the transcript is still being made by the RNA polymerase, ribosomes can already initiate translation. So, ribosomes should not move faster along the transcript than the elongation rate of a RNA polymerase as otherwise collisions would occur. Ribosomes move at a rate of 18 $\frac{aa}{s}$ which amounts to 54 $\frac{nucleotides}{sec}$ which is indeed slightly faster than RNA polymerase, but due to the time delay in translation initiation, which is a couple of seconds, collisions are not expected to occur.

3. **How much time does it take to make one protein from its mRNA?** The average protein length is 300 $\frac{aa}{mRNA}$ and the ribosome elongation rate is 18 $\frac{aa}{s\times ribosome}$, the time equals $\frac{300}{18}\frac{\frac{aa}{mRNA}}{\frac{aa}{s\times ribosome}} = 17 \frac{\text{s}}{\text{mRNA}}$. So, starting from gene induction, the time to make a protein takes at least: $18\ s + 17\ s = 35\ s$.

4. **Consider a mutant cell that grows 5% faster than the wild type cell it derived from. Assume we have 1 mutant cell and 10^6 wild type cells. How many generations does it take before an equal number of mutant and wild type cells is obtained?** Thus we ask for the time t that obeys $1e^{1.05\mu t} = 10^6 e^{\mu t}$. Now we have to solve for this time:

$$\begin{aligned} \ln(1e^{1.05\mu t}) &= \ln(10^6 e^{\mu t}) \Rightarrow \ln(1) + 1.05\mu t = \ln(10^6) + \mu t \\ \Rightarrow t &= \frac{\ln(10^6) - \ln(1)}{0.05\mu} \end{aligned}$$

And we know that $\mu = \ln 2/\tau_g$ with τ_g as the generation time. So,

$$t = \frac{\ln(10^6) - \ln(1)}{0.05\mu} = \frac{\ln(10^6)\tau_g}{0.05\ln(2)} = 398.6\tau_g, \quad (2.23)$$

thus the answer is that we have to wait for 398.6 generations. If the generation time is 1 hr then this means 16.6 days.

2.13 Using the numbers: Protein synthesis and partitioning

How many ribosomes are required to synthesise all cellular protein? An *E. coli* contains about 3×10^6 $\frac{proteins}{cell}$ and this amount needs to be synthesised within the doubling time of an *E. coli* cell, the time that it takes for *E. coli* to double its content and divide into two cells. Say we want to make 3×10^6 proteins – this is the amount we calculated above for the protein content of *E. coli* – in one hour (so, the growth rate equals $\ln 2$ hr^{-1}). An average protein contains 300 amino acids and ribosomes synthesise proteins at 18 aa/s. So, now we can ask how many ribosomes are required to attain this protein synthesis rate: 3×10^6 $\frac{proteins}{hr} 300 \frac{aa}{protein} \frac{1}{60} \frac{hr}{min} \frac{1}{60} \frac{min}{sec} = \frac{3\times10^6\times300}{60\times60} \frac{aa}{s} = 2.5 \times 10^5 \frac{aa}{s}$. This should be achieved by N ribosomes each working at rate $18 \frac{aa}{s\times ribosome}$. So, $2.5\times10^5 \frac{aa}{s} = N\ ribosomes \times 18 \frac{aa}{s\times ribosomes}$; therefore, $N = \frac{2.5\times10^5}{18} ribosomes \approx 14 \times 10^3\ ribosomes$. We know that ribosomes are only 80% active and 20% are always maturating; so, we expect about 17×10^3 $\frac{ribosomes}{cell}$ when the doubling time equals 1 hr. Again this is a realistic value.

2.14 Exercises

1. Calculate how many proteins fit in *E. coli*'s cell membrane assuming that their radius is 5 nm. Assume a radius of an *E. coli* cell of 1 μm. How many proteins fit in its periplasm if this compartment is 15 nm thick? What is the ratio of the protein numbers in the membrane and periplasm over the number of proteins in its cytoplasm?

2. Find the area of the earth and the distance of the moon to earth on Google. Calculate the time that it takes for a bacterium to cover the area of planet earth if the bacterium has an area of 1 μm^2 and grows exponentially at a rate of 1 hr^{-1}. How much time does it takes for this bacterium to fill the distance between the earth and the moon if it is 1 μm thick if all cells are stacked on top of each other?

3. **Stoichiometry of nutrient uptake fluxes during steady-state growth.** The elemental composition of an *E. coli* cell equals

$CH_{1.77}O_{0.49}N_{0.24}$. These elements are components of the molecules making up biomass such as DNA, RNA, lipids and proteins.

(a) Calculate the weight of mol of $CH_{1.77}O_{0.49}N_{0.24}$.

(b) A realistic value for the weight of *E. coli* cell 0.95 $pg = 0.95 \times 10^{-12} gram$. How many carbon and nitrogen atoms does a single *E. coli* cell contain?

(c) How many glucose and ammonium molecules are minimally required to make a single *E. coli* cell?

(d) When growing on glucose in mineral medium *E. coli* requires 5.9×10^9 ATP molecules to synthesise one cell. How many glucose molecules are required to attain this amount of ATP via respiration? How many via fermentation?

4. **The number of proteins made per mRNA.** Thinking about translation is a bit similar to thinking about a conveyor belt that breaks down quickly. So during the life time of the conveyor belt, during which it carries boxes from the left to the right, the conveyor belt manages to transport a number of boxes. Multiple boxes are one the belt, and the distance between the boxes on the belt is determined by the speed of the belt and the times between consecutive placements of boxes onto the belt. If boxes are placed on the belt at a higher rate then boxes are closer to one another and if the belt runs slower the boxes are also closer to one another. A box resembles a ribosome and the moving belt corresponds to a ribosome walking over mRNA. The distance between ribosomes is determined by the translation initiation rate and the moving rate of ribosomes. The life time of the mRNA correspond to the operating time of the conveyor belt before it breaks down and stops working.

   (a) Say a ribosome produces peptide chains of a length of 20 amino acids per second. How many mRNA nucleotides does it pass in a second?

   (b) What is the distance in nucleotides between neighbouring ribosomes on the mRNA if every two seconds a new ribosome hops into the mRNA?

   (c) If every two seconds a ribosome hops on to start translation at steady state, what is time period between ribosomes leaving the mRNA transcript?

   (d) What is then the protein synthesis rate?

   (e) What is the protein synthesis rate if two such mRNAs occur?

   (f) If an mRNA lives 16 seconds how many proteins are made from it?

   (g) Is the rate set by the elongation rate or the initiation rate in this exercise?

2.15 Extra Exercises (Tougher than exam questions!)

1. **Transcription and translation data in human cells.** In Schwanhäuger et al. [20, 19], the number of mRNA and protein copies per cell for more than 5000 genes of a human cell line is reported. The analysis of this data gives insight into the steady-state rates of transcription, mRNA degradation, translation, and protein degradation. This is a very insightful paper about the levels of thousands of gene products in a cell.

   (a) Assume a volume of 1800 μm^3 of an average human cell, which is realistic value. Calculate the mean concentration of the mRNA and protein in human cells given their mean number per cell: 17 $\frac{\text{mRNA molecules}}{\text{cell}}$ and 50000 $\frac{\text{protein molecules}}{\text{cell}}$.

   (b) The median translation rate is 180 $\frac{\text{proteins}}{\text{mRNA}\times\text{hour}}$. Consider the mean number of mRNA and protein molecules per cell and calculate the degradation rate constant of protein.

   (c) Calculate the mean transcription rate, given a mean mRNA half life of 9 hr and a mean mRNA concentration of 17 $\frac{\text{transcripts}}{\text{cell}}$.

   (d) Why do you think that the most stable mRNAs and proteins are involved in translation, respiration and central metabolism?

2. **Transcription and translation data in *E. coli*.** Taniguchi et al. [24] report data on 137 mRNAs and their cognate proteins. The mRNA number ranges from $0.05 - 5$ $\frac{\text{transcripts}}{\text{cell}}$ and the protein number from $0.1 - 5000$ $\frac{\text{proteins}}{\text{cell}}$.

   (a) How can a cell contain on average 0.05 mRNA molecules?

   (b) Assume a volume of 1 fl per cell and calculate the concentration ranges of mRNA and protein.

   (c) They found that essential proteins tend to have higher expression levels than nonessential proteins. Why does this make sense?

   (d) Do you think that a mother cells that divides and contains nine mRNA molecules will divide those evenly over its daughter cells at cell division? Consider 5 identical mothers cells with 10 molecules each. Do you think that division of those molecules at division over 2 daughter cells will always occur in such a manner that each daughter gets 5 molecules? What kind of statistics does molecule partitioning during cell division follow? (Hint: will you obtain 5 heads and 5 tails each time when you flip a coin 10 times?)

3. **Spontaneous genetic variation and competing bacteria.** Genetic variation occurs spontaneously in bacterial populations because copying errors are always made during replication. This leads to the spontaneous generation of mutants in bacterial populations that have slightly different properties. Some of those mutants are better adapted to the current environment, purely by chance. Those mutants will outgrow the resident population and can overtake the entire population such that their genotype becomes the dominant one at the expense of the resident genotype.

   We define the mutation probability as,

   $$p \quad = \quad \text{probability for a single base pair change in DNA} \qquad (2.24)$$

   Let's apply some elementary probability reasoning to make some inferences about mutation and selection.

   (a) What is expected genome size with a single base pair change?

   (b) What is the expected number of mutations in a genome of length N?

   (c) What should be the population size of bacteria with genome length N that contains all single mutations?

   (d) Bionumbers tells us that a *E. coli* culture at $OD_{600} = 0.1$ contains $10^8 \frac{cells}{ml}$. Take as the mutation rate $10^{-10} \frac{mutations}{nucleotide \times generation}$ (Lee, ..., Foster, PNAS, 2012). What is the expected OD_{600} of a 1 ml culture to contain all single mutations?

   Given these numbers, we conclude that during serial-batch evolution experiments should therefore quickly sample all single mutations, as they involve growth experiments in shake flasks of 100 ml that are grown to $OD_{600} = 1$, diluted by a factor of 10, regrown in a new flask to $OD_{600} = 1$, which repeated for 100s of times.

2.16 Key messages of this chapter

1. Many quantities are known for biological systems that are useful to make estimates of cellular properties using quite straightforward envelope calculations.

2. Many of those quantities can be found at `http://bionumbers.hms.harvard.edu`.

3. Careful usage of units of quantities allows you to construct simple equations for the calculations of insightful cellular properties.

4. Basic knowledge of physics and biochemistry is often enough to make useful approximations.

5. Thinking about cells from a quantitative perspective can give rise to new ideas, aid in experimental design and often indicate the relevance and feasibility of biological hypotheses.

Chapter 3

Kinetics of biochemical reactions

3.1 Reactions between molecules are the basic processes of life

Understanding how the molecules inside cells bring about cellular behaviour requires understanding of cellular activities in molecular terms such as signaling, metabolism, and gene expression. 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. So, genome sequencing alone is insufficient for understanding the working of a living organism.

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.

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 signaling networks. With this toolkit, you will already be able to understand the principles of a whole range of unexpected and sophisticated behaviours of molecular systems. We will first focus on the basic theory of mass-action kinetics in this chapter.

3.2 The quantitative description of molecular reactions

3.2.1 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 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 upon the formation of the complex or its dissociation. How fast those concentrations change will depend on the rate of these two reactions, 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 or 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 us 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 financial 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 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 it 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,

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

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, 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.

3.2.2 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 mass-action kinetics 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:

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

If the unit of the reaction rate is expressed in terms of mM/min, the unit of the concentration needs to be mM and the unit of the rate constants are then necessarily 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,

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

Confirm that: the unit of the rate constant k^+ should now be $min^{-1}mM^{-1}$. This rate constant is an example of a second-order rate constant, as its associated 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 $min^{-1}mM^{-2}$. 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 mM/min as unit. Obviously, a first-order reaction means $X \rightarrow Y$ and has a rate constant with 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,

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

The dissociation rate is given by $-v$.

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

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

the following rate equation for the reaction,

$$v = k^+ \prod_{i=1}^{s} x_i^{n_i} - k^- \prod_{j=1}^{p} y_j^{m_j} \tag{3.6}$$

The symbol $\prod$ means product,

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

Consider equation 3.6 again, it 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,

$$\begin{aligned} \frac{dx}{dt} &= -2(k^+ x^2 - k^- x_2) = -2v \\ \frac{dx_2}{dt} &= k^+ x^2 - k^- x_2 = v \end{aligned} \tag{3.8}$$

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 expect 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: $x_T = x(0) + 2x_2(0) = x(t) + 2x_2(t)$. This means that the consumption rate of x equals twice the production rate of x_2 and therefore $0 = dx/dt + 2dx_2/dt$ and $-dx/dt = 2dx_2/dt$ and this is true because $dx/dt + 2dx_2/dt = -2v + 2v$ (see equation 3.8)! These tricks we will apply very often to reaction systems.

3.2.3 Exercises

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

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

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

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

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

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

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

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

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

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

3.3 Linear growth

We consider the following reaction,

$$\xrightarrow{k} X, \tag{3.9}$$

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

$$\frac{dx}{dt} = k \tag{3.10}$$

So, per unit time we always 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,

$$\begin{aligned} \int_{x(0)}^{x(t)} dx &= \int_0^t k dt \\ \Rightarrow x(t) - x(0) &= kt \\ \Rightarrow x(t) &= x(0) + kt \end{aligned} \tag{3.11}$$

(See Figure 3.1.) 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.

3.4 Exponential growth

Now, we consider autocatalysis. For instance, a cell making a copy of 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,

$$\frac{d}{dt}x = kx, \tag{3.12}$$

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

$$\begin{aligned} \int_{x(0)}^{x(t)} \frac{1}{x}dx &= \int_0^t kdt \\ \Rightarrow \ln x(t) - \ln x(0) &= \ln\frac{x(t)}{x(0)} = kt \\ \Rightarrow x(t) &= x(0)e^{kt} \end{aligned} \tag{3.13}$$

(See Figure 3.1.)

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

$$\frac{x(t)}{x(0)} = 2 = e^{kt_d} \quad \Rightarrow \quad t_d = \frac{\ln 2}{k} \tag{3.14}$$

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

$$\begin{aligned} n_x(t_d) &= 2^1 n_x(0) \\ n_x(2t_d) &= 2^1 n_x(t_d) = 2^2 n_x(0) \\ n_x(3t_d) &= 2^1 n_x(2t_d) = 2^2 n_x(t_d) = 2^3 n_x(0) \\ &.... \\ n_x(gt_d) &= 2^g n_x(0) \end{aligned} \tag{3.15}$$

with g as the number of doublings or 'generations'.

3.5 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,

$$\frac{dx}{dt} = -kx \tag{3.16}$$

Solving this equation gives,

$$\begin{aligned} \int_{x(0)}^{x(t)} \frac{dx}{x} &= \int_0^t -k dt \\ \ln x(t) - \ln x(0) &= -kt \\ \ln \frac{x(t)}{x(0)} &= -k \cdot x \\ \frac{x(t)}{x(0)} &= e^{-kt} \end{aligned} \tag{3.17}$$

(See Figure 3.1.) 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.

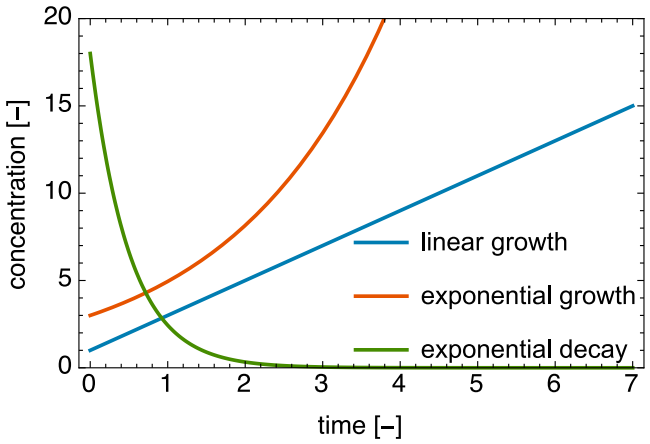


Figure 3.1: **Linear growth, exponential growth, and exponential decay.** Here we plot the following functions: linear growth $x(t) = 1 + 2t$, exponential growth $x(t) = 3e^{0.5t}$, and exponential decay $x(t) = 18e^{-2t}$.

3.6 Zeroth-order synthesis and first-order degradation

Clearly, the following models has broad applicability,

$$\begin{aligned} &\xrightarrow{k_1} X \xrightarrow{k_2 x}, \\ &\frac{dx}{dt} = k_1 - k_2 x \end{aligned} \tag{3.18}$$

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

$$x(t) = x(0)e^{-k2t} + (1 - e^{-k_2 t})\frac{k_1}{k_2} \tag{3.19}$$

We also know that when $dx/dt = 0$ that $x = \frac{k_1}{k_2} \equiv x_s$ with x_s as the steady state concentration of x, it will become clear in a moment what this means, so now have,

$$x(t) = x(0)e^{-k2t} + (1 - e^{-k_2 t})x_s. \tag{3.20}$$

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 ($t \to \infty$, or $t >> 1/k_2$) 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. 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,

$$x(t) = \underbrace{x(0)e^{-k2t}}_{\substack{\text{Exponential decay} \\ \text{of initial concentration}}} + \underbrace{(1 - e^{-k_2 t})x_s}_{\substack{\text{Combined synthesis and degradation} \\ \text{and eventually } x_s \text{ is reached}}} . \tag{3.21}$$

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!

3.6.1 Exercise

Consider the following system,

$$\underline{S} \underset{k_1^- y}{\overset{k_1^+ s}{\rightleftharpoons}} Y \underset{k_2^- p}{\overset{k_2^+ y}{\rightleftharpoons}} \underline{P} \tag{3.22}$$

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 3.2 should be helpful while doing this exercise.

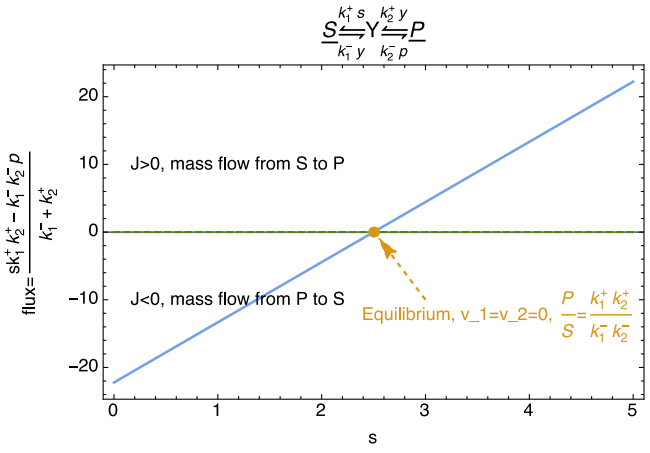


Figure 3.2: **Two reversible reactions in series: steady state, equilibrium state, and flux reversal.**

3.7 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,

$$\underline{S} \overset{1}{\rightleftharpoons} X \overset{2}{\rightleftharpoons} \underline{P} \tag{3.23}$$

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,

$$\frac{dx}{dt} = v_1 - v_2 = \underbrace{k_1^+ s - k_1^- x}_{v_1} - \underbrace{(k_2^+ x - k_2^- p)}_{v_2} \tag{3.24}$$

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 3.3D 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 3.3 depict the rates of the reactions as function of x. When x equals 0 the rate of reaction 1 equals $k_1^+ s$ and $-k_2^- p$ for reaction 2. The two rates equal zero at different concentration of x; reaction 1 at $k_1^+ s/k_1^-$ and 2 at $k_2^- p/k_1^+$. This you can conclude by setting each of the rates of zero and solving for x.

Suppose you supply an initial amount of x slightly larger than the intersection of the rate characteristic of the first reaction with the x-axis. At that concentration of X, $v_2 > v_1$ and the concentration of X will decrease because $dx/dt < 0$. The rate with which x decreases becomes smaller as it approaches the intersection between the two rate curves because dx/dt gets smaller. This allows a sketch of the dynamics of x, in a plot of 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,

$$x_S = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} \tag{3.25}$$

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 rates 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 3.3 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,

$$\frac{p}{s} = \frac{k_1^+ k_2^+}{k_1^- k_2^-} \tag{3.26}$$

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.

3.7.1 Exercises

1. Sketch the dynamics of X as function of time on the basis of the rate characteristic; take $k_1^+ = 5, k_1^- = 1, k_2^+ = 3, k_2^- = 2$. Show that equation 3.26 indeed causes the system to settle to an equilibrium state where all reactions rate equal zero. Show that X then has the same stationary concentration as for the system $\underline{s} \rightleftharpoons x$. Show that the time to reach half the steady-state concentration is halved when all rate constants are doubled in value.

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

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

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

3.8 Rate characteristics, attracting states, and dynamics

Using Figure 3.3 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 3.3, the left figures refer to the system $\underline{S} \rightleftharpoons X$,

$$\begin{aligned} \frac{dx}{dt} &= v = k^+ s - k^- x \\ x_T &= s(t) + x(t) = s(0) + x(0) \quad \Rightarrow \quad 0 \leq x \leq x_T \end{aligned} \tag{3.27}$$

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 3.3A 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 3.3. There $v_1 = v_2 \neq 0$ at the stationary state and, hence, a steady state occurs.

Figure 3.3B & 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 3.3C & 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 3.3 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.

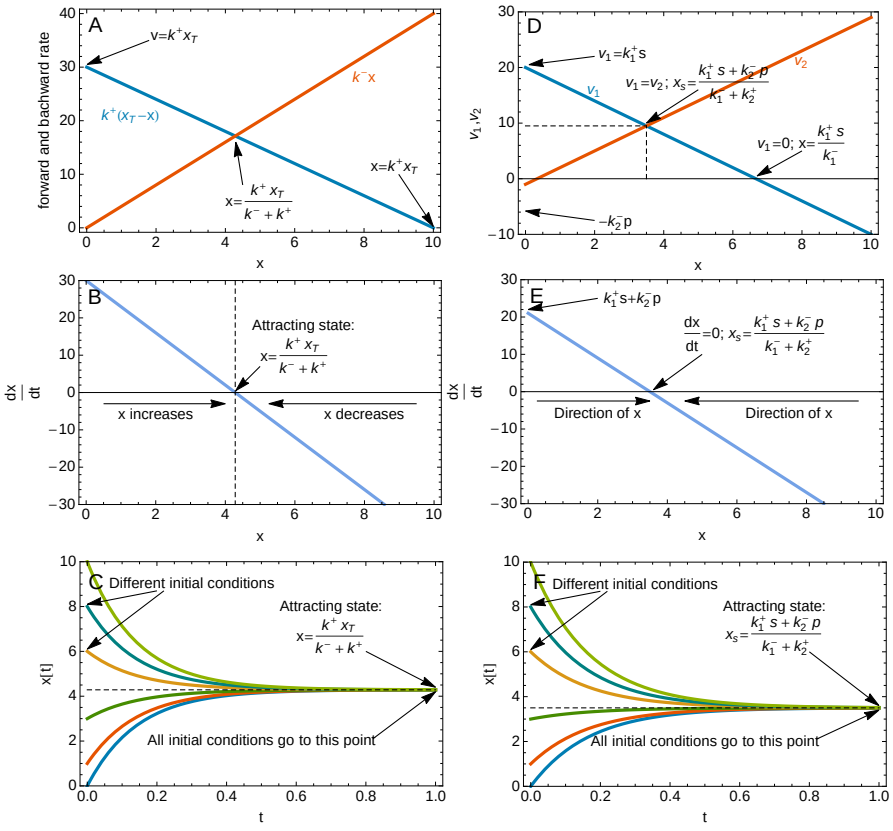


Figure 3.3: **Rate characteristics, stability of the final state, and dynamics.** On the left the system $\underline{S} \rightleftharpoons X$ is considered and on the right $\underline{S} \rightleftharpoons X \rightleftharpoons \underline{P}$ The upper figures show the rate characteristics; the dependency 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.

3.9 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 cytosolic protein and a membrane receptor, that can form a complex,

$$A + B \rightleftharpoons AB \tag{3.28}$$

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 assume that B is in excess. This means we only have to deal with the conservation of A molecules: $a_T = a + ab$ (because $ab << b_T$ 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,

$$\frac{da}{dt} = k_1^-(a_T - a) - k_1^+ a \cdot b \tag{3.29}$$

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,

$$a_T = a + \frac{k_1^+ a \cdot b}{k_1^-} = a\left(1 + \frac{b}{K_D}\right) \tag{3.30}$$

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,

$$a = \frac{a_T}{1 + \frac{b}{K_D}} \tag{3.31}$$

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,

$$ab = \frac{a_T \frac{b}{K_D}}{1 + \frac{b}{K_D}} = a_T \frac{b}{K_D + b} \tag{3.32}$$

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 $1/K_D$.

3.9.1 Exercises

1. Plot ab as function of b. What type of relationship do you find? What is the ratio of b/K_D where 10% and 90% of A is in the complex?
2. The K_D of a transcription factor for a DNA binding site is 1 nM. What is the concentration of the transcription factor such that bound fraction of binding sites is by 10%, 50% and 90%?
3. Consider the following reactions:

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

   Define a K_D for the first reaction and the second reaction. Do you understand that those can indeed be different? Assume 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:

$$\begin{aligned} A + B &\rightleftharpoons AB \\ A + AB &\rightleftharpoons A_2B \\ A + A_2B &\rightleftharpoons A_3B \end{aligned} \tag{3.33}$$

   (a) At what concentration of A is 50% of B in the A_3B complex?
   (b) At what concentration of A is 50% of B in the A_2B complex?
   (c) What is then the fraction of B in the AB and the A_3B complex?

3.9.2 Exercises

1. **Steady state versus equilibrium state** Many biological systems attain steady states. The principles of steady states are therefore very important and you will study them in this exercise. Consider the following reactions and assume them to follow mass action kinetics

$$S \overset{1}{\rightleftharpoons} X \tag{3.34}$$

$$X \overset{2}{\rightleftharpoons} P \tag{3.35}$$

We will focus on X. The rate of reactions 1 and 2 are denoted respectively by v_1 and v_2. Here we use the convention that the concentration of molecules are written in small font and the name of the molecule in capitol font.

(a) Give the mass balance for the concentration of X.

(b) When has X attained a steady state?

(c) Can this happen when the concentrations of S and P are not fixed?

(d) What is the name of the state that the system attains when S and P are not fixed?

(e) Calculate the concentration of X in the final state when S and P are not fixed, assuming reversible mass-action kinetics.

(f) When S and P are not fixed, the concentration X can only become constant if the rates of reactions 1 and 2 are each equal to zero. Derive the two equations that relate the concentration ratio of $\frac{x}{s}$ of $\frac{p}{x}$ in this equilibrium state. Define those ratio's as equilibrium constants K_1 and K_2

(g) Rewrite the rate of reaction 1 and 2 in terms of the constant K_1 and K_2 and show that the rates are negative when $\frac{x}{s} > K_1$ and $\frac{p}{x} < K_2$ and positive when $\frac{x}{s} < K_1$ and $\frac{p}{x} < K_2$. What happens if a reaction rate changes sign?

(h) Fix S and P and solve the differential equation with $x(0) = x_0$ as initial state.

(i) What is the concentration of X when time become very large?

(j) Are the rates of the reactions necessarily unequal to zero in the final state or does it depend on the choice of the value of the fixed concentrations of S and P?

(k) Express the concentration of X in the final state in terms of the system parameters.

(l) Show that at the steady state, when S and P are fixed and the rate of X synthesis and degradation are equal, that the direction of mass flow – so from S to P or vice versa – depends on the ratio of $\frac{p}{s}$.

3.10 Extra Exercises (Tougher than Exam Questions!)

1. **Enzyme kinetics.** Enzymes are the workhorses of a cell. Essentially all reactions are catalysed by them. They speed up reactions by offering a favourable physicochemical environment in their catalytic site for the reaction to occur. Without the enzyme the reaction would also take place, as enzymes cannot change the equilibrium constant of a reaction, but the reaction rate would be orders of magnitude slower. So one way to envision cellular metabolism is that a cell selects reactions that are favourable for its fitness, by expressing the associated enzymes that can catalyse those reactions, out of all possible reactions. In this exercise, we will think about the principles of enzyme catalyses. Consider the following enzyme-catalysed conversion,

$$\underline{S} + E \underset{v_1^-}{\overset{v_1^+}{\rightleftharpoons}} ES \xrightarrow{v_2} E + \underline{P} \tag{3.36}$$

   The underlines indicate that the concentrations of S and P are held fixed.

   (a) Give the mass balances for the concentrations of the variable species in the model in terms of rates of the reactions.

   (b) An enzyme is a catalyst that is not spent during the reaction. This has one particular consequence for the concentration of enzyme in the system. What is this consequence?

   (c) Why is $\frac{de}{dt} + \frac{des}{dt} = 0$?

   (d) Express the rate of the reactions in terms of mass action kinetics.

   (e) Give the units of all the terms appearing in the mass balance equation with mass action kinetics.

   (f) Solve for es at steady state.

   (g) The steady-state rate of the enzyme is defined as $v = k_2 es$. Express this enzyme rate in terms of s and identify the combination of constants that you have to make in order to write this enzyme rate equation into its more familiar form,

$$v = V_{max} \frac{s}{K_M + s} \tag{3.37}$$

   (h) Studying $\frac{1}{v}$ leads to an intuitive understanding of how an enzyme works (we set e_T to 1),

$$\frac{1}{v} = \frac{k_2}{k_2 k_1^+ s} + \frac{k_1^-}{k_2 k_1^+ s} + \frac{k_1^+ s}{k_2 k_1^+ s} = \frac{1}{k_1^+ s}\left(1 + \frac{k_1^-}{k_2}\right) + \frac{1}{k_2} \tag{3.38}$$

$\frac{1}{v}$ now corresponds to the waiting time for 1 enzyme to convert one molecule of S into P. If $k_2 >> k_1^-$ then $ES \rightarrow E + P$ nearly always occurs, rather than $ES \rightarrow E+S$, and the waiting time for the reaction becomes,

$$\frac{1}{v} = \underbrace{\frac{1}{k_1^+ s}}_{\substack{\text{time to} \\ \text{bind}}} + \underbrace{\frac{1}{k_2}}_{\substack{\text{catalysis} \\ \text{time}}} \tag{3.39}$$

Which makes intuitive sense, both reactions have to occur before a P molecule appears. Why does the $\left(1 + \frac{k_1^-}{k_2}\right)$ factor appear when k_2 is not much larger than k_1^-?

2. **The chemostat for culturing of cells.** The chemostat is a bioreactor set up that allows for the continuous steady-state cultivation of cells, it keeps the cells at a steady-state growth rate. The chemostat state is sometimes called a continuous culture. The concept is that medium flows into the reactor from a medium vessel at a fixed flow rate F, expressed in $\frac{liter}{hr}$. The volume of the culture V is kept fixed by flowing out medium, including cells, from the bioreactor into an exhaust vessel at the same rate. The dilution rate D is now defined as $D = \frac{F}{V}$. Medium leaves and enters the vessel at this rate. You can therefore think of D as a rate constant. The medium vessel contains the growth substrate, which limits growth, e.g. glucose, at a concentration s_m. Since the cells in the bioreactor consume this substrate the concentration in the bioreactor, s will be smaller than the concentration in the medium vessel: $s < s_m$. The cells have a specific growth rate μ that depends on the concentration of the growth-limiting substrate in the reactor. We model it as,

$$\mu = \mu_{max} \frac{s}{K_s + s} \tag{3.40}$$

with $1/K_s$ as the affinity of the organism for the growth-limiting substrate. This equation is often called the Monod equation with μ_{max} as the maximal growth rate and K_s as the Monod constant. A model of the chemostat contains minimally two variables: the concentration of substrate (in mol/l) and cells (in $gram/liter$) in the bioreactor. For those variables we can write the following mass balances,

$$\frac{ds}{dt} = D(s_m - s) - \frac{1}{Y_{x/s}} \mu x \tag{3.41}$$

$$\frac{dx}{dt} = (\mu - D)x \tag{3.42}$$

with $Y_{x/s}$ as the yield of biomass on substrate in $\frac{gram\ biomass}{mol\ substrate}$.

(a) Explain the meaning of all the terms in the balances: Ds_m, $-Ds$, $-\frac{1}{Y_{xs}}\mu x$, μx and of $-Dx$.

(b) At steady state the specific growth rate μ equals the dilution rate D. Why does this make sense?

(c) Express the steady state concentration of biomass x_s and of growth substrates$_s$ in terms of model parameters and plot their dependences on the dilution rate.

(d) What is the maximal possible value of D at which cells still occur in the chemostat?

(e) What is concentration of growth-limiting substrate in the bioreactor when the maximal D is reached?

3. **Persister cells in bacterial populations** In the last few years, it has become clear that populations of many bacterial species consist of two subpopulations. One that is growing and another that is non-growing – 'dormant'. The non-growing cells are more stress resistant than the growing cells. Antibiotic resistant bacteria are often dormant cells and since they persist antibiotic conditions, or stress conditions, they are generally referred to as 'persister cells'. A single growing cell can switch to become persister and a persister can switch to become a growing cell. Clearly, the persister cells do not grow. We therefore have the following three processes: i. growth of a cell in the growing state, ii. a cell in the growing states that switches to the persister state, and iii. a persister state cell that turns into a growing cell state. We use the following notation: $\mu =$ growth rate, $k_p =$ the rate constant for switching from the growing to the persister state, $k_g =$ the rate constant for switching from the persister state to the growing state, the concentration of growing and persister cells equals g and p, respectively.

   (a) What are the mass balances for the concentrations of growing and persister cells?

   (b) Why does this system never settle to a steady-state concentration of the cell states?

   (c) Do you think the fraction of persister and growing cells becomes fixed over time?

   (d) The balance for the fraction of persister cells, ϕ, is given by,

$$\frac{d\phi}{dt} = \frac{d}{dt}\underbrace{\left(\frac{p(t)}{g(t)+p(t)}\right)}_{\phi} = \underbrace{k_p + \mu\phi^2}_{\text{Synthesis rate}} - \underbrace{(k_p + k_g + \mu)\phi}_{\text{Degradation rate}} \tag{3.43}$$

   The steady state fraction equals,

$$\phi_s = \frac{k_d + k_g + \mu - \sqrt{(-k_d - k_g - \mu)^2 - 4k_d\mu}}{2\mu} \tag{3.44}$$

This fraction was obtained by setting the previous equation to zero and solving for ϕ. Derive this equation yourself.

(e) Calculate the steady state fraction. Use realistic parameters: $\mu = 1\ hr^{-1}$, $k_g = \frac{1}{10}\ hr^{-1}$, and $k_d = \frac{1}{100}\ hr^{-1}$.

(f) Show that the persister fraction indeed moves towards a stable steady state by plotting the synthesis rate and degradation rate as function of ϕ. Use the same parameters as in the previous question.

3.11 Key messages of this chapter

1. Signal transduction is the process by which cells sense and integrate environmental cues and integrate this with the cellular state.
2. Signal transduction done by sophisticated regulatory proteins that change in activity by way of covalent modification (e.g. phosphorylation) and conformation changes.
3. Such regulatory proteins form together the cellular signal transduction network that senses and integrates information and induces and monitors the progress of cellular responses, for instance involving changes in gene expression.
4. Mathematical models are very useful in helping you think about the capabilities and limitations of molecular mechanisms of signal transduction.

Chapter 4

Signal transduction

4.1 Introduction

All cells experience changes in their environment that they have to respond to in order to stay alive and competitive as single cells, or perform their functions in a multicellular context. Changes in nutrients force cells to change their metabolism to be able to sustain themselves and grow. Changes in temperature, osmotic pressure or other stresses force cells to make compensatory responses to repair and protect themselves. These changes are all imposed on cells from their environment. Therefore, cells need to sense their environment, integrate this information, and induce adaptive responses. Such responses often involve regulation of gene expression, which we will discuss in the next chapter. In this chapter, we focus on the quantification of molecular processes underlying signal perception and transduction.

4.2 Exercises

1. In signal transduction, proteins are often activated by phosphorylation, this causes them to change shape and have different binding affinities and enzymatic activities. They are then inactivated by dephosphorylation. Kinase phosphorylate proteins and phosphatase dephosphorylate them.

   (a) What is the reaction catalysed by a kinase?

   (b) What is the reaction catalysed by a phosphatase?

   (c) If we keep the concentrations of ATP, ADP, Pi, kinase and phosphatase constant which differential equations would you need to model the activation of protein by a kinase and its inactivation by a phosphatase?

(d) What would be suitable enzyme kinetics for the kinase and the phosphatase?

(e) Explain why one differential equation is sufficient for the dynamic description of this process.

(f) Do you think that this system reaches a steady state or an equilibrium state?

(g) Is this state stable? What if is not?

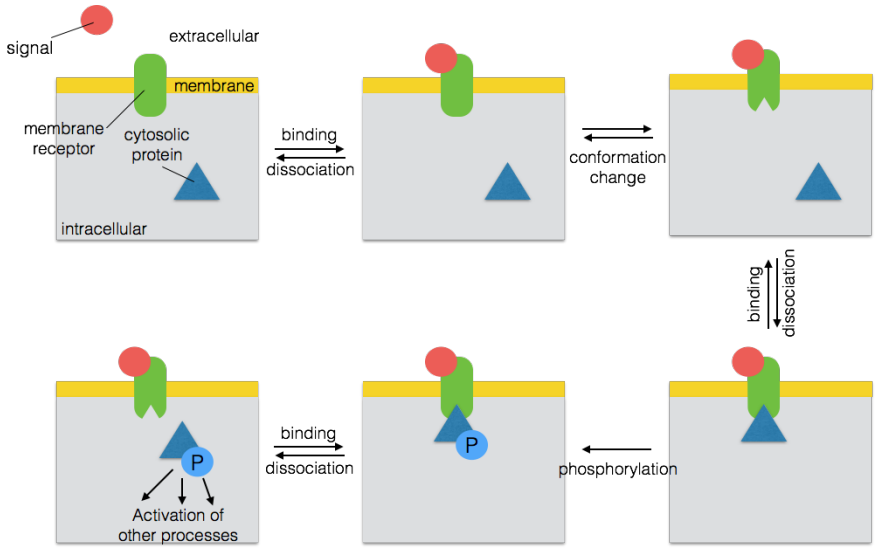


Figure 4.1: **Basic mechanism of receptor activation and signal transmission.** Although the precise molecular details differ between signalling systems, the basic principles are captured in this scheme.

4.3 Activation of membrane receptors and cytosolic signalling proteins

4.3.1 Signal binding and receptor affinity

We consider a signal S and a receptor R at concentrations s and r, respectively. Their association and dissociation is captured by the following reaction,

$$\underline{S} + R \underset{-}{\overset{+}{\rightleftharpoons}} SR \tag{4.1}$$

$$v^+ = k^+ \times s \times r \qquad \text{Rate of binding} \tag{4.2}$$

$$v^- = k^- \times sr \qquad \text{Rate of dissociation} \tag{4.3}$$

$$v = v^+ - v^- \qquad \text{Net rate of binding} \tag{4.4}$$

We assume S to be in excess, which we denote by an underline in the reaction. The rate constant k^+ and k^- are fixed and are properties of the molecules R and S. The net rate of binding gives the number of SR molecules formed per unit time. The mass balances are given by,

$$\frac{d}{dt}r = -v \tag{4.5}$$

$$\frac{d}{dt}rs = v \tag{4.6}$$

Since R is not in infinite supply and S exceeds R in amount, we should also consider that the total concentration of R (r_T) remains fixed:

$$r_T = r + rs \quad \Rightarrow -\frac{d}{dt}r = \frac{d}{dt}rs \quad : \text{Loss of } R = \text{Gain of } RS \tag{4.7}$$

Taken together we arrive at,

$$\frac{d}{dt}rs = \underbrace{k^+ \times (r_T - rs) \times s}_{\text{binding rate}} - \underbrace{k^- \times rs}_{\text{dissociation rate}} \tag{4.8}$$

When you have never seen such a 'mass-balance equation' before, then it is instructive to compare the concentration rs to money in a bank account:

$$\text{change in money amount in your bank account at time } t = \frac{d}{dt}\text{money}$$

$$\frac{d}{dt}\text{money} = \text{earning rate} - \text{spending rate} = v_{earning} - v_{spending}$$

Clearly, when $v_{earning} > v_{spending}$ you gain money at time t, i.e. $\frac{d}{dt}money > 0$. Alternatively, $\frac{d}{dt}money < 0$ if $v_{earning} < v_{spending}$ or $\frac{d}{dt}money = 0$ if $v_{earning} = v_{spending}$. In the latter case, your bank account is "in balance". We can make this analogy between money and concentration because both quantities are additive. Likewise,

1. the concentration rs drops when the dissociation rate is higher than the binding rate: $\frac{d}{dt}rs < 0$ if $v_{dissociation} > v_{binding}$,

2. the concentration rs stays constant when the binding and the dissociation rate are equal: $\frac{d}{dt}rs = 0$ if $v_{binding} = v_{dissociation}$, and

3. the concentration rs rises when the binding rate exceeds the dissociation rate: $\frac{d}{dt}rs > 0$ if $v_{binding} > v_{dissociation}$.

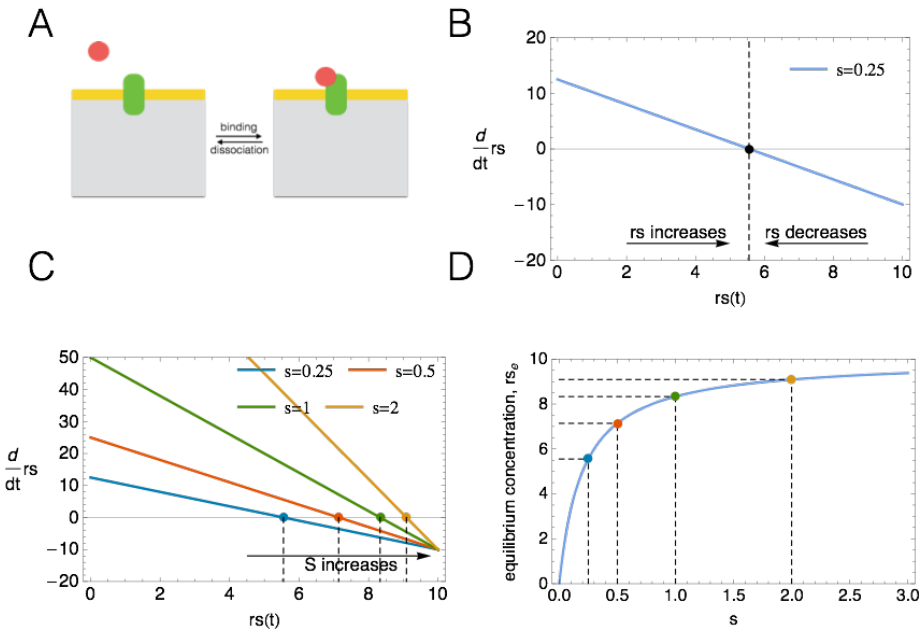


Figure 4.2: **Receptor-signal binding.** **A**. Picture of the binding reaction. **B.** For all starting values of rs, the dynamics goes to the same end state – the equilibrium concentration – indicated by the black point. We choose $r_T = 10$; so, $0 \leq rs \leq 10$. **C.** The equilibrium point increases when more S is present. **D.** The dose-response curve of rs as function of s. Here we considered: $k^+ = 5,\ r_T = 10,\ k^- = 1$.

Interestingly, for a fixed signal and total receptor concentration the reaction $S + R \underset{-}{\overset{+}{\rightleftharpoons}} SR$ always evolves to the same concentrations r and rs at which it then stays. This is very simple to understand from the plot of $\frac{d}{dt}rs$ as function of rs shown in figure 4.2B: when rs is smaller than the value of rs at the black dot then $\frac{d}{dt}rs > 0$ and rs rises until $\frac{d}{dt}rs = 0$ at the black dot; whereas if rs is higher than the black dot value of rs, $\frac{d}{dt}rs < 0$ and rs drops until it reaches the black dot value. So, for all starting concentration of rs the same end state is reached eventually. This is also illustrated in figure 4.3.

For the parameter values used in figure 4.2B it is easy to calculate the equilibrium value of rs denoted by rs_e where $v^+ = v^-$:

$$\overbrace{\underbrace{5}_{k^+} \times \underbrace{0.25}_{s} \times (\underbrace{10}_{r_T} - rs_e)}^{v^+} = \overbrace{\underbrace{1}_{k^-} \times rs_e}^{v^-} \quad \Rightarrow rs_e = 5.55 \tag{4.9}$$

We can use the same trick to solve rs as function of s,

$$\overbrace{k^+ \times s \times (r_T - rs_e)}^{v^+} = \overbrace{k^- \times rs_e}^{v^-}$$
$$\Rightarrow rs_e = r_T \frac{s}{K_D + s} \quad \text{with} \quad K_D = \frac{k^-}{k^+} \tag{4.10}$$

K_D is called the dissociation constant and the affinity of receptor for the signal is defined as $\frac{1}{K_D}$. When the signal concentration equals the dissociation constant, i.e. when $s = K_D$ then $rs = \frac{r_T}{2}$. Hence, the K_D is sometimes called the half-saturation constant or EC_{50} (50% effect concentration). The equilibrium concentration as function of s, the equation 5.6, is shown in figure 4.2D.

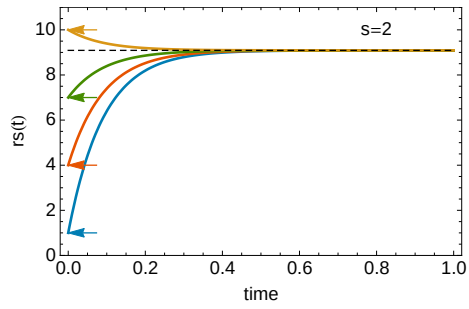


Figure 4.3: **The dynamics of rs as function of time: regardless of the starting values of rs (marked with the arrows), it always ends up at the same equilibrium concentration rs_e.** Here $s = 2$. The horizontal black dotted line is the vertical dotted line as in figure 4.2B.

It is important that you take a moment to study the differences between the plots of figure 4.2 and the conclusions:

1. **Figure B:** Indicates whether rs increases or decreases at a particular value. Regardless of the values for rs that you take, it always ends up in the equilibrium state where $v^+ = v^-$ and $rs = rs_e$ (see also figure 4.3).

2. **Figure C:** The equilibrium value of rs depends on s and increases with s.

3. **Figure D:** The equilibrium value of rs, rs_e increases in a hyperbolic manner with s.

4.3.2 Conformation change equilibrium

Many regulatory proteins, such as membrane sensors and DNA-binding transcription factors, change conformation upon activation. In our example (figure 4.1) this means that RS changes conformation to a state that the triangular cytosolic protein can bind to. This means that some fraction of the total amount of RS, which equals $rs_e = r_T \frac{s}{K_D+s}$, is in the active state and the remainder is the inactive state,

$$rs_e = \underbrace{rs^i}_{\text{inactive}} + \underbrace{rs^a}_{\text{active}} = rs^a \left(1 + \frac{rs^i}{rs^a}\right) \tag{4.11}$$

We consider the following reversible conformation change reaction,

$$\begin{aligned} rs^i &\underset{inactivation}{\overset{activation}{\rightleftharpoons}} rs^a \\ \text{activation rate} &= v_a = k_a \times rs^i \\ \text{inactivation rate} &= v_i = k_i \times rs^a \\ \text{net activation rate} &= v = v_a - v_i \end{aligned} \tag{4.12}$$

When $v_a = v_i$, we can solve for the ratio $\frac{rs^i}{rs^a}$ that appears in equation 4.11,

$$k_a \times rs^i = k_i \times rs^a \quad \Rightarrow \frac{rs^i}{rs^a} = \frac{k_i}{k_a} = L \tag{4.13}$$

The constant L is called the conformation equilibrium constant. Combining equations 4.11 and 4.13 leads to expression of the active concentration of the receptor rs^a in terms of the receptor-signal affinity (K_D), the signal concentration (s), and conformation equilibrium constant (L),

$$rs^a = rs_e \frac{1}{1+L} = r_T \frac{s}{K_D + s} \frac{1}{1+L} \tag{4.14}$$

If you are interested in how such models are being applied in drug discovery read Wootten et al. [27]. For a more fundamental perspective see for instance Kasai and Kusumi [9].

4.4 Kinase and phosphatase pairs: ultrasensitivity

4.4.1 Phosporylation-dephosphorylation steady state

Phosphorylation of regulatory proteins occurs often in cell biology. Simple models have contributed a lot to our understanding of how phosphorylation of proteins can give rise to variable degrees of signal sensitivities of cells. To understand this, we have to study the following two reactions where a target protein

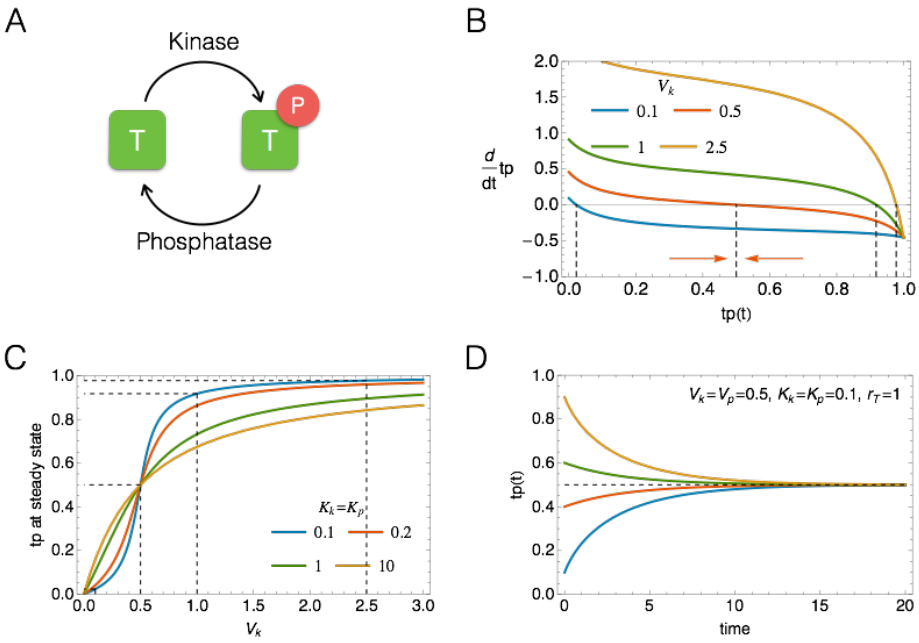


Figure 4.4: **Target protein phosphorylation and dephosporylation. A**. Picture of the reaction system. **B.** Plot of $\frac{d}{dt}tp$ as function of tp for four different values of the maximal rate of the kinase (V_k). In each V_k-case, the system evolves to the same final concentration of tp where $\frac{d}{dt}tp = 0$. This is explicitly shown for the red scenario with the arrows that indicate when the concentration of tp rises or falls. The dashed lines indicate the end concentration of tp – the steady-state concentration – for the four V_k values. At higher V_k values, higher tp values are attained at steady state. Here we considered: $K_k = 0.1$, $K_p = 0.1$, $V_p = 0.5$, $r_T = 1$. **C.** The dose-response curve of the steady-state value of tp as function of V_k for different values of the K_k and K_p. When K_p and K_k are small the dose response curve becomes steeper and the system becomes more sensitive to changes in the maximal rate of the kinase (V_k). Here: $V_p = 0.5$, $r_T = 1$. **D.** Illustration of the fact that the steady state concentration reached does not depend on the initial concentration of TP; the steady state concentration only depends on the parameters of the enzymes.

T is phosphorylated and dephosphorylated,

$$\begin{aligned} \text{Phosphorylation:} \quad & T + \underline{ATP} \xrightarrow{kinase} \underline{ADP} + TP \\ \text{Dephosphorylation:} \quad & TP \xrightarrow{phosphatase} T + \underline{P} \end{aligned} \tag{4.15}$$

The phosphorylation and dephosphorylation reaction do not occur spontaneously, but are catalysed by a kinase and phosphatase respectively. We do not consider the concentrations of ATP, ADP, and P and omit those from now onwards. Since we are considering enzyme catalysed reactions, we have to deal with enzyme kinetics to describe the phosphorylation and dephosphorylation rates. We

take the simplest description, called Michaelis-Menten kinetics,

$$\text{Kinase kinetics, phosphorylation rate: } v_k = V_k \frac{t}{t + K_k} \quad (4.16)$$

$$\text{Phosphatase kinetics, dephosphorylation rate: } v_p = V_p \frac{tp}{tp + K_p} \quad (4.17)$$

We note that the total target protein concentration $t_T = t + tp$ remains fixed because t is only interconverted between two forms; therefore, $t = t_T - tp$. Taken together, the rate of change of the phosphorylated concentration is given by,

$$\frac{d}{dt}tp = v_k - v_p = V_k \frac{t_T - tp}{t_T - tp + K_k} - V_p \frac{tp}{tp + K_p} \quad (4.18)$$

Again we can study $\frac{d}{dt}tp$ as function of tp to figure out at which concentrations of tp the concentration of tp rises ($\frac{d}{dt}tp > 0$), remains constant ($\frac{d}{dt}tp = 0$), and drops ($\frac{d}{dt}tp < 0$). We consider different maximal activities of the kinase, V_k, to simulate the effect of some activating signal on this enzyme. The plot of $\frac{d}{dt}tp$ as function of tp is shown in figure 4.4B. Figures 4.4B,D show that the system always evolves to the same final concentration of tp. At this concentration $v_p = v_k$ and $\frac{d}{dt}tp = 0$; since $v_p = v_k \neq 0$ this state is called a steady state (and not an equilibrium state then $v_p = v_k = 0$). At this steady-state state holds that,

$$v_k = v_p \quad \Rightarrow \quad V_k \frac{t_T - tp_s}{t_T - tp_s + K_k} = V_p \frac{tp_s}{tp_s + K_p} \quad (4.19)$$

with tp_s as the steady-state concentration of TP. In principle, we can solve the previous equation for tp_s and obtain an expression as function of V_k that describes the dependency shown in figure 4.4D. This is not very complicated but we do not do this here.

4.4.2 Signal ultrasensitivity

Figure 4.4C shows the dose-response curve of the steady-state concentration of tp as function of V_k. The blue line corresponds to the parameter settings of figure 4.4A. Figure 4.4C indicates that for lower values of the K_p and K_s values the dose-response curve becomes steeper. In fact, the steepness of the curve depends on the K values relative to r_T. We will not analyse this in detail here.

Ultrasensitivity occurs when $\frac{V_k}{tp}\frac{\partial tp}{\partial V_k} > 1$; indicating that a change in V_k of 1% leads to a response in the steady-state concentration of tp that exceeds 1%. So, the system acts an amplifier of signal changes. This illustrates that a cell can change the sensitivity of its signalling networks by changing expression of its target proteins: when their concentrations exceed the K_M values of their kinases and phosphatases, the phosphorylated concentration of the target protein becomes ultrasensitive to changes in regulators of kinase activity.

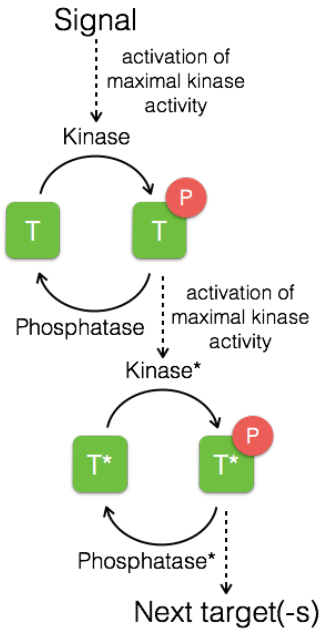


Figure 4.5: **A signal transduction cascade.** A signal change activates a first kinase that phosphorylates a target protein. As a result, a new steady-state concentration of tp is established resulting in a changes tp concentration, δtp. The change in the concentration of the phosphorylated target protein, which acts as a kinase or activates a second kinase, brings about a change in the maximal phosphorylation activity V_k^*, at which a new target protein T^* can be phosphorylated. The resulting increased phosphorylated rate causes the steady state concentration of tp^* to change by δtp^*.

4.5 Signal transduction cascades: sensitivity amplification

The slope in figure 4.4D translates a change in V_k, δV_k into a change in tp, δtp,

$$\delta tp = \underbrace{\frac{\partial tp}{\partial V_k}}_{\text{slope}} \delta V_k \tag{4.20}$$

If tp activates another kinase by changing its maximal rate, V_K^*, then the change in this kinase activity parameter is,

$$\delta V_k^* = \frac{\partial V_k^*}{\partial tp} \delta tp \tag{4.21}$$

If this kinase, together with a phosphatese, activates another target protein tp^* then the change in the concentration of this next target protein upon a change

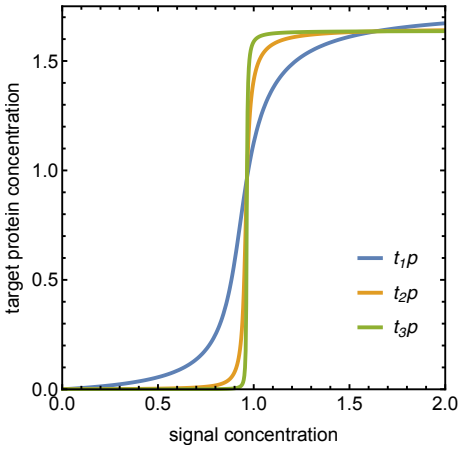


Figure 4.6: **Sensivity amplification along a signalling cascade:** $s \rightarrow t_1p \rightarrow t_2p \rightarrow t_3p$**.** Here we considered a cascade of three phosphorylation events, each with a kinase and phosphatase pair, as shown in figure 4.5. The kinase and phosphatase of all three tiers in the cascade has the same kinetic properties, so they are identical. What this figure shows is that when you move down the cascade the response gets more and more sensitive to a change in the signal, the slope of the curves become steeper. The final response curve, i.e. of t_3p, is very close to a step function – resembling a "light switch".

in V_k^* is given by,

$$\delta tp^* = \frac{\partial tp^*}{\partial V_k^*} \delta V_k^* \tag{4.22}$$

The $\frac{\partial tp^*}{\partial V_k^*}$ factor is the slope of the curve of tp^* as function of V_k^*, like figure 4.4D. So, we have been considering the following cascade of phosphorylations,

$$\delta signal \xrightarrow{\frac{\partial V_k}{\partial \text{signal}}} \delta V_k \xrightarrow{\frac{\partial tp}{\partial V_k}} \delta tp \xrightarrow{\frac{\partial V_k^*}{\partial tp}} \delta V_k^* \xrightarrow{\frac{\partial tp^*}{\partial V_k^*}} \delta tp^* \tag{4.23}$$

as shown in figure 4.5. So, the net outcome of the signaling cascade is the following multiplication of sensitivities (slopes),

$$\delta tp^* = \underbrace{\frac{\partial tp^*}{\partial V_k^*} \frac{\partial V_k^*}{\partial tp} \frac{\partial tp}{\partial V_k} \frac{\partial V_k}{\partial \text{signal}}}_{\text{sensivity amplification}} \delta \text{signal} \tag{4.24}$$

The product of the sensitivities that occurs in this equation indicates the sensitivity amplification occurs along a signal transduction cascade (see figure 4.6). Depending on the values of those sensitivities small changes in the signal can bring about either large changes in the output, tp^*, or small changes. This has been postulated to be one of the functions of signaling cascades; in addition, to acting as information integrators – an interesting property that we do not consider here any further.

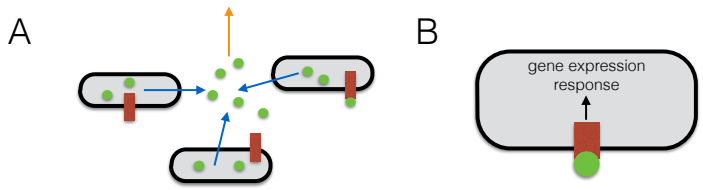


Figure 4.7: **Principle of quorum sensing.** **A.** Growing bacteria produce a chemical compound, the green circle, at a fixed rate, indicated by the blue arrows. The green compound can diffuse away, indicated by the orange arrow, or bind to a dedicated receptor, red rectangle, when the concentration of green compound passes a threshold. **B.** The green compound can bind to a sensor embedded in the bacterial membrane that induces a cellular response such as motility. But binding only occurs to the sensor if the concentration of the green compound is high enough. So at low concentrations of the green compound, the bacteria do not respond.

4.6 Exercises

1. **Dimerisation kinetics of a membrane receptor.** Kasai et al. [10] studied in a very detailed and patient manner the dimerisation equilibrium of a membrane receptor, the N-formyl peptide receptor (FPR), in CHO cells. They found 6000 receptor proteins per cell membrane. The life time of a dimer is 91 ms (ms=milliseconds) and every 150 ms a dimer is formed from two monomers.

   (a) The authors state that the total number of receptor proteins (n_T) in the membrane equals: $n_T = 2 \times n_D + n_M$ with n_D as the number of dimers and the n_M as the number of monomers. Why is this correct?

   (b) The rate of dimer formation equals $v_a = k_a \times m^2$ with m as the monomer concentration. The rate of dimer dissociation equals $v_d = k_d \times m_2$ with the m_2 as the dimer concentration. Give the mass balance for the number of dimers. How is this mass balance related to the mass balance for the number of monomers?

   (c) Even though the dimer is the functional form they only found a small fraction of the receptors in the dimerised form. They found $\frac{\text{number of monomers in dimers}}{\text{number of free monomers}} = \frac{2500}{3500}$. Why do you think that this ratio is so small?

2. **Kinase and phosphatase kinetics and the steepness of the input/output relationship of a kinase-phosphatase pair.** Open the file "kinase and phosphatase.cdf", after having downloaded Wolfram

player (see the instructions on Canvas). Convince yourself that the follow statements are true:

- Increasing the Km of the kinase and the phosphatase makes the input/output relationship, so the steady-state concentration of EP as function of the maximal rate of the kinase V_k, less steep.
- Increasing the total concentration of the target protein makes the input/output relationship steeper.
- The ratio of the total concentration of the target protein over the mean value of the Km's determines the steepness.
- A higher Vmax of the kinase increases the steady-state concentration of EP and a higher value of the Vmax of the phosphatase lowers it.

4.7 Additional exercises (Tougher than exam questions)

1. **How negative feedback in a signaling cascade causes insensitivity to cancer drugs.** In engineering, negative feedback is used to make engineered systems robust against undesired, unpredictable disturbances. Remarkably, it turns out that cells exploit negative feedback for the same purpose. We are considering the famous signaling MAPK-pathway made up out of the following four proteins that sequentially activate each other: $EGFR \rightarrow RAF \rightarrow MEK \rightarrow ERK$. This pathway is overly active in cancer cells and drugs are needed to suppress its activity. In some cells, activated ERK inhibits RAF's activity to activate MEK and in other cells this does not happen. The cells with feedback are remarkably resistant to a MEK-inhibitor whereas the cells without feedback are not. This is a well-known problem in oncology and was recently studied using a systems biology approach by two papers [23, 5]. For the cells with feedback the following response equation for ERK can be derived (the r's denote strengths of protein interactions: so, a 1% change in RAF gives a r^{MEK}_{RAF}% change in MEK),

$$\underbrace{\frac{d\ln ERK}{}}_{\substack{\text{Output}\\ \text{response, } y}} = \frac{\overbrace{r^{ERK}_{MEK} \times r^{MEK}_{RAF}}^{\text{Amplification factor of cascade, } A}}{1 - \underbrace{r^{RAF}_{ERK} \times r^{ERK}_{MEK} \times r^{MEK}_{RAF}}_{\substack{\text{Feedback strength } (-f) \text{ and}\\ \text{cascade amplification: } -fA}}} \times \underbrace{r^{RAF}_{EGFR} \times d\ln EGFR}_{\substack{\text{Change in cascade}\\ \text{activation/input, } u}}$$

$$\Rightarrow \quad y = \frac{A}{1 + f \times A} \times u \qquad \text{(Used in control engineering)} \quad (4.25)$$

So, feedback action in engineered and biological systems are very similar and we can learn from engineers in studying the robustness of biological systems!

(a) Draw the network with and without feedback and indicate the strength of the interactions, the r's, on top of the interaction arrows. Investigate the equation given above once more given this network diagram. The equation should now make more sense.

(b) What is the response equation, so y as function of u, in the absence of the feedback (so when $f = 0$)? What is effect of the feedback?

(c) Why does feedback (when it is strong) make the system insensitive to drug-induced changes in the cascade, A?

(d) Does feedback also protect against any drug-induced changes in u? Hence, do you think that $EGFR$ inhibitors should be more potent than MEK inhibitors?

(e) If you want to understand how the feedback works then rewrite the engineering equation to $y = A \times (u - f \times y)$. So, the feedback substracts the system output from the input. This is indeed how engineers make their control devices. Say, y is the temperature change in a room relative to a desired setpoint, A is related to the capacity of the room to absorb heat, u is the change in outside temperature, and f is the action of the feedback in your thermostat. When does the thermostat work properly? What does this mean for the value of f? So, biological systems even work as thermostats sometimes!

2. **Bacteria communicating with each other and deciding together.** Bacteria are much smarter than we often think. For instance, they use a mechanism called quorum sensing to keep track of the number of bacteria in the population. In this way, bacteria can 'measure' their population size and initiate a response when this number passes a threshold. Some bacteria even use this mechanism to measure the abundance of competing or harmful bacteria. The principle of quorum sensing is shown in figure 4.7. (We note that not all bacteria use sensors to measure quorum-sensing compounds, you can also think of the sensor as a transcription factor inside the bacterium, and the green compound as freely diffusing over membranes.)

   Say that the production rate of green compound per cell equals, k_p, in $\frac{\text{nM}}{\text{minute} \times \text{cell}}$, that we have N cells, and that the diffusion rate equals k_d in $\frac{1}{\text{minute}}$.

   (a) Write down the mass balance of the extracellular concentration of the green compound, denoted by g, and express this equation in terms of k_p, N, and k_d.

(b) Express the steady state concentration of g, called g_s, in terms of k_p, N, and k_d.

(c) Suppose that the bacteria grow exponentially, such that their number increases as $N(t) = N(0) \times e^{\mu \times t}$, with μ as the growth rate and t as time. We also know that when g_s exceeds the threshold concentration g^* that the cells respond with gene expression. Express the time at which the cells initiate a response in terms of k_p, k_d, $N(0)$, and μ. Assume that at every moment in time $g \approx g_s$.

(d) Calculate the time that it takes before the threshold concentration is reached when we start with 1 cell, $k_p = 5\ \frac{\text{nM}}{\text{minute} \times \text{cell}}$, $k_d = 100\ \frac{1}{\text{minute}}$, $\mu = 1\ \text{hour}^{-1}$ and $g^* = 100$ nM.

(e) What is the threshold number of cells?

3. **Does a cell need to be covered completely with sensors for maximal sensing or witg transporters for a maximal uptake rate of nutrients?** We consider a cell with radius a that has N transporters or sensors on its surface, each with radius s. The transporters transport nutrient molecules that find the cell by diffusion whereas the sensors bind molecules and initiate signaling. The nutrient or signal concentration far away from the cell is c_∞. The nutrients and signals diffuse with a diffusion coefficient D. The relation between the uptake flux, J, (or sensor binding rate) and those parameters is (Berg & Purcell, Biophysical Journal, 1977),

$$J = \underbrace{4\pi D c_\infty a}_{J_{max}} \frac{Ns}{Ns + a\pi} \quad \Rightarrow \quad \frac{J}{J_{max}} = \frac{Ns}{Ns + a\pi} \tag{4.26}$$

with J_{max} as the maximal uptake rate. N can be considered as the only variable in this equation. Consider a transporter of 5 nm and the cell of 1 μm in radius. Those are realistic numbers for bacteria.

(a) How many receptor are required for half maximal uptake?

(b) Consider the transporters as disks and the cell as a sphere. What is percentage of area covered by transporters for half-maximal uptake? Write down your expectation first.

(c) Plot the $\frac{J}{J_{max}}$ as function of the cell-surface area that is occupied by transporters.

These were very surprising results and also exact results, no approximations were made in the derivation of the equations mentioned above that are not biologically realistic.

4.8 Key messages of this chapter

1. Signal transduction is the process by which cells sense and integrate environmental cues and integrate this with the cellular state.
2. Signal transduction done by sophisticated regulatory proteins that change in activity by way of covalent modification (e.g. phosphorylation) and conformation changes.
3. Such regulatory proteins form together the cellular signal transduction network that senses and integrates information and induces and monitors the progress of cellular responses, for instance involving changes in gene expression.
4. Mathematical models are very useful in helping you think about the capabilities and limitations of molecular mechanisms of signal transduction.

Chapter 5

Gene regulation

5.1 Introduction

Gene regulation is often considered the most important decision making process by many biologists. I guess this believe is so common because gene regulation acts directly on DNA, and DNA is by many biologists believed to be the central molecule. This is not entirely correct, especially not when one considers bacteria or other autonomously living cells (such as yeasts or cancer cells): such cells can be best perceived as selfish, self-replicating, highly-evolved sensors and adapters; they carefully sense their environment and adapt themselves to maximise their fitness and survival prospects. In this context, the genome acts as a "cookbook" that prescribes how cellular components can be made when they are needed. So, DNA is an equal part of the decision-making and execution machinery of the cell as metabolism and signal transduction are. These three subsystems are highly coupled and none of them is really "in charge" of the others. This is precisely why cells are so complicated to understand: it is a huge network of molecular interactions, thriving only to reach one goal, which is to make copies of itself across as many conditions as possible.

We know now that gene regulation follows very simple principles, at least in the best understood organisms, i.e. bacteria. In higher eukaryotes, such as humans, gene regulation is much more complicated and even though general principles are slowly emerging they are still very sketchy and actively being studied. So, to "play" with the basic molecular interactions and kinetics of gene regulation we limit ourselves to bacterial systems. You do not have to be disappointed; many cool things occur in these systems that are transferable to the more complicated, less-well understood gene regulation systems of mammals.

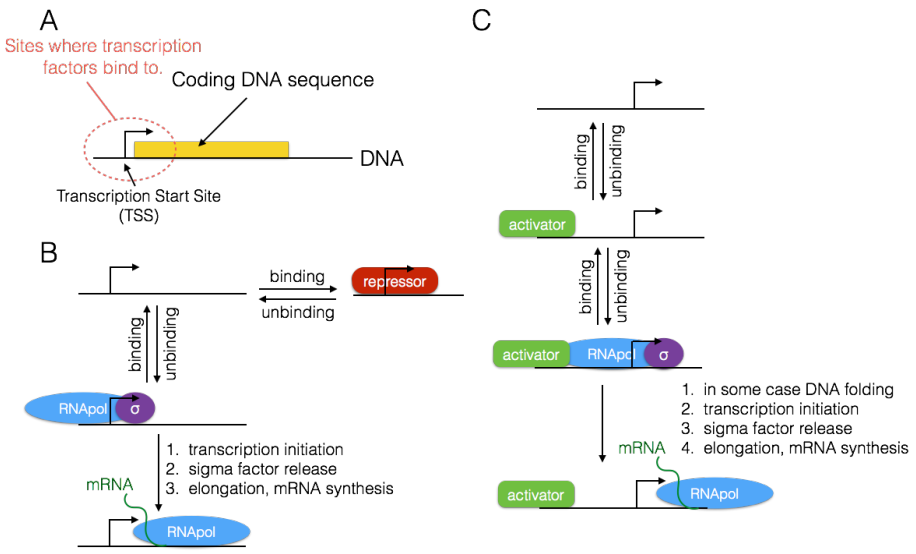


Figure 5.1: **Basic principles of gene regulation by a repressor and an activator. A**. Basic gene structure. **B.** A repressor blocks the binding of the RNA polymerase. When the RNA polymerase binds it is complexed with a sigma factor. Transcription initiation involves open complex formation – opening of DNA followed by RNA polymerase movement, dissociation of the sigma factor, and progression of the elongation RNA polymerase along the gene. **C.** An activator stabilises the binding of RNA polymerase to the DNA and enhances the rate of transcription initiation. In some cases, the activator binds more distant and a DNA loop is formed to facilitate stabilising activator-RNA polymerase interactions.

5.2 Activation and inhibition of a single gene

Gene activity v_g is simply defined as the number of mRNA molecules (deriving from a single gene) produced per unit time by all the gene copies in a cell,

$$\text{gene activity} = \frac{d}{dt}m = v_g \tag{5.1}$$

with m denoting the mRNA concentration; the gene activity (or transcription rate) v_g is proportional to the copy number of the gene encoding the mRNA of interest.

The rate v_g is dependent on many factors: concentrations of RNA polymerase, auxiliary proteins (such as sigma factors), transcription factors, and nucleotides; and on kinetic properties that are determined by the promoter sequence of the gene and the associated proteins. All of this we can describe by

the following equation,

$$v_g = \overbrace{\underbrace{n_g}_{\substack{\text{copy}\\ \text{number}}} \times \underbrace{k_g}_{\substack{\text{rate constant}\\ \text{per copy}}}}^{\text{maximal transcription rate}} \times \underbrace{f_g(\text{concentration of transcription factors})}_{\text{Gene Regulatory Function (GRF); } 0 \leq f_g \leq 1} \quad (5.2)$$

The concentrations of RNA polymerase, sigma factors, and nucleotides we consider fixed, to keep things simple and consider them as part of the transcription rate constant (k_g). This constant varies between $25 - 80$ $\frac{\text{nucleotides}}{\text{second}}$ in *E. coli*. The gene copy number per cell varies with the growth rate and ranges from 1 to 7 from low to high growth rates.

Basic mechanisms of transcription regulation are schematically shown figure 5.1. Transcription factors are proteins that either stimulate or repress transcription initiation. This they achieve by binding to the promoter sequence of the gene of interest and the influence that they have on RNA polymerase activity. Transcription-factor binding sites can be adjacent to the transcription start site where the RNA polymerase binds to or they can be distant. In the latter case, DNA folding is required to achieve direct physical interaction of the transcription factor with RNA polymerase.

In prokaryotes, sigma factors (proteins; *E. coli* has several of them) are required for transcription initiation and they are specific for classes of genes. So, stress response genes can all be associated with the same sigma factors (i.e. σ^H, σ^E, and σ^S), whereas metabolism and growth-related genes are dependent on another sigma factors (i.e. σ^{70} and σ^{54}). When the RNA polymerase/sigma factor complex is bound to the transcription start site, transcription initiation starts when DNA is "melted": the DNA is opened by the RNA polymerase to make the template DNA strength accessible and usable as a template for RNA synthesis. Next, RNA polymerase moves a little into the body of the gene, the sigma factor is released, and the RNA polymerase continues until it reaches the end of the gene where it releases the finished mRNA and falls off the DNA. This process takes about 15 seconds whereas the time between consecutive transcription initiation is a lot more variable and can be minutes.

Several binding sites can occur in the promoter of a gene where several different or identical transcription factors bind. The net outcome is that they determine together the rate of transcription initiation. How they do this is quantitatively described by the gene regulatory function (see equation 5.2).

5.3 Gene regulatory functions describe influence of transcription factors on transcription rate

Gene regulatory functions (GRFs) describe the influence of the transcription factor concentrations and interactions on the promoter region of a gene. They describe the "saturation" degree of the promoter with activating transcription factors: at high transcription factor concentrations all promoters are bound to

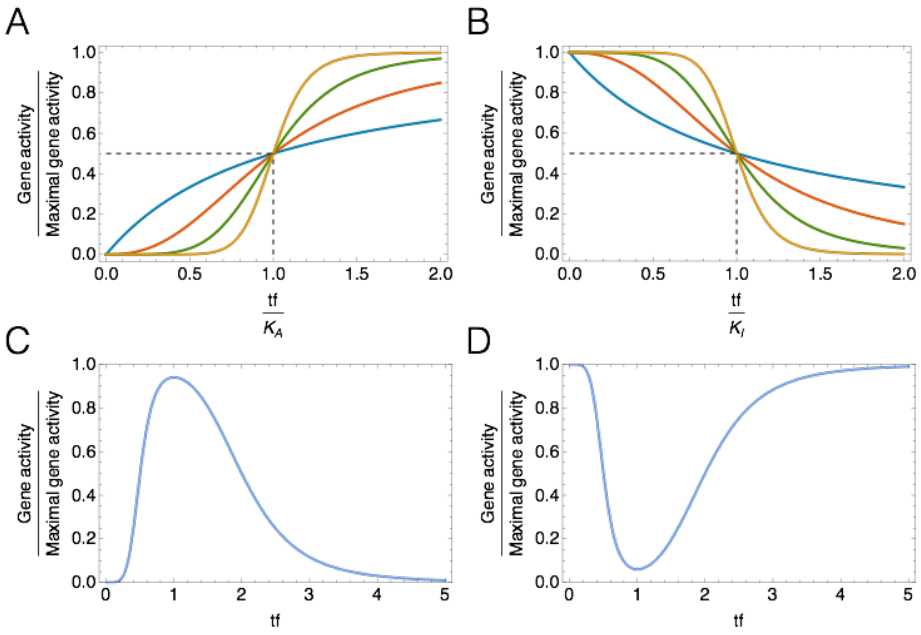


Figure 5.2: **Examples of gene regulatory functions. A.** Examples of GRFs of activating transcription factors. **B.** Examples of GRFs of inhibiting transcription factors **C.** Examples of a GRF for a bandwidth-activating transcription factor. **D.** Examples of a GRF for a bandwidth-inhibiting transcription factor.

transcription factors and the promoter is said to be "saturated" then $f_g = 1$, if 50% of the promoters are occupied then $f_g = 0.5$ and if none of the promoters are bound then $f_g = 0$. If the transcription factor in figure 5.3 is an activator of transcription when two copies of it bind to the promoter then $f_g = \frac{tfptf}{p_T} = \frac{\frac{tf^2}{\alpha \times K_1 \times K_2}}{1+\frac{tf}{K_1}+\frac{tf}{K_2}+\frac{tf^2}{\alpha \times K_1 \times K_2}}$; else if this transcription would be an inhibitor and the empty promoter state would be active then $f_g = \frac{p}{p_T} = \frac{1}{1+\frac{tf}{K_1}+\frac{tf}{K_2}+\frac{tf^2}{\alpha \times K_1 \times K_2}}$.

Figure 5.2 shows various possible GRFs that vary in their sensitivity to the transcription factor and show examples of transcription factors that have an effect in a particular concentration region ("bandwidth").

Figure 5.3 shows a transcription factor binding mechanism where an identical transcription factor binds on two adjacent sites on the promoter. In such a case, an allosteric interaction can occur that either sensitises or desensitises the promoter for transcription factor binding when one of the two sites is occupied.

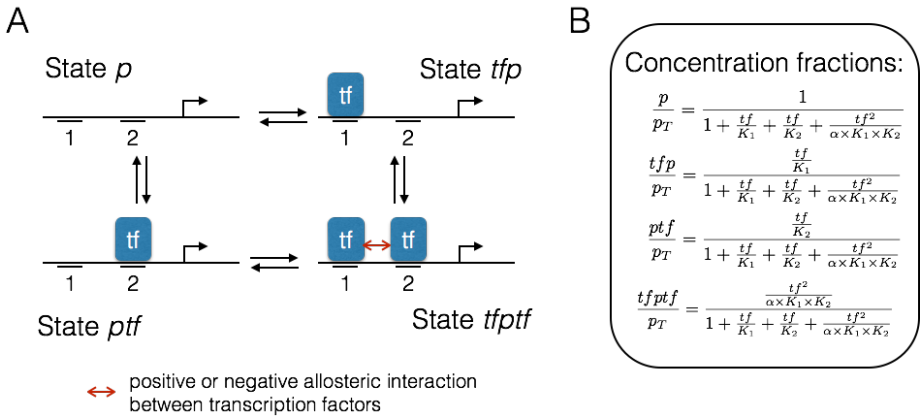


Figure 5.3: **Allosteric interaction between transcription factors that bind to adjacent sites in the promoter region of a gene. A**. Transcription factor (tf) binding and interaction state diagram. **B.** The concentration fractions of the four promoter states. The concentrations of the promoter states are: p, tfp, ptf, and $tfptf$. The total promoter concentrations, p_T, is defined as $p_T = p+tfp+ptf+tfptf$. K_1 and K_2 denote the dissociation constants of the binding sites 1 and 2, respectively. The α factor describes the interaction between the transcription factors: in case of positive allostery $0 \leq \alpha \leq 1$, negative allostery $\alpha > 1$ and if no interaction occurs then $\alpha = 0$.

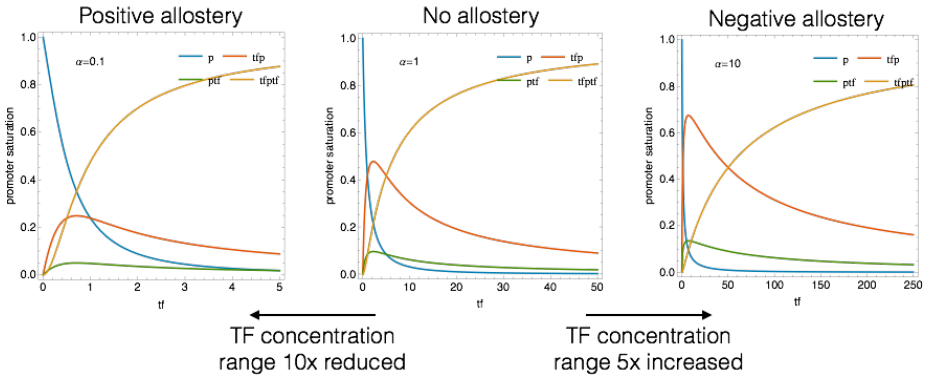


Figure 5.4: **Effect of an allosteric interaction on the transcription factor concentration range.**

5.4 Interpretation of the dissociation constant of a transcription factor for its DNA binding site

It is instructive to take a moment and think about the physical basis of a dissociation constant, the K_1 and K_2 in figure 5.3. They have concentration as unit and are defined by the kinetics of the association and dissociation reaction of the transcription factor to the DNA site,

$$\begin{aligned} TF + P \quad & \underset{dissociation, v_d}{\overset{association, v_a}{\rightleftharpoons}} \quad TFP \\ v_a &= k_a \times p \times tf \\ v_d &= k_d \times tfp \\ \text{At equilibrium: } v_a = v_d \quad & \Rightarrow \tfrac{tf \times p}{tfp} = \tfrac{k_d}{k_a} = K_D \end{aligned} \tag{5.3}$$

The association rate constant can be related to the diffusion coefficient of the transcription factor as,

$$k_a \approx 4 \times \pi \times D_{tf} \times (r_{tf} + r_{site}) \times N_{avogadro} \times 10^{-9} \tag{5.4}$$

When the D is given in dm^2/s and the radii in dm, the unit of k_a is $\frac{1}{nM \times s}$. We know those numbers for *E. coli*,

$$\begin{aligned} k_a \quad \approx \quad & 4 \times \pi \times 5 \frac{(10^{-5} dm)^2}{s} \times 2 \times 5 \times 10^{-8} \frac{\text{dm}}{\text{molecule}} \times 6 \times 10^{23} \frac{\text{molecules}}{\text{mol}} \\ & \times 10^{-9} \frac{\text{mol}}{\text{nmol}} = 0.4 \ \frac{1}{\text{nM} \times \text{s}} \end{aligned} \tag{5.5}$$

A characteristic free concentration of transcription factors is $10 - 30\ nM = 10 - 30\ \frac{\text{molecules}}{\text{cell}}$ in *E. coli*. In order to saturate the promoter we require $\frac{tf}{K_D} \approx 10$. So, a K_D of a 3 nM is a good estimate. Since $k_a = 0.4\ \frac{1}{\text{nM} \times \text{s}}$, k_d has to be $k_d = k_a \times K_d = 0.4\ \frac{1}{\text{nM} \times \text{s}} \times 3\ \text{nM} = 1.2\ \frac{1}{\text{s}}$. The residence time of the transcription factor on the DNA site equals $\frac{1}{k_d} = 0.8\ s$. It then takes about $\frac{1}{k_a \times tf \times p} = \frac{1}{0.4 \times 20 \times 1} = 0.125\ s$ before the promoter is occupied again assuming that $tf = 20\ nM$ and $p = 1\ nM$. This time increases ten fold when $tf = 2\ nM$. At those conditions, $\frac{tf}{K_D} = \frac{20}{3} = 6.7$ and the saturation of the promoter (assuming it has one binding site) equals,

$$\frac{\text{gene activity}}{\text{maximal gene activity}} = \frac{tf}{tf + K_D} \tag{5.6}$$

$$= \frac{20}{20 + 3} = 0.87 \tag{5.7}$$

In figure 5.5 the saturation of the promoter with the transcription factor is shown as function of the free transcription factor concentration.

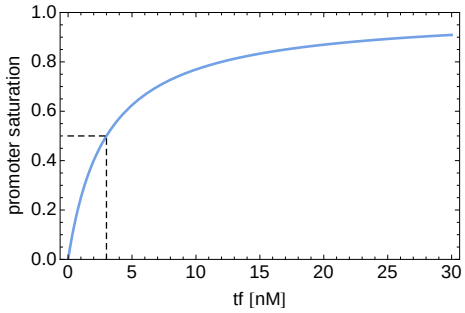


Figure 5.5: **Saturation curve of a promoter with a single transcription factor binding site.** This dependency is described by equation 5.6.

Next, we add transcription initiation to the model,

$$TF + P \underset{k_d=1.2\ \frac{1}{s}}{\overset{k_a=0.4\ \frac{1}{\mathrm{nM}\times s}}{\rightleftharpoons}} TFP \xrightarrow{k_i} \text{elongating RNA polymerase} \tag{5.8}$$

The initiation probability, p_i, when the promoter is in the bound state TFP equals $p_i = \frac{k_i}{k_i+k_d}$. So, it takes on average $\frac{1}{p_i}$ trails before an elongation RNA polymerase appears from a TFP state. In total this takes $\frac{1}{p_i} \times \frac{1}{k_i}$ time and, hence, it is likely that TFP falls apart before an elongation RNA polymerase has appeared. Then we have to wait at least $\frac{1}{k_a \times p \times tf} = 0.125\ s$ before a new TFP state appears. So, the rate of elongating RNA polymerase is a rather complex function. At steady state its value can be derived from the steady state condition. We denote that steady state TFP concentration by tfp_s and we assume that tf is fixed and the total promoter amount equals $p_T = p + tfp$,

$$\begin{aligned} k_a \quad &\times \quad tf \times (p - tfp_s) - k_d \times tfp_s - k_i \times tfp_s = 0 \\ \Rightarrow tfp_s \quad &= \quad \frac{k_a \times p_T \times tf}{k_d + k_i + k_a \times tf} = p_T \frac{tf}{\frac{k_d+k_i}{k_a} + tf} \\ v_{initiation} \quad &= \quad k_i \times tfp_s \end{aligned} \tag{5.9}$$

Note that when we define $K_M = \frac{k_d+k_i}{k_a}$, a Michaelis-Menten relation results for gene expression rate as function of the transcription factor concentration,

$$v_{initiation} = \underbrace{V_{MAX}}_{k_i \times p_T} \times \frac{tf}{K_M + tf} \tag{5.10}$$

This derivation underlies most mathematical models of gene expression regulation. A realistic value for k_i is $0.5\ \frac{1}{\mathrm{min}}$. It is insightful now to see how low the probability for initiation actually becomes!,

$$p_i = \frac{k_i}{k_i + k_d} = \frac{0.5 \frac{1}{\mathrm{min}} \frac{1}{60} \frac{\mathrm{min}}{\mathrm{s}}}{0.5 \frac{1}{\mathrm{min}} \frac{1}{60} \frac{\mathrm{min}}{\mathrm{s}} + 0.8 \frac{1}{\mathrm{s}}} = 0.01 \tag{5.11}$$

We can now also calculate the K_M,

$$K_M = \frac{k_d + k_i}{k_a} = \frac{1.2\ \frac{1}{\text{s}} + 0.5 \frac{1}{\text{min}} \frac{1}{60} \frac{\text{min}}{\text{s}}}{0.4\ \frac{1}{\text{nM} \times \text{s}}} = \frac{1.2 + 0.008}{0.4} \text{nM} = 3.02\ \text{nM} \quad (5.12)$$

If you had problems following this section then it's main insights are summarised in figure 5.6.

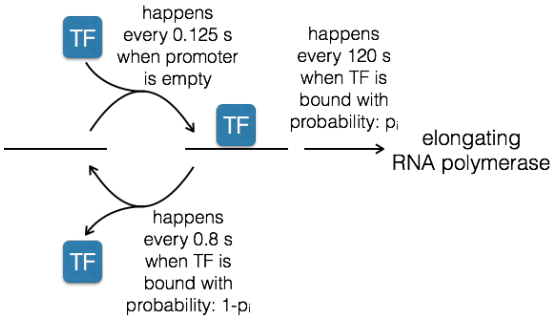


Figure 5.6: **Overview of transcription regulation calculations made in the text.**

5.5 Adding mRNA degradation to the picture

Like any other molecule in the cell, transcripts are also degraded. So, we have to extend the picture of mRNA synthesis with degradation,

$$\frac{d}{dt} m = v_g - k_d \times m = k_i \times p_T \times \frac{tf}{K_M + tf} - k_d \times m \quad (5.13)$$

Here we take $v_g = v_{initiation}$ which is a good approximation for most genes.[1] The steady state mRNA concentration m_s equals,

$$m_s = \frac{k_i}{k_d} \times p_T \times \frac{tf}{K_M + tf} \quad (5.15)$$

This equation indicates that the steady state transcript concentration depends in a hyperbolic manner on the transcription factor concentration. A realistic

[1] Some genes however are so packed with RNA polymerases that the steady state is not determined by k_i but by the delays introduced by RNA polymerases collisions such that $k_g \neq k_i$; then;

$$\frac{1}{k_g} \approx \frac{1}{k_i} + n_c \times \tau_c \quad (5.14)$$

With n_c and τ_d as the number of collisions and the delay time per collision such that their product equals the total delay time. As a result the time between consecutive mRNA production events $\frac{1}{k_g}$ equals the times between consecutive initiations and the time delay introduced by polymerase collisions.

value for $k_d = \frac{1}{10}\frac{1}{\text{min}}$ indicating that every 10 minutes a mRNA is degraded (on average). Taking all the number together we arrive at,

$$m_s = \frac{0.5}{0.1} \times 1 \times \frac{tf}{3.02 + tf} \quad \Rightarrow 0 \leq m_s \leq 5 \; nM \tag{5.16}$$

(Note that 1 $nM = 1 \; \frac{\text{molecule}}{\text{cell}}$.)

Following induction of gene expression, after a sudden increase in the (active) transcription factor concentration, mRNA synthesis follows and after some time a steady state mRNA level is reached. We shall now characterise the "response time", the waiting time for the mRNA concentration to reach 50% of its steady state value. To achieve this, we first have to solve equation 5.13 (assuming that at time zero there is no mRNA),

$$m(t) = m_s \left(1 - e^{-k_d \times t}\right) \tag{5.17}$$

We note that m_s is a function of tf (equation 5.15). The response time $t_{1/2}$ is now defined as,

$$\frac{1}{2} = \left(1 - e^{-k_d \times t_{1/2}}\right) \quad \Rightarrow t_{1/2} = \frac{\ln 2}{k_d} \tag{5.18}$$

So, a mRNA that lives short is quickly degraded and will have a fast response time.

5.6 Autoregulatory gene expression: bistability and memory

Many transcription factors regulate their own transcription rate, either positively or negatively, which is known as "autoregulation". In the case of positive autoregulation, the dynamics of the transcription factor concentration could result from the following mass balance equation,

$$\frac{d}{dt}tf = \underbrace{k_b + k_g \times p_T \times \frac{tf^3}{K^3 + tf^3}}_{v_g} - \underbrace{k_d \times tf}_{v_d} \tag{5.19}$$

This system is studied in figure 5.7. As function of the basal level of transcription activity k_b multiple steady states can occur. This phenomenon is known as "bistability" and occurs surprisingly often in cell biology – it is for instance mechanims that plays an important role in cellular differentiation. Generally, bistability arises in more complicated systems than considered here, making this system a prime example of the basic features of bistability.

Bistability provides a simple form of memory. This becomes clear when you study figure 5.7D. Say, we start at a lower value for the basal rate constant. Suppose we can vary this rate constant experimentally. When we increase the value of this constant and monitor the steady-state value of the transcription factor then we track the lower blue line upto the point where it becomes red.

Then we jump to the higher steady states (upper blue line), as the red steady state are unstable and, therefore, inaccessible. So, we now have arrived at the upper blue line. When we decrease the k_b value we move to the left over the upper blue curve! Thus, we find different values of tf_s for the same value of k_b that we passed earlier! This is a form of memory: if you start at low k_b then you remain at low tf_s values in the region with three steady state, whereas if you start at high k_b values you remain at high tf_s values.

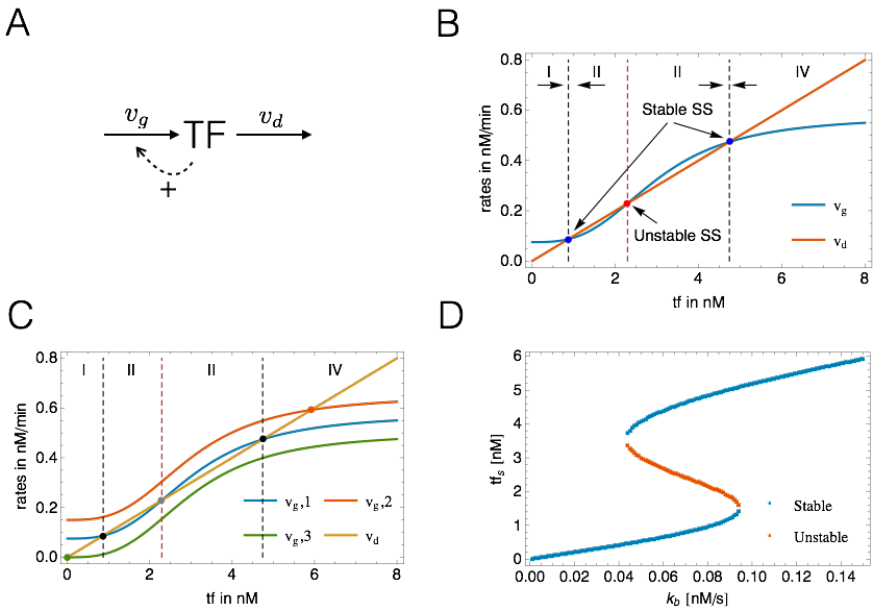


Figure 5.7: **Overview of transcription regulation calculations made in the text. A.** Overview of the network diagram of autoregulatory transcription factor. **B.** Three steady-states can occur in this system, as indicated by the dots. The red dot marks a steady state that cannot be reached by the system as the dynamics always move in the opposite direction as indicated by the arrows. The blue steady state are accessible by the system. The left steady state is reached for concentration of the transcription factor below the red dashed line whereas the right state is reached for concentration higher than the red dashed line. **C.** Whether three steady states or only one occur depends on the value of k_b when it is zero only the green steady state is possible, when it is high the red state is possible, and for intermediate values multiple three steady states are possible. **D.** The admissible steady states are shown as function of the value of k_b. Here we used the realistic parameters derived in the main text.

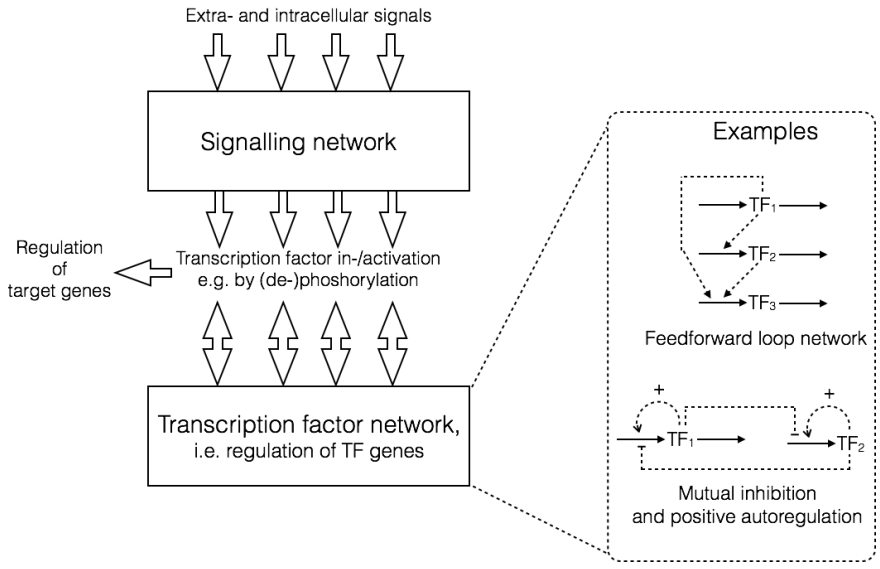


Figure 5.8: **Transcription factor networks.** Activation of transcription factors occurs by signalling networks that respond to extra- and intracellular signals, such as toxins, nutrients, and growth factors. The gene activity of transcription factors is regulated by transcription factors in the gene network, which sets the total concentration of transcription factors. So, the fraction of active transcription factor is set by the signaling network, whereas the total amount is set by the transcription factor network. Finally, the resulting active transcription factor concentration determines the activity of target genes. Two examples of transcription factor networks that occur frequently are shown: i) a feedforward loop network and ii) a mutual inhibition network that is often occurring in gene control of cell differentiation – i.e. the formation of different cell types in multicellular organisms.

5.7 Transcription factor networks: mutual inhibition/activation and feedforward loops

When the transcription rates of transcription factors are controlled by other transcription factors then we speak of a transcription factor network. The concentrations of those transcription factors in such networks are interdependent and the regulation of the target genes of these transcription factors are regulated indirectly by all those transcription factors. Generally those transcription factor are controlled in activity by signaling networks; so, a complicated network of sensing, integration and cellular response results as shown in Figure 5.8.

5.8 Exercises

1. **Induction and repression of gene activity.** When the yeast *Saccharomyces cerevisiae* grows on methionine and sulphate as sulphur sources, it prefers methionine over sulphate. On methionine it reaches a higher growth rate and saves itself the expression of a metabolic pathway that synthesises methionine from sulphate. When methionine runs out, or when the medium is shifted to one without methionine, the methionine biosynthesis pathways needs to be expressed. A central enzyme in this pathway is sulphite reductase. When subsequently a change is made to a medium with excess methionine, the expression of this enzyme needs to be repressed again. We carried out this experiment and determined the number of transcript molecules per cell during those two transitions [18]. The experimental data is shown in Table 5.1 & 5.2.

   (a) Plot the two experimental datasets; for instance, in Microsoft Excel.

   (b) What could explain the overshoot in the induction dynamics?

   (c) Fit the repression dynamics to the theoretical relation $m(t) = m(0) \times e^{-k_d \times t}$ to determine the k_d value, the degradation rate constant. You can do this by hand in Excel, by selecting a value of the k_d such that the theoretical relation overlaps best with the experimental data. What is the unit of k_d? What is the life time of the mRNA? Is the repression dynamics slightly delayed or does it start immediately at time zero? What do you expect and what does the data suggest?

   (d) Fit the experimental data of the induction experiment to the theoretical relation $m(t) = m_s \times \left(1 - e^{-k_d \times t}\right)$ with $m_s = \frac{k_s}{k_d}$ to determine the transcription rate constant. Use the k_d value obtained from the previous exercise. This equation corresponds to the simplest hypothesis about the mechanism:

      - mRNA synthesis is induced at time zero from a synthesis rate of zero to a rate with value k_s,
      - induction starts from 0 mRNA molecules,
      - synthesis and degradation follow the simplest kinetics: $\frac{d}{dt}m = k_s - k_d \times m$

      What is the unit of k_s? Do you think that this model describes the induction mechanism properly? If not, which two features that are apparent in the data should be added to the model to improve the fit?

2. **Dynamics of feedforward loop gene networks [1].** Consider the feedforward loop network in figure 5.8. An interesting feature of such networks is their dynamics. The feedforward loop is the activation of the synthesis

of transcription factor TF_3 by TF_1. Consider the system at steady state at time point 0 when suddenly the synthesis of TF_1 is increased. We start from TF's concentrations that are close to zero. Sketch the dynamics of TF_1, TF_2, and TF_3 – so, their concentrations as function of time – for two different cases. Those cases only differ in the sign of the activation of TF_3 synthesis by TF_2; for one system it is positive and activating whereas for the other system it is negative and inhibiting. In the activating case, the synthesis rate of TF_3 is proportional to the product of the concentration of TF_1 and TF_2; so, then we effectively model an AND relation between these two transcription factors. Which of the two models is expected to show a peak in the concentration of TF_3 as function of time and which one is expected to show a delay in the activation of TF_3 (when you compare the dynamics of TF_1 and TF_3)?

3. **The *lac* operon of *Escherichia coli*.** The *lac* operon of *E. coli* encodes three genes required for catabolism and transport of the sugar molecule lactose. The transcriptional activity of this operon is controlled by two transcription factors: the repressor LacI and the activator CRP. CRP becomes active when it binds to cAMP. The concentration of cAMP is high when the glucose concentration is low. So, glucose suppresses cAMP production. When LacI binds IPTG (a lactose analogue), it no longer binds to the promoter of the *lac* operon. Setty et al. [21] have studied the gene regulatory function of the operon by titration of the concentration of cAMP and IPTG followed by measurement of the transcription activity of the *lac* operon. They fitted an equation to their data, which is plotted in figure 5.9. When is the *lac* operon activated?: a) when lactose is high and glucose is high, b) when lactose is low and glucose is high, c) when lactose is high and glucose is low, or d) when lactose is low and glucose is low? Given you answer, what do you expect what *E. coli* does when it grows on a mixture of glucose and lactose?

4. **Transcription factor binding to DNA** Transcription factors bind to DNA to influence the rate of mRNA synthesis. We consider the following reaction,

$$F + D \underset{\text{dissociation}}{\overset{\text{association}}{\rightleftharpoons}} DF, \tag{5.20}$$

with F as the transcription factor and D as the DNA binding site. Concentrations are in small font and names of molecules in capitol font.

   (a) Why are the total concentrations of F, which is equal to $f_T = f + df$, with subscript 'T' for total, and of D, equal to $d_T = d + df$, constant when the binding reaction occurs in a test tube? Which condition(-s) needs to be met?

(b) If this reaction can be described in terms of mass action kinetics then what would be the rate equation for this reaction and what be the units of all the terms occurring in it?

(c) Consider the situation where we keep the concentration of the transcription factor fixed. Give the mass balances for d and df.

(d) Explain why the following relation holds: $\frac{d}{dt}d = -\frac{d}{dt}df$.

(e) Show that at thermodynamic equilibrium the following relation holds,

$$df = d_T \frac{f}{\frac{k_d}{k_a} + f} \tag{5.21}$$

(f) The ratio $\frac{k_d}{k_a}$ is called the dissociation constant and often written as K_D. What is its unit?

(g) What is the unit of $\frac{f}{K_D+f}$?

(h) Say the volume of a bacterial cell is 1 fl and 15 molecules of f occur in this cell. What is the concentration of f in nM?

(i) If one DNA site exists in this bacterial cell then what would be its concentration in nM?

(j) Plot the occupancy fraction of the DNA binding site, defined as $\frac{df}{d_T}$, as function of f. Set K_D=2 nM. When is $df = \frac{d_T}{2}$?

(k) If f is maximally 15 nM then what is the maximal occupancy fraction?

(l) If the transcription rate equals $v = k_t df$ and k_t equals $5\frac{\text{nM mRNA}}{\text{nM DNA site}\times\text{min}}$ then what is the rate of mRNA synthesis when f equals 0.1 nM, 1 nM, or 10 nM. What is the maximal rate?

(m) If the lifetime of mRNA equals 10 min, which is a realistic number for bacteria, what is the steady state concentration of mRNA when $f = 1$ nM?

5. **Sliding of transcription factors along DNA shortens the promoter search time.** In bacteria, transcription factors find their target DNA sites, from which they regulate transcription, via shortly sliding in 1-D along the DNA, starting from a random DNA site that they encountered after a 3-D diffusive search for DNA in the cytoplasm. If during a single slide the target site is not found the transcription factor falls off and the process starts again. The search time for the target DNA site by a single transcription factor molecule is given by the following equation

$$\tau_s = \left(\frac{V}{4\pi DL} + \frac{l_s^2}{2D_1}\right)\frac{L}{l_s}, \tag{5.22}$$

with V as the cell volume, D as the cytosolic diffusion coefficient, L as the DNA length, l_s as the sliding length and D_1 as the sliding diffusion coefficient. The search time for the target site without sliding equals

$$\tau_{ws} = \frac{V}{4\pi D a}, \tag{5.23}$$

with a as the reaction radius of the target site on the DNA. In this question consider the following parameters: $D = D_1 = 5\ \mu m^2/s$, $a = 5\ nm$, $L = 1.5 \times 10^3\ \mu m$, $l_s = 30\ nm$ and $V = (2\ \mu m)^3$.

(a) What is the length of DNA in mm? Note that this length is a realistic value for *E. coli*). What is the length of an *E. coli* cell? How many times should the DNA be folded to place it in a cytosolic compartment of half the length of an *E. coli* cell? (Can you imagine now how weird it is that during DNA replication the two genome copies have to unwind and separate? This is a longstanding issue in *E. coli* and currently believed to occur spontaneously due to entropic forces.)

(b) Calculate the search time, τ_{ws}, for the DNA target site without sliding.

(c) Calculate the search time, τ_s, for the DNA target site with sliding.

(d) Explain the meaning of $\frac{V}{4\pi DL}$, $\frac{l_s^2}{2D_1}$, and $\frac{L}{l_s}$ in equation 7.78.

(e) When is sliding along DNA by transcription factors advantageous?

Table 5.1: Induction data.

time (min)	$\frac{\text{transcripts}}{\text{cell}}$
0	1.9
10	1.8
20	2.00
30	3.6
40	20.1
50	26.7
60	31.1
70	26.4
80	23.6
90	24.3
100	22.3
120	18.2

Table 5.2: Repression data.

time (min)	$\frac{\text{transcripts}}{\text{cell}}$
0	18.0
5	15.8
10	10.7
20	4.3
30	2.1
45	1.3
60	1.3

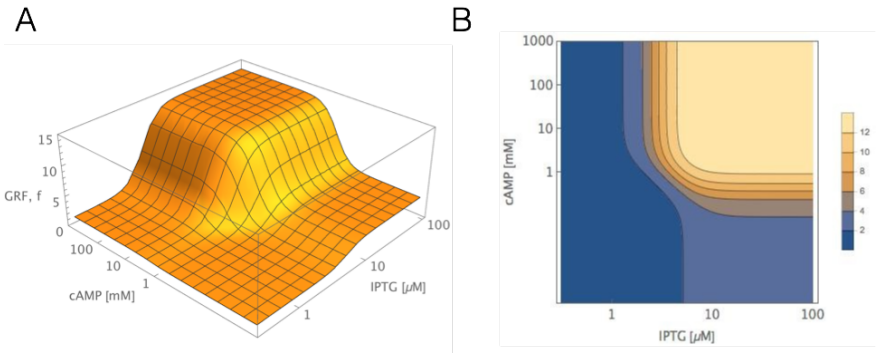


Figure 5.9: **The gene regulatory function of the *lac* operon of *E. coli*.** The transcription activity of the *lac* operon was measured as function of the IPTG and cAMP concentrations. The resulting experimental data was fitted to a function which is plotted in 3D in figure A and in 2D in figure B. So, figure B is figure A when viewed from the top. This data comes from Setty et al. [21].

5.9 Key messages of this chapter

1. Genes are not dictating cellular behaviour as some say but gene activity rather results from environmental sensing and the integration of this information with the cellular state.

2. Individual gene activity is determined by a gene regulatory function that depends on the interplay between transcription factors, which are each dependent on the cellular state.

3. Transcription factors can form networks to give rise to puzzling and astonishing decision behaviour by genetic circuits underlying cellular differ-

entiation, adaptation to new conditions and phenotypic diversification of cellular populations (bistability).

Chapter 6

Metabolism and cell growth

6.1 Introduction

Signal transduction, metabolism and gene expression are all coupled to ensure that cells carry out their function in multicellular organisms or that they grow and withstand stresses in case of autonomous unicellular organisms. In this chapter, we will investigate some of the basic properties of metabolism and growth:

1. what is exponential growth and why is it so important?,
2. why does exponential growth implicate that metabolism operates at a steady state?,
3. short-lived proteins reach their steady state concentration faster at steady-state growth than long-lived proteins,
4. basics of metabolic pathways,
5. how does metabolic regulation work and bring about changes in cellular growth when conditions change?

6.2 The basics of exponential growth of cells

At exponential growth, the number of cells X and the total cell volume V of the population both increase exponentially at a rate μ, called the specific growth rate,

$$\begin{aligned} \frac{d}{dt}X &= \mu \times X \quad \text{with } X \text{ as the number of cells} \\ \frac{d}{dt}V &= \mu \times V \quad \text{with } V \text{ as the total volume of a cell population} \end{aligned} \tag{6.1}$$

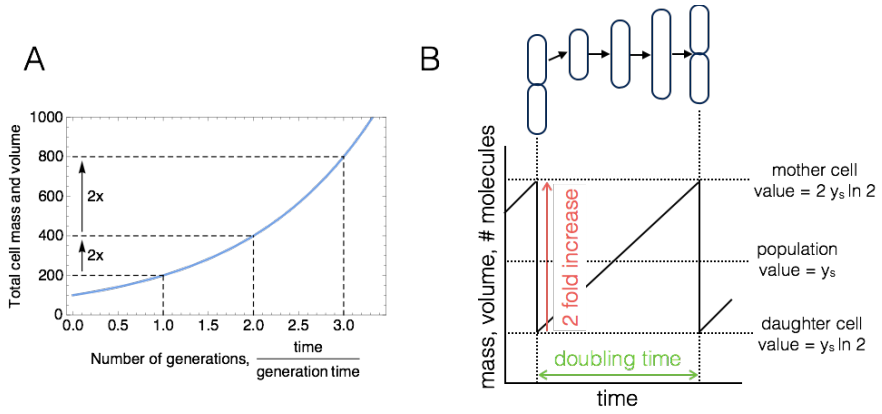


Figure 6.1: **Exponential growth of cells A.** Population perspective: the number of cells, number of molecules and total cell volume all double within one generation time and increase exponentially in time. **B.** Single-cell perspective: every daughter cell doubles its own content and size within one generation time. A daughter has $\ln 2$ times the value measured at the population level and a mother has $2 \ln 2$ times this value.

These equations indicate that the ratio $\frac{X}{V}$ remains fixed because they rise equally fast in time.[1] We will call this ratio χ.

How does the concentration of a protein change with time at exponential growth? The concentration c is defined as the number of molecules divided by the volume. Since, the number of molecules and the cell volume both change, their ratio, the concentration, changes due to both,

$$\begin{aligned} c &= \frac{n}{V} \\ \frac{d}{dt}c &= \frac{d}{dt}\left(n \times \frac{1}{V}\right) = \frac{1}{V}\frac{d}{dt}n - \frac{n}{V^2}\frac{d}{dt}V \end{aligned} \tag{6.2}$$

Here we have used the product rule of differentiation: $\frac{d}{dt}(f(t) \times g(t)) = g \times \frac{d}{dt}f + f \times \frac{d}{dt}g$.

The change in the volume is set by the growth process; exponential growth in this case. The change in the number of molecules is set by the activity of the reactions that this molecule is involved in. We limit ourselves to the simplest case: the protein of interest is synthesised at a constant rate per cell, κ in $\frac{\text{molecules}}{\text{cell} \times \text{time}}$, such that the total number of proteins increases as,

$$\frac{d}{dt}n = \kappa \cdot X. \tag{6.3}$$

[1] You can verify this by using the product rule of differentiation: $\frac{d}{dt}\left(X \times \frac{1}{V}\right) = \frac{1}{V}\frac{d}{dt}X - \frac{X}{V^2}\frac{d}{dt}V = \mu \times \frac{X}{V} - \frac{X}{V^2} \times \mu \times V = 0$! So, the ratio of X and V does not change with time at exponential growth!

Substitution of this equation into equation 6.2 gives,

$$\frac{d}{dt}c = \underbrace{\kappa \times \overbrace{\frac{X}{V}}^{\chi}}_{\substack{\text{synthesis in} \\ \text{concentration units}}} - \underbrace{\mu \times \frac{n}{V}}_{\substack{\text{degradation in} \\ \text{concentration units}}} = k - \mu \times c \tag{6.4}$$

The last equation is an insightful result; it arises from a basic description of exponential growth and it is indeed the simplest equation that one would write for protein synthesis with a constant rate k and the "growth-dilution effect", which lowers the concentration c in time at rate $\mu \times c$. Dilution by growth means that the concentration of a protein decreases due to volume growth of the cell. This concentration reduction needs to be compensated for by protein synthesis to maintain a fixed protein concentration.

Say, a population of cells is growing exponentially and we suddenly induce the synthesis of a protein at a cell state where the concentration of this protein is zero. So, we have,

$$\begin{aligned} \frac{d}{dt}c &= k - \mu \times c \\ c(0) &= 0 \end{aligned} \tag{6.5}$$

We can solve this equation by hand to give rise to[2],

$$c(t) = \frac{k}{\mu}\left(1 - e^{-\mu \times t}\right) = \frac{k}{\mu}\left(1 - \frac{1}{e^{\mu \times t}}\right) \tag{6.6}$$

So, for large times the concentration becomes constant and reaches a steady state value $c_s = \frac{k}{\mu}$ because then $\frac{1}{e^{\mu \times t}} \approx 0$. Therefore, while the cells are growing the concentration of this protein eventually becomes constant! This is the basis of the statement of steady-state metabolism at exponential growth and this is what makes exponential growth the best-defined growth state of cells; simply because it gives rise to the same protein concentration in independent experiments as long as the cells continue to grow exponentially. So, we have

$$c(t) = c_s\left(1 - e^{-\mu \times t}\right) \tag{6.7}$$

To capture how long it takes to reach the steady state protein concentration we determine the half time,

$$\frac{c(t_{1/2})}{c_s} = \frac{1}{2} = 1 - e^{-\mu \times t_{1/2}} \quad \Rightarrow \quad t_{1/2} = \frac{\ln 2}{\mu} \tag{6.8}$$

This is another surprise: $\frac{\ln 2}{\mu}$ is also the generation time of the cells denoted by t_g.[3] The generation time is the time between consecutive cell divisions and

[2] You can test your basic mathematics skills by solving this equation yourself.

[3] Cell growth is given by $\frac{d}{dt}X = \mu \times X$ such that $X(t) = X(0)e^{\mu \times t}$ and the generation time, t_g, is defined as $\frac{X(t_g)}{X(0)} = 2 = e^{\mu \times t_g}$; solving for t_g gives $\frac{\ln 2}{\mu}$.

during this time the cell doubles its volume and divides. So, it appears that several doublings are needed to reach the steady state. How many do we need to reach 95% of the steady-state protein concentration?

$$\begin{aligned} \frac{c(t_{0.95})}{c_s} = 0.95 &= 1 - e^{-\mu \times t_{0.95}} \quad \Rightarrow t_{0.95} = \frac{\ln 0.05}{-\mu} \\ \text{Number of doublings: } \frac{t_{0.95}}{t_{1/2}} &= 4.3 \end{aligned} \tag{6.9}$$

So, after 4.3 doublings the concentration of the protein has reached 95% of its final value. Thus, if *E. coli* doubles every hour you have to wait for about 4.3 hours.

So far, we have considered a protein that lives long, as its concentration decreases only due to dilution by growth. If we contrast this situation with a protein that is more unstable then equation 6.3 changes into,

$$\frac{d}{dt}n = \kappa \times X - k_d \times n \tag{6.10}$$

such that equation 6.4 becomes,

$$\frac{d}{dt}c = \frac{1}{V}(\kappa \times X - k_d \times n) - \mu \times \frac{n}{V} = k - (\mu + k_d) \times c \tag{6.11}$$

Again the equation that one would expect. Solving it gives,

$$c(t) = c_s \left(1 - e^{-(\mu + k_d) \times t}\right) \quad \text{with: } c_s = \frac{k}{\mu + k_d} \tag{6.12}$$

Thus in this case $t_{1/2} = \frac{\ln 2}{\mu + k_d}$, which is a smaller time than the previous case (then: $t_{1/2} = \frac{\ln 2}{\mu}$). So, from this we can conclude that short-lived proteins respond much faster than long-lived proteins. Short-lived proteins are mostly regulatory proteins, such as transcription factors and signaling proteins, whereas long-lived proteins are found most in metabolism and cellular growth. It makes sense that the latter proteins live longer: i) they are most abundant, about a 1000 fold, and their rapid turnover would introduce major costs for the cell causing it to grow slower, and ii) their effects on the cell occur on the time scale of the generation time, so having them only makes sense if they live at least as long as the generation time. So, a protein with a very short life time reaches a steady state earlier than one generation time![4]

[4] In fact you can show that in this case $t_{0.95} = \frac{\ln 0.05}{-(\mu + k_d)} = \frac{\ln 0.05}{-\mu\left(1 + \frac{k_d}{\mu}\right)} = \frac{1}{1 + \frac{k_d}{\mu}} \frac{\ln 0.05}{-\mu} = \frac{\mu}{\mu + k_d} \frac{\ln 0.05}{-\mu}$. Accordingly, the required number of doublings becomes $4.3 \times \frac{\mu}{\mu + k_d}$ and fast degradation shortens the response time. If $\frac{k_d}{\mu} = 3.3$ then 1 doubling is need to reach steady state. So, a protein or mRNA with a life time of 10 minutes (such that $k_d = \frac{1}{10}$ min^{-1}) and a growth rate of $\frac{\ln 2}{t_g} = \frac{\ln 2}{60 \text{ min}} = 0.005$ min^{-1} reaches steady state within $4.3 \times \frac{0.005}{0.005 + 0.1} = 0.2$ doublings which equals 12.3 minutes.

Now we shift focus to a single cell. When it just arose out of a division of its mother cell, its volume and molecule content is half of the volume and content of the mother cell. At the end of the generation time, this cell has doubled its volume and molecule content and will divide itself. Let's look at this in a bit more detail. We know that immediately after division the number of protein molecules equals $n_0 = \frac{n_{t_g}}{2}$ with t_g as the generation time; so, have

$$\begin{aligned} \frac{d}{dt}n &= \kappa \\ n(0) &= n_0 \quad \left(= \frac{1}{2}n(t_g)\right) \end{aligned} \tag{6.13}$$

Solving this gives,

$$n(t) = n_0 + \kappa \times t \tag{6.14}$$

We know that:

$$n(t_g) = 2 \times n_0 = n_0 + \kappa \times t_g \quad \Rightarrow t_g = \frac{n(0)}{\kappa} = \frac{\ln 2}{\mu} \tag{6.15}$$

So, we can conclude:

$$\begin{aligned} n(0) &= \ln 2 \times \frac{\kappa}{\mu} \\ n(t_g) &= 2 \times n(0) = 2 \times \ln 2 \times \frac{\kappa}{\mu} \end{aligned} \tag{6.16}$$

What is meaning of $\frac{\kappa}{\mu}$? It is the mean number of protein molecules, n_s, that you find in a growing population of cells, because

$$c_s = \frac{k}{\mu} = \frac{X}{V}\frac{\kappa}{\mu} = \text{number of cells} \times \frac{\text{steady-state number of molecules per cell}}{\text{total cell volume}} \tag{6.17}$$

So, again using elementary relations and a little mathematics tells you a lot: at exponential growth,

1. the total number of cells, total number of molecules and total cell volume increase exponentially,
2. the protein concentration remain fixed at $c_s = \frac{k}{\mu}$,
3. it takes 4.3 doubling to reach 95% of the protein concentration steady-state when you induce a gene at steady state growth that encodes for a stable protein,
4. short-lived proteins reach their steady state much faster than stable proteins, and
5. after cell division the number of protein per cell equals $\ln 2 \times n_s = 0.69 \times n_s$ and at the time of division the number of molecules is $2 \ln 2 n_s = 1.39 \times n_s$ while the average number of molecules per cell in an exponentially growing population is $n_s = \frac{\kappa}{\mu}$.

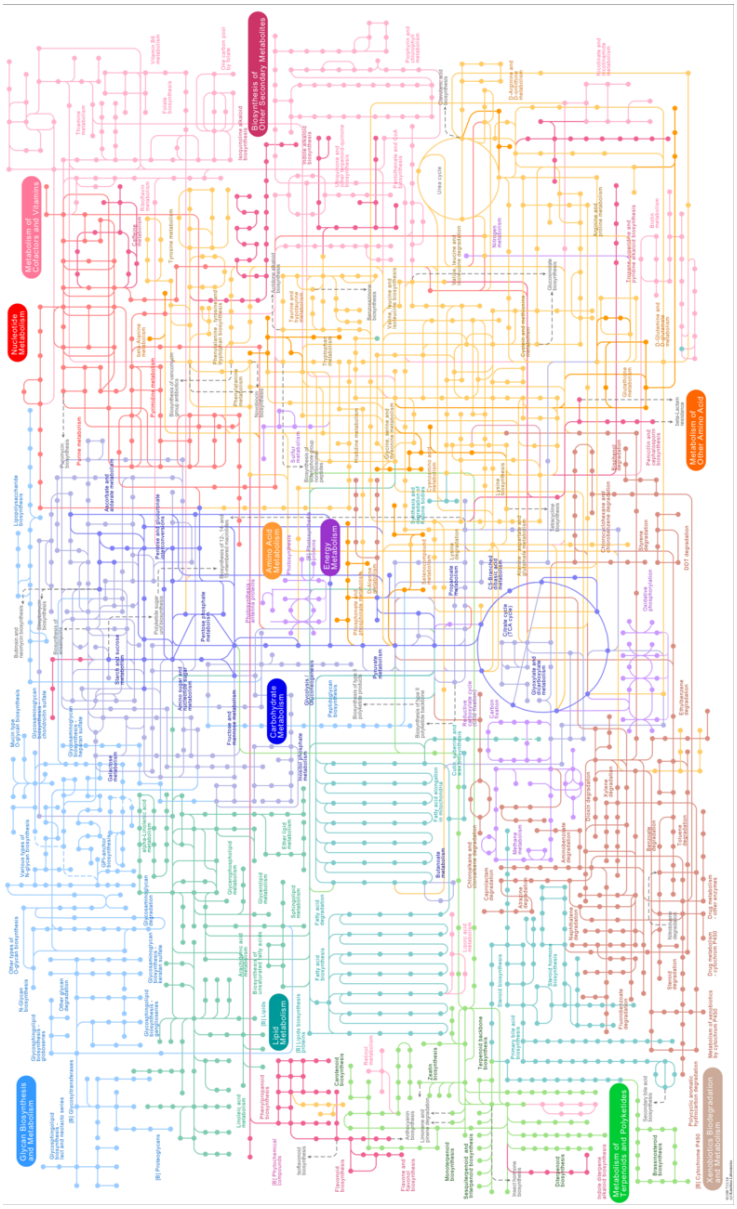


Figure 6.2: **Overview of whole-cell metabolism.** Thousands of enzyme-catalysed reactions are involved in metabolism that together make all the cellular components of a new cell.

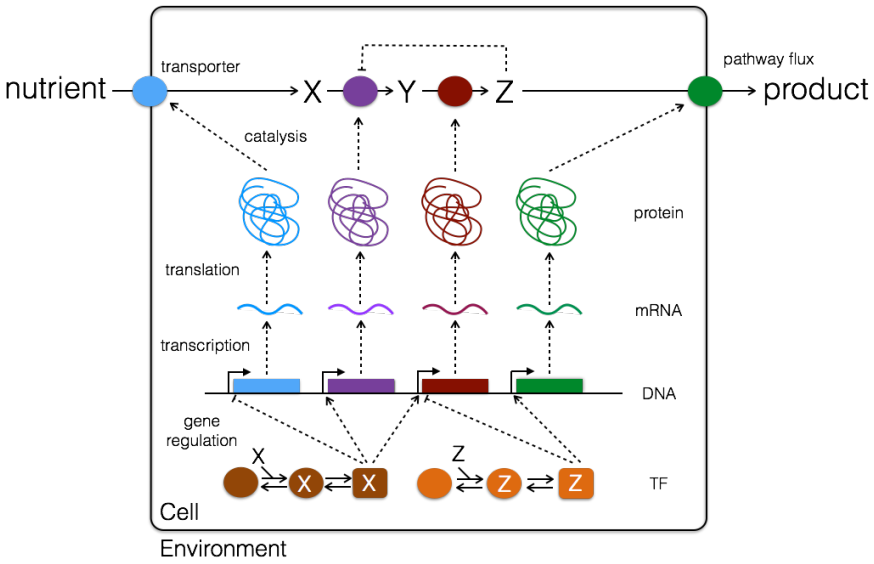


Figure 6.3: **Regulation of metabolism.** Transcription factor activity is dependent on concentrations of intracellular molecules. In this way, the gene expression of metabolic enzyme is adjusted to the demand of the cell. Within metabolism itself, regulatory interactions between metabolites and enzymes give rise to feedback and feedforward circuitry giving rise to fast adjustments of enzyme rates.

6.3 Steady state metabolism

The entire metabolism of a cell is the complete set of reactions (generally catalysed by enzymes) that produce all cellular components from nutrients supplied in the growth medium. At exponential growth, the composition of a daughter cells is doubled within one generation time which requires the uptake of nutrients and their conversion into cellular building blocks, such as DNA, RNA, and protein. The metabolic network that achieves this is composed out of hundreds to thousands of enzymes – depending on the organism and the growth conditions (figure 6.2).

We have already seen in the previous section that all molecules reach a steady state concentration during exponential growth. The time that this takes depends on the life time of these molecules. So, after a change in a nutrient concentration, the perception of this change by the cell and the induction of a response, it may take several generation times before the cell has attained a new steady-state growth state. Changing enzyme concentrations requires the regulation of gene expression activity as shown in figure 6.3.

We will limit ourselves to metabolism at steady state. In this state, the rate of synthesis of all metabolites (the reactants) equals their degradation rates. So,

for instance for glucose-6-phosphate (G6P) in figure 6.4 we have at steady state that,

$$\frac{d}{dt}g6p = v_{HK} - v_{PGI} - v_{PGM1} - v_{TPS1} = 0 \tag{6.18}$$

with the v's as the rates of the reactions catalysed by HK, PGI, PGM1, and TPS1. And for pyruvate (PYR),

$$\frac{d}{dt}pyr = v_{PYK} - v_{PDC} = 0 \tag{6.19}$$

The mass balance equations for all reactants are zero at steady state. This is achieved because all the reactant concentrations reach values such that all balance equations are zero. Such a steady state is often an "attracting state" such that for all initial conditions a steady state is reached after some time of dynamics of reactant concentrations.

When changes in enzyme concentrations occur, due to changes in gene activity or protein degradation, the steady state of metabolism is perturbed and the balances no longer equal zero such that some metabolites concentrations increase whereas others drop. Eventually, when the enzyme concentrations have attained fixed values, the reactant concentration converge to their steady state values and the rates of reactions become fixed. So, metabolism moves in principles from steady state to steady state. Such transitions are induced by environmental changes such as nutrient changes.

6.4 Metabolic pathway expression is regulated by demand

Cells, in particular those that compete with others for nutrients, tend to express only the enzymes that they need for growth. This strategy reduces wasteful usage of resources and makes sure that the enzymes that are expressed are being used. In this manner, cells attain high growth rates and become good competitors when they compete for nutrients. To understand this argument you have to realise the risk of extinction that an organism takes when it expresses the wrong enzymes and grows at a rate that is 1% slower than a competitor. The dynamics of the number of fitter cell, n_f over the less-fit cells, n_l, is most easily found when we realise that,

$$\begin{aligned} n_f(t) &= n_f(0) \times e^{\mu_f \times t} \\ n_l(t) &= n_l(0) \times e^{\mu_l \times t} \\ \frac{n_f(t)}{n_l(t)} &= \frac{n_f(0) \times e^{\mu_f \times t}}{n_l(0) \times e^{\mu_l \times t}} = \frac{n_f(0)}{n_l(0)} \times e^{(\mu_f - \mu_l) \times t} \end{aligned} \tag{6.20}$$

So, the ratio of fitter over less-fit cells increases exponentially in time! This ratio doubles every $\frac{\ln 2}{\Delta\mu}$ time units! So, if the ratio starts at 1, the generation time of the fast cell is 1 hour, and the slow cell grows 1% slower, such that

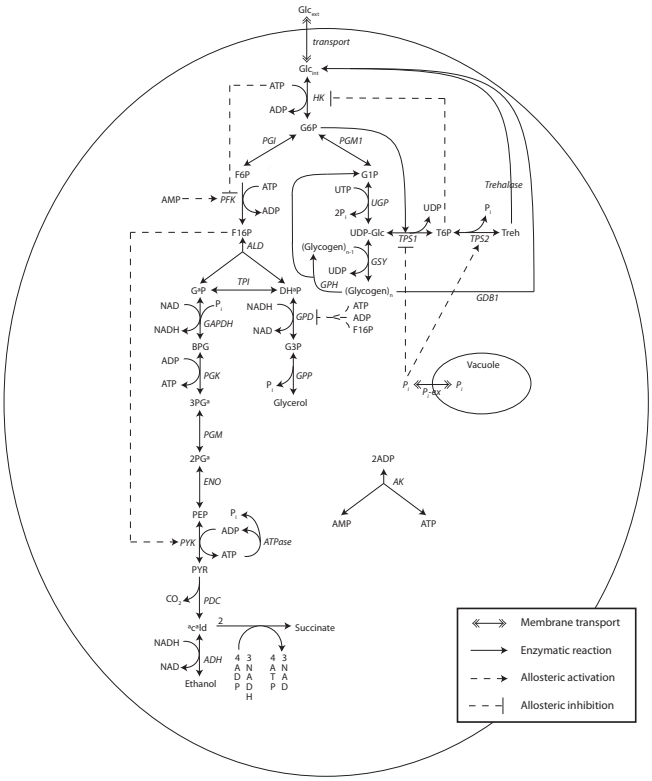


Figure 6.4: **Glycolysis: the best understood metabolic pathway.** Full arrows indicate reaction catalysed by enzymes. So the second reaction from the top is catalysed by hexokinase (HK): $Glc + ATP \rightleftharpoons ADP + G6P$ with Glc as glucose, and G6P as glucose-6-phosphate. The activity of hexokinase is inhibited by trehalose-6-phosphate as indicated by the dashed arrow. Glycolysis is the pathway drawn at the left whereas the reactions at the right – emanating from $G1P$ are involved in the synthesis of glycogen and trehalose. Metabolites are the substrates and products of reactions.

$\Delta\mu = 0.001$ min^{-1}, then after $10 \times \frac{\ln 2}{0.001 \text{ min}} = 6600$ min $= 4.6$ days, the ratio has grown to $2^{10} = 1024$. We started at 50% of each phenotype and after 4.6 days the less-fit cell makes up about 0.1% of the population. Thus, expressing proteins in such a way that the highest growth rate is achieved makes a lot of sense, because competition at exponential growth rate makes you loose fast! Therefore, cells should sense their environment actively, express metabolic pathways only when needed, and tune enzyme levels to prevent overexpression. This we call "metabolic pathway expression by demand".

In a perfect world, cells achieve optimal states but they may not do so in reality; because of all kinds of reasons, but one can easily imagine that optimal

states are hard to reach by evolutionary changes and that trade-offs occur. A trade-off means that if an organism is good at carrying out one task it is the same time bad at another. If trade-offs are strong then they provide a driving force for speciation in evolution.

A well-known example of metabolic pathway expression by demand is so-called diauxic growth; first studied in detail by Jacques Monod, François Jacob, and André Wolff which got them the Nobel Prize for Physiology and Medicine in 1965. Diauxic growth means sequential growth where one "preferred nutrient" is consumed prior to the second. This requires a repression of the preferred nutrient metabolism when the metabolism of the second nutrient is induced. This takes time and the growth of an organism therefore can display a lag phase. Figure 6.5 shows Monod's 1965 visualisation of diauxic growth with varying durations of lag phase in his Nobel lecture (available from: `http://www.nobelprize.org/nobel_prizes/medicine/laureates/1965/monod-lecture.pdf`). The mechanism of repression is quite well understood in molecular terms for the glucose-lactose diauxy of *E. coli*. The activity of the genes encoding the enzymes for lactose metabolism and transport are all part of the so-called *lac* operon. We have earlier studied this system (figure 5.9) and concluded that this operon is active when the lactose concentration is high and the glucose concentration is low. So, at high glucose and high lactose concentration, this operon is off. Thus, *E. coli* prefers glucose over lactose and will first consume glucose and subsequently lactose when it grows in the presence of both.

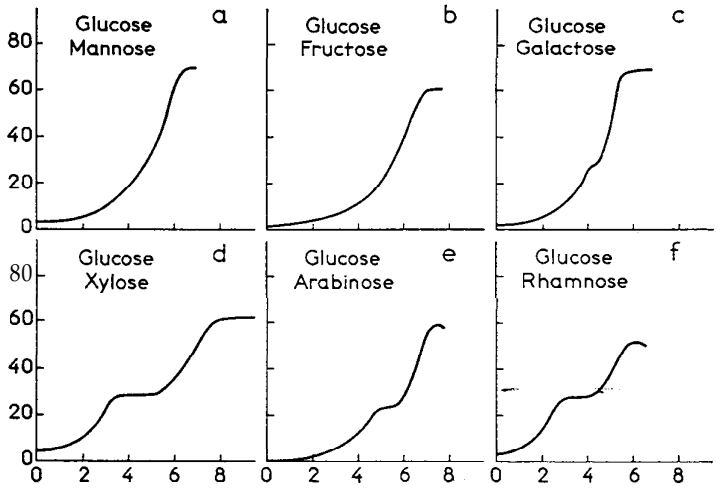


Fig.1. Growth of *Esherichia coli* in the presence of different carbohydrate pairs serving as the only source of carbon in a synthetic medium[50].

Figure 6.5: **Diauxic growth figure from Monod's Nobel prize lecture in 1965.**

6.5 Regulation of metabolic enzyme activity

In addition to gene regulation of the activity of metabolic enzymes, which gives rise to changes in their concentration, metabolic enzyme activity is regulated by the concentrations of reactants and effectors. Effectors are molecules that are not converted by the enzymes that they regulate. So, AMP is an effector of PFK in figure 6.4. Taken together, enzyme activity, v, is effected by metabolic and gene regulation,

$$v = V_{max} \times \frac{s}{K_M + s} = k_{cat} \times e \times \frac{s}{K_M + s} \tag{6.21}$$

The total enzyme concentration, e_T, is regulated by gene expression. This total concentration is either the same or higher than the concentration of enzyme that carries out catalysis of the reaction, e. The ratio of $\frac{e}{e_T}$ can be controlled by effectors. Effector can also influence the K_M of the enzyme – so, the affinity ($\frac{1}{K_M}$) of the enzyme for its substrate – or the k_{cat}.

6.6 Which metabolic enzyme should be regulated?

Now that we know that effectors and gene expression can both influence the rate of metabolic enzymes, we need to better understand what characterises an enzyme that should be regulated by either of these two mechanisms. This is a relevant question because we know that not all metabolic enzymes are regulated by effectors. Why are only some metabolic enzymes subject to effector regulation and others not? Generally, the function of a metabolic pathway is to generate its product at a certain rate. Regulation occurs to tune this rate with respect to the environment or other intercellular processes; so, to tune pathway activity to its demand. We limit ourselves to steady state again.

So, the question we ask is: "Which enzyme in the metabolic pathway is the best target for a regulatory interaction with an effector molecule to bring about a desired change in steady-state pathway flux?". For convenience, we denote the steady-state flux by J. Clearly, the regulated enzyme should have an influence on the flux when its rate is changed by the effector molecule. If this does not hold in general for all enzymes in the pathway then we can already exclude these enzymes. Several simple situations can occur:

1. **When the first enzyme activity is not inhibited by the concentration of its product.** In general, enzymes are inhibited by their product concentration; for instance,

$$\begin{aligned} \text{Reaction:} \quad & S \rightleftharpoons P \\ \text{Enzyme rate equation:} \quad & v = V_{max} \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \end{aligned} \tag{6.22}$$

the rate v of conversion of S into P decreases with the product concentration. However, if this is not the case for the first enzyme in a pathway then this enzyme sets the flux through the pathway. Then this enzyme acts like a pump and continuesly "pumps" product into the pathway. So, a change in flux can only be achieved by an activity change of the first enzyme in the pathway.

2. **When one enzyme in the pathway has a much lower maximal rate.** An enzyme cannot catalyse a reaction faster than its maximal rate, its V_{max}. If one of the enzymes has a much lower V_{max} than all others and the pathway flux is nearly equal to this V_{max}-value then this enzyme has the largest influence on the flux when its activity is regulated. Also this can be intuitively understood. This enzyme works at maximal capacity and a decrease in this V_{max} will lead to an almost proportional change in the flux (J); since, $V_{max} \approx J$.

3. **Enzymes that are in great excess such that $V_{max} >> J$ have no influence on the flux.** This follows immediately from the previous argument.

6.7 Quantification of metabolic enzyme susceptibilty for regulation

From the last section it has become clear that enzymes that are suitable as regulatory targets should have an influence on the flux when their rate is changed by a regulator; so, a change in a regulator concentration of this enzyme should bring about a change in the flux,

$$\Delta[\text{regulator}] \xrightarrow{1} \Delta \text{ enzyme activity} \xrightarrow{2} \Delta J \tag{6.23}$$

When $\frac{\Delta J}{\Delta[\text{regulator}]}$ is high then this regulator is very potent and the enzyme affected by the regulator is a potent regulatory target. The regulator can be the enzyme concentration – in case of gene expression regulator – or a concentration of an effector – e.g. in case of regulation by metabolic feedback.

Let's start with the enzyme case (gene expression regulation): so, the influence Δ enzyme activity $\xrightarrow{2} \Delta J$. An enzyme is very susceptible to regulation if the value $\frac{\Delta J}{\Delta[\text{enzyme}]}$ for this enzyme is large. This is often quantified by a so-called flux control coefficient [7][5] – defined for (differentially) small change in enzyme concentration;

$$\frac{dJ_k}{J_k} = \frac{e_i}{J_k}\frac{\partial J_k}{\partial e_i}\frac{de_i}{e_i} \quad \Rightarrow \quad d\ln J_k = \underbrace{\frac{\partial \ln J_k}{\partial \ln e_i}}_{\substack{\text{flux control} \\ \text{coefficient, } C_i^{J_k}}} d\ln e_i \tag{6.24}$$

[5] This is a great paper and definitely worth reading! An updated version of this original work was published in 1995 [8].

Different scenarios for the control of flux by enzymes are shown in figure 6.6.

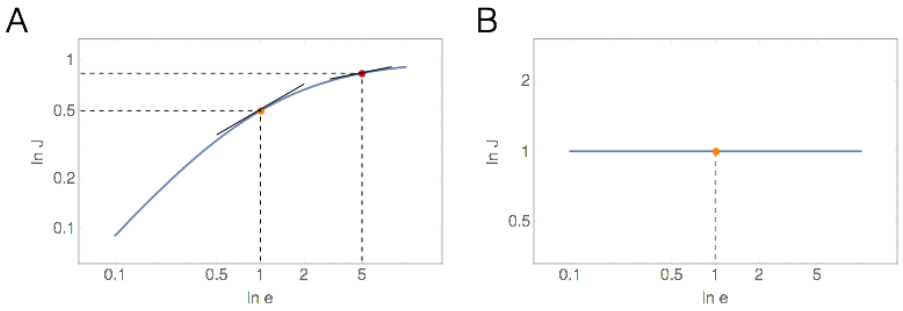


Figure 6.6: **Changes in enzyme concentration can have variable influences on metabolic pathway flux. A.** The sensitivity of the flux through a metabolic pathway J can depend on the concentration of the enzyme; rule of thumb: when an enzyme is increased in concentration its influence on the flux decreases. **B.** Some enzymes have no effect on the flux at all.

The influence of a regulator on the rate of the enzyme Δ[regulator] $\xrightarrow{1}$ Δ enzyme activity is quantified by a so-called elasticity coefficient of the enzyme (with r =[regulator]) [7],

$$\frac{dv_i}{v_i} = \frac{r}{v_i}\frac{\partial v_i}{\partial r}\frac{dr}{r} \quad \Rightarrow \quad d\ln v_i = \underbrace{\frac{\partial \ln v_i}{\partial \ln r}}_{\substack{\text{elasticity} \\ \text{coefficient, } \epsilon_r^{v_i}}} d\ln r \tag{6.25}$$

To determine this coefficient we make the same curves as shown in figure 6.6 but then with $\ln v$ on the y-axis and $\ln r$ in the x-axis and determine the slope.

Finally, an enzyme is susceptible to regulation if both C_i^J and $\epsilon_r^{v_i}$ are high; in fact, the flux response to a regulator is,

$$d\ln J = C_i^J \epsilon_r^{v_i} d\ln r \quad \Rightarrow \quad \text{multiplication of two slopes} \tag{6.26}$$

In other words, we found that:

$$\underbrace{d\ln[\text{regulator}]}_{\substack{\text{fractional change} \\ \text{in regulator} \\ \text{concentration}}} \xrightarrow{\epsilon_r^v} \underbrace{d\ln \text{enzyme activity}}_{\substack{\text{fractional change} \\ \text{in enzyme} \\ \text{activity}}} \xrightarrow{C_v^J} \underbrace{d\ln J}_{\substack{\text{fractional change} \\ \text{in flux}}} \tag{6.27}$$

This causal chain of events in the quantification of the qualitative description in equation 6.23. Enzyme properties – "biochemistry" – sets the first sensitivity coefficient ϵ_r^v whereas the second sensitivity coefficient C_v^J is set by the *entire* metabolic network; so, the kinetic properties of all the enzyme in the network! Other interesting aspects about the functioning of metabolic systems can be found in Hofmeyr & Cornish-Bowden [6].

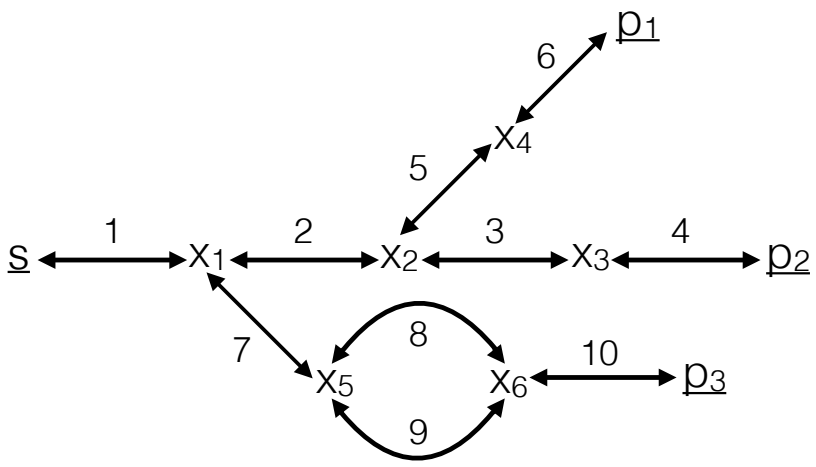


Figure 6.7: **Example of a branched metabolic network.** Arrows denote enzyme catalysed reactions, underlined metabolites are held fixed to allow for a steady state, and all reactions are reversible. Reaction rates are positive when mass flows from left to right.

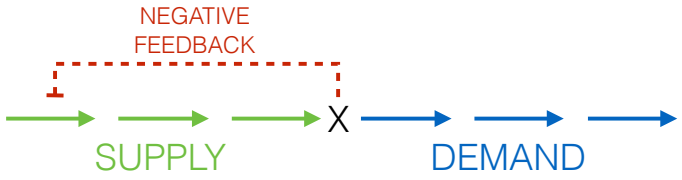


Figure 6.8: **Supply-demand model of metabolism with a negative feedback loop.** The green reactions preceding the feedback metabolite x together from the supply system and the demand system comprises the reactions following x. The concentration of x inhibits a reaction in the supply system.

6.8 Exercises

1. **Synthesis of cellular components by metabolism and adaptations of metabolism upon a nutrient change.** Metabolic networks are composed out of hundreds to thousands of coupled reactions that together synthesise cellular components from extracellular nutrients. Metabolic map

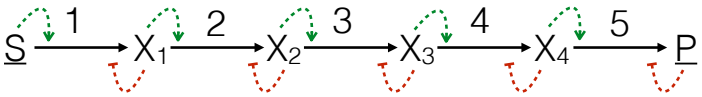


Figure 6.9: **Principles of a metabolic pathway.** Enzyme biochemistry dictates that the rate of every metabolic enzyme is generally activated by its substrate and inhibited by its product. The rate of the enzyme therefore increases when the substrate concentration rises and decreases when the product concentration rises.

visualisation are a very human-friendly tool to get an impression of the metabolic activity of an entire cell. Maarleveld et al. [17] have published a framework for visualisation of metabolic networks together with experimental data, such as transcript levels and rates of metabolic reactions, and computational results. The map was made for the metabolism of the cyanobacterium *Synechocystis spp. PCC 6803* used for biofuel production in biotechnology. We will the map of this microorganism to study a few properties of metabolic networks.

(a) First use Wikipedia to learn a bit more about cyanobacteria. How do they acquire energy? Why does their metabolism resemble the metabolism of plants? How do they survive during the night?

(b) Study the metabolic map of *Synechocystis spp. PCC 6803* when it is growing on carbon dioxide (CO_2) by downloading it from `http://bruggemanlab.nl/?page_id=229` and opening it in a pdf-reader. Find the biomass reaction on the map; this the reaction that makes biomass and has on the right hand side of the reaction arrow: "Biomass specific growth rate" and on the left hand side of this arrow the cellular components. The value above the arrow is the growth rate of *Synechocystis spp. PCC 6803* in units hr^{-1}. The synthesis fluxes of the cellular components are the numbers above their synthesis arrows and have as unit $\frac{\mathrm{mmol}}{\mathrm{gram\ cells} \times \mathrm{hr}}$.

i. Calculate the generation time of the bacterium.

ii. How much protein, DNA, RNA, and lipid in mmol per gram biomass per generation time needs to be synthesised by this bacterium?

iii. Identify the molecular precursors of DNA and RNA by scanning the metabolic network for reactions that have DNA or RNA as product and writing down the substrates – which are the molecular precursors – of those reactions. What is the main difference between the RNA and DNA precursors?

iv. Use the metabolic map to figure out what the main molecular components of lipids are.

(c) Study the metabolic map of *Synechocystis spp. PCC 6803* when it is growing on glycogen by downloading it from `http://bruggemanlab.nl/?page_id=229` and opening it in a pdf-reader. Does the organism grow faster all slower on glycogen than on carbon dioxide? Which parts of metabolism change in activity when this organism shift from carbon dioxide to glycogen growth?

2. **Measurement of the promoter activity of genes during steady-state cell growth (Advanced question).** Keren et al. [11] have measured the promoter activity of thousands of genes in *E. coli* and *S. cerevisiae* during steady-state exponential growth at different nutrient sources. Although this is not immediately clear form their text, they determined the rate constant $k = \kappa \times \frac{X}{V}$ as defined in equation 6.4 upto a factor. So they determined $k \times \alpha$ with α as a constant, its meaning because clear below. We know that at steady state: $k = \mu \times c_s$. Keren et al. [11] measured the growth rate μ and the fluorescence "concentration" $\alpha \times c_s = \frac{\text{fluorescence}}{\text{OD}}$ (with OD as the optical density and proportional to total cell volume); the following argument shows that fluorescence per OD is proportional to a protein concentration,

$$\begin{aligned} \frac{\text{total fluorescence (f)}}{\text{OD}} &= \frac{\frac{\text{fluorescence}}{\text{cell}} \times \text{number of cells}}{\frac{\text{volume}}{\text{cell}} \times \text{number of cells}} \\ &= \frac{\frac{\text{fluorescence}}{\text{protein}} \times \frac{\text{protein}}{\text{cell}}}{\frac{\text{volume}}{\text{cell}}} \\ &= \underbrace{\frac{\text{fluorescence}}{\text{protein}}}_{\alpha} \times \underbrace{\frac{\text{protein}}{\text{volume}}}_{c} \end{aligned} \tag{6.28}$$

(a) Show that the equation used by Keren et al. [11] (shown in Figure 1),

$$\text{promoter activity} = \rho = \frac{f(t_2) - f(t_1)}{\int_{t_1}^{t_2} OD(t)dt} \tag{6.29}$$

indeed equals $\alpha \times c_s \times \mu$ when the cells grow exponentially. So, they really measured the promoter activity as $\alpha \times k$.

(b) They compared the promoter activity of each gene at two different growth conditions, say growth on glucose in mineral medium (condition A) versus growth in the same medium supplemented with amino acids (condition B) (Figure 2C in [11]). They found that two classes of promoter activity differences can be distinguished: i) class 1: gene activity giving rise to the same fluorescence concentration at steady

state at the two different conditions despite a different growth rate and ii) class 2: where the fluorescence concentration and the growth rate are both different between the two conditions. Show that all the class 1 promoter activities fall on the same line in a plot of $\ln k_A$ (so, promoter activity at condition A) as function of $\ln k_B$ and that this line is described by:

$$\ln k_B = \underbrace{1}_{\substack{\text{line}\\ \text{slope}}} \times \ln k_A - \underbrace{(\ln \mu_A - \ln \mu_B)}_{\Delta \ln \mu, \text{ intercept of the line}} . \tag{6.30}$$

(c) Why do class 2 genes not fall on this line?

(d) How would you characterise class 1 and class 2? So, what is their main difference?

3. **Metabolism and growth responses of single cells exposed to nutrient changes.** Read the abstract of the following five papers:

   i) A Solopova, J van Gestel, F J Weissing, H Bachmann, B Teusink, J Kok, and O P Kuipers. Bet-hedging during bacterial diauxic shift. *Proc Natl Acad Sci U S A*, 111(20):7427–7432, May 2014

   ii) O Kotte, B Volkmer, J L Radzikowski, and M Heinemann. Phenotypic bistability in escherichia coli's central carbon metabolism. *Mol Syst Biol*, 10(7):736–736, 2014

   iii) J H van Heerden, M T Wortel, F J Bruggeman, J J Heijnen, Y J Bollen, R Planqué, J Hulshof, T G O'Toole, S A Wahl, and B Teusink. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114–1245114, Feb 2014

   iv) P J Choi, L Cai, K Frieda, and X S Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446, Oct 2008.

   v) A Schwabe and F J Bruggeman. Single yeast cells vary in transcription activity not in delay time after a metabolic shift. *Nat Commun*, 5:4798–4798, 2014

   What do these papers tell you about the responses of single cells to nutrient changes? Do all cells respond the same or are single cell responses (unexpectedly) variable between individual cells? What do you think can be origin of the differences between single cells responses to environmental changes?

4. **What sets the maximal growth rate of a bacterium?** During steady-state exponential growth – 'balanced growth' – the total number of cells, N, increases exponentially in time as $N(t) = N(0)e^{\mu t}$ with μ as the specific growth rate in hr^{-1}. The total culture volume and total cell mass

increase exponentially as well. Since metabolism is operating at steady state during balanced growth, the concentrations of all molecular species remain fixed over time such that the total number of molecules of every molecular species increases equally fast as the total volume. If we focus on the concentration of ribosome then we have the following balance equation at balanced growth, with f_{ribo} as the ribosome synthesis rate per ribosome and ϕ_{ribo} as the fraction of ribosome that is synthesising ribosomes (i.e. $\phi_{ribo} = \frac{\text{number of ribosomes synthesising ribosomes}}{\text{number of ribosomes synthesising protein}}$),

$$\begin{aligned} \frac{d}{dt}c_{ribo} &= v_{synthesis} - \mu c_{ribo} = 0 && (6.31)\\ &= f_{ribo}\phi_{ribo}c_{ribo} - \mu c_{ribo} && (6.32)\\ \Rightarrow \mu &= f_{ribo}\phi_{ribo} && (6.33) \end{aligned}$$

The last equation is a definition of the growth rate in terms of biochemistry.

(a) Calculate f_{ribo} in ribosomes per hour, given that a single ribosome contains 7459 amino acids and that the translation rate is 20 amino acids per ribosome per second.

(b) Calculate the maximal growth rate.

(c) How many doublings do you have per hour with this maximal growth rate?

(d) Calculate the maximal doubling rate in $\frac{doublings}{hr}$.

(e) The experimentally-determined minimal doubling time (= minimal generation time) that the bacterium *Escherichia coli* can attain equals 20 $\frac{min}{doubling}$. Calculate the fraction of ribosome synthesising ribosomes at this growth rate.

(f) In reality, ϕ_{ribo} will always be smaller than 1 because the cell needs always other proteins than ribosomes. Examples of the other proteins are metabolic proteins that make amino acids out of nutrients, in order to allow ribosomes making proteins. Therefore if $\phi_{ribo} < 1$ other proteins are made in addition to ribosomes, such that ribosomal protein fraction of $\Phi_{ribo} = \frac{[ribosomes]}{[proteins]}$ is smaller than 1. Show that the fraction of ribosomes making ribosomes equals the ribosomal protein fraction: $\phi_{ribo} = \Phi_{ribo}$.

(g) Relate the ribosomal protein fraction to the growth rate.

(h) In reality some percentage of the ribosomes is always inactive – they are 'maturating' – such that the total ribosome concentration equal the sum of active and inactive ribosome: $c_{ribo} = c^A_{ribo} + c^I_{ribo}$ and therefore $\mu = f_{ribo}\frac{c^A_{ribo}}{c^A_{ribo}+c^I_{ribo}+c_m} = f_{ribo}\frac{c_{ribo}-c^I_{ribo}}{c^A_{ribo}+c^I_{ribo}+c_m} = f_r(\Phi_{ribo} -$

Φ^I_{ribo}). When we use experimental data for *E. coli* to study the relationship between the ribosomal protein fraction (y-axis) as function of the growth rate (x-axis), we find a linear relation (Figure 1A, in Scott et al, Science, 2010). Use the $\mu = f_{ribo}(\Phi_{ribo} - \Phi^I_{ribo})$ to identify what the slope and intercept of this linear relation is.

(i) How does the previous relation change when we inhibit translation using a translation inhibitor (chloramphenicol) that only affect f_{ribo} (Figure 1B, in Scott et al, Science, 2010)?

5. **Mass flow through a steady state metabolic network.** Metabolic networks are responsible for the synthesis of energy and precursor molecules for the construction of cellular macromolecules such as proteins, DNA, RNA and membranes. Metabolic networks are highly branched and can contain upto thousands of reactions. Navigating through those networks is therefore not straightforward and computational tools are often used to simplify this process. Since mass balance depend linearly on reaction rates, we can use linear algebra to study these huge networks. This linearity is reflected in the mass balance equations that depend on the reaction rates as a linear function: $\frac{dx}{dt} = \sum_{i=1}^{r} n_i v_i$, with the n_i as a (positive or negative) fixed stoichiometry coefficient, v_i as a concentrations-depend reaction rate and r as the number of the reactions in the network. The relation would be nonlinear when, for instance, $\frac{dx}{dt} = \sum_{i=1}^{r} n_i v_i^m$, with m as a number greater than 1 and then linear algebra would not be as useful. In this exercise you will realise the consequence of this linear relation, its use and how it relates to linear algebra. We study a simplified, toy network that has particular illustrative features shown in figure 6.7.

(a) Give the mass balances of the variable metabolite concentrations.

(b) Consider the network at steady state with reaction rate 1 equal to 10 $\frac{mM}{min}$. Give the reaction rate values of the reactions when the effluxes are equal to $v_6 = 5\ \frac{mM}{min}$, $v_4 = 2\ \frac{mM}{min}$, and $v_{10} = 3\ \frac{mM}{min}$.

(c) Think a bit more about the cycle, containing reaction 8 and 9, why do you not know their rates at steady state when you only know $v_1,\ v_4,\ v_6$, and v_{10}?

(d) Say you would know reaction rates v_3 , v_5 , v_7 and v_9 would this allow you to unambiguously determine all flux values?

(e) Another way to look at mass flow through this network at steady state is as if it is the superposition of 4 subnetworks, which each can attain a steady state on their own. Which reactions would those subnetworks contain?

(f) The flux vector $\mathbf{j}$ of this network is defined as the vector that contains

all the values of the reaction rates, so

$$\mathbf{j} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \end{pmatrix} \tag{6.34}$$

Write down the flux vectors, $\mathbf{j}_i$ of the 4 subnetworks, each having 1 column and 10 rows, and put zeros at entries that correspond to reactions that are not used in this subnetwork.

(g) Now the statement is that the following relation holds,

$$\mathbf{j} = \sum_{i=1}^{5} \alpha_i \mathbf{j}_i \tag{6.35}$$

with α_i as the contribution of the subnetwork to the flux through the entire network. Use the flux vector that you have identified in (b) to find the values of α_i when you set all the rates in the subnetworks to 1.

(h) Do the α_i values change when the effluxes change in value?

6. **Operon organization in bacterial genomes.** Genes are often contained in groups, called 'operons', in bacterial genomes. Consider again figure 6.7. We do not really understand why operon structure is as it is, one hypothesis worked out in this exercise. We know however that this cannot be the whole story because of all kinds of processes that distort the gene order on bacterial genomes such as horizontal gene transfer and because of gene shuffling due recombination events to transposons.

   (a) Why would it make sense to have enzyme 5 and 6 in one operon and 3 and 4 in another? Why are those two operons expected not to have enzyme 2 in it?

   (b) Why are enzyme 1 and 2 most likely not in the same operon?

   (c) If the cycle has an additional function, not related to this pathway, then it is expected be part of a different operon. Why could reaction 7 and 10 then still be part of the same operon?

   (d) What is the assignment of metabolic-enzyme genes into operons leading to the smallest number of operons (and the smallest number of

genes; so do not use the same gene twice) that still allows for complete flux flexibility through the network?

(e) Why do *you* think – because I (Frank) do not know the answer – that genes do not occur in multiple copies such that they can participate in different operons?

7. **Why some enzymes couple ATP hydrolysis to biochemical conversions.** In metabolism, many reactions are coupled to the hydrolysis of ATP into ADP and Pi. In this question you will study why this is so common. Consider the following reactions,

$$S \overset{1}{\rightleftharpoons} P, \quad v_1 = V_{1,MAX} \frac{s \cdot \left(1 - \frac{p}{s \cdot K_{eq,1}}\right)}{1 + s + p}$$

$$S + ATP \overset{2}{\rightleftharpoons} P + ADP + Pi, \quad v_2 = V_{2,MAX} \frac{s \cdot atp \cdot \left(1 - \frac{p \cdot adp \cdot pi}{s \cdot atp \cdot K_{eq,2}}\right)}{(1 + s + p)(1 + pi)(1 + atp + adp)}$$

The difference between reaction 1 and 2 is that enzyme 2 couples the hydrolysis of ATP to the formation of P out of S and enzyme 1 does not do this. The following constraints apply in the cell: i. P should be formed out S to allow for growth, ii. in the cell, the concentrations of P, S, ATP, ADP, and Pi fall within strict physiological bounds, they all vary between 0.1 and 10 mM, iii. the equilibrium constant of reaction 1 equals $K_{eq,1} = 10^{-3}$ and for reaction 2 it equals $K_{eq,2} = 10^3\ mM$. (In this question, do not consider that at the level of this reaction the adenosine and phosphate in ATP, ADP and Pi remain fixed.) Show that reaction 2 can form P out of S inside the cell and that reaction 1 cannot achieve this.

8. **Regulation of flux at branch points in metabolism.** In metabolism it occurs very often that a metabolic pathways branches into two directions. The balance for the metabolite concentration at this junction equals, the difference between the synthesis rate v and the two consumption rates v_1 and v_2 corresponding to the branch rates.

$$\frac{dx}{dt} = v_1 - v_2 - v_3 \tag{6.36}$$

$$v_1 = V_1 \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{x}{K_1}} \tag{6.37}$$

$$v_2 = V_2 \frac{x}{K_2 + x} \tag{6.38}$$

$$v_3 = V_3 \frac{x}{K_3 + x} \tag{6.39}$$

At steady state we denotes rates as fluxes and we get,

$$\begin{aligned} 0 &= J_1 - J_2 - J_3 && (6.40)\\ v_1 &= J_1 && (6.41)\\ v_2 &= J_2 && (6.42)\\ v_3 &= J_3 && (6.43)\\ x &= x_s && (6.44)\\ \Rightarrow V_1 \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{x_s}{K_1}} &= V_2 \frac{x_s}{K_2 + x_s} + V_3 \frac{x_s}{K_3 + x_s} && (6.45) \end{aligned}$$

Set $s = 10,\ K_s = 1,\ V_1 = 10,\ K_1 = 10,\ V_2 = 10,\ K_2 = 1, V_3 = 10$ and $K_3 = 10$.

(a) Use a plotting program, e.g. Excel, to plot the rates of the reaction as function of x. Determine the steady state value of x.

(b) Plot the ratio v_2/v_3 as function of x. Why is reaction 2 more active than 3 at low concentrations of X? Explain your result.

(c) Double the maximal rate of third reaction. What happens to the steady state flux and the concentration of x? Explain you result.

(d) Half the maximal rate of third reaction. What happens to the steady state flux and the concentration of x? Explain you result.

(e) What happens when you reduce K_3 by a factor of 2? Explain your result.

9. **Supply and demand analysis of metabolic pathways.** Negative feedback occurs often in metabolic pathways. The negative feedback metabolite cuts the pathway into blocks or systems: a supply and a demand system (figure 6.9). In this exercise you will study the basic consequences of negative feedback in metabolic pathways, using a simplified model that captures the main effects of negative feedback,

$$\frac{dx}{dt} = \underbrace{V_s \frac{1}{1 + \left(\frac{x}{K}\right)^n}}_{\text{supply rate}} - \underbrace{V_d \frac{x}{1+x}}_{\text{demand rate}} \qquad (6.46)$$

The basic parameter setting is that all parameters equal 1. In all plots below, plot always the supply rate and the demand rate.

(a) Investigate the influence of n. Use Excel to plot the supply rate and degradation rate as function of x. Vary n: set it to 1, 2, 4, 6, and 8.

(b) Investigate the influence of K. Set n to 4 and vary K: set it to 0.1, 0.5, 1, 2, 4.

(c) Investigate the influence of V_1. Set n to 1 and to 6. Vary for those two cases V_1: set it to 0.5, 1, and 2.

(d) Investigate the influence of V_2. Set n to 1 and to 6. Vary for those two cases V_2: set it to 0.5, 1, and 2.

(e) Is it true that the effect of a strong feedback is that the concentration of x hardly changes when changes are made to either V_1 or V_2 (metabolite homeostasis)? And that the influence of the maximal rate of the demand system on the steady-state flux can greatly exceed that of the supply system if metabolite homeostasis occurs (flux control by demand)?

Metabolite homeostasis and flux control by demand have been identified as the main functional consequences of negative feedback in metabolic pathways.

10. **Steady-state responses of metabolic pathways to changes in metabolic enzyme concentration through gene expression.** When the environment changes many metabolic pathways are adapted to the new condition via increased or reduced expression of genes coding for metabolic enzymes.

(a) Say the steady-state flux in the pathway increases because the first enzyme is increased in concentration. What happens to the steady state concentrations of X_1, X_2, X_3 and X_4?

(b) Consider the change in the steady state of the metabolic pathway when enzyme 2 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?

(c) Consider the change in the steady state of the metabolic pathway when enzyme 3 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?

(d) Consider the change in the steady state of the metabolic pathway when enzyme 4 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?

(e) Consider the change in the steady state of the metabolic pathway when enzyme 5 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?

11. **A different look at a classical metabolic pathway: glycolysis in yeast** Central to most biochemical pathways is glycolysis, the breakdown of glucose to pyruvate in 10 enzymatic steps. The details of the pathway

can be looked up in any textbook of biochemistry, or on wikipedia. Here we simplify and look at some interesting aspects from a systems biology point of view. The simplified pathway is depicted in figure 6.10

(a) make a stoichiometry matrix N of this pathway: take glucose and pyruvate fixed, so they do not need to be balanced and are therefore not part of the matrix. Each row represent a balance for a metabolite, each column correpsonds to a reaction, such that $N \cdot v$ form the set of balances for this system.

(b) apply Gaussian elimination on this matrix to transform it into its reduced row echelon form and decide on the number of row dependencies in this pathway. Think about what such a row dependency means in biochemical terms. For example, you should readily see that the rows for ATP and ADP sum up to 0 (and so these two rows are dependent). What does this mean? Provide a similar explanation for the other dependency.

(c) in the reduced row echelon form you can read off the solution to the equation: $N \cdot v = 0$. This will give you dependencies between the steady state rates, which we call fluxes.

(d) Show that indeed, in steady state, glycolysis yields 2 ATP per glucose So we can conclude that $v5$, the rate of ATP expenditure, is two times the rate of glucose conversion, $v1$. So 2 ATP was made per Glc.

(e) Sketch the solution space (or null space) of this system in the $v1$,$v5$ plane under the constraints that $v1 > 0$ and $v5 > 0$.

(f) inspect in the reduced row echelon form of N what would happen if you would not have reaction 5 in, i.e. no reaction that consumes ATP? Do this by simply removing the last column of N. Link the mathematical consequence to biochemical reasoning: can ATP be in steady state, then, and why (not)?

(g) now we will inspect a peculiar state that glycolysis get in. We put $v5$ back in, and let's assume that $v2$ and $v3$ act really fast and FbP, DHAP and GAP form one pool, and effectively $v1$ produces 2 GAP molecules. $v2$ and $v3$ are then simply thrown out of the model. We remain with 3 reactions with the following kinetics:

$$v1 = V_{m1} \cdot Glc \frac{atp}{atp + 0.5} \tag{6.47}$$

$$v4 = V_{m4} \frac{gap}{gap + 1} \tag{6.48}$$

$$v5 = k_5 \cdot atp \tag{6.49}$$

Suppose now that V_{m4} is really low, say 0.1 (in reality this is caused by a low inorganic phosphate concentration whose action we model through V_{m4}). V_{m1} and k_5 are set to 1.

i. in this state gap will accumulate to very high levels because of the low activity of V_{m4}. What will then be the rate of $v4$?

ii. construct a *rate characteristic* by plotting ATP production ($v4$) and ATP consumption ($v1 + v5$) as a function of *atp*. Where is the steady state, at what flux and *atp* concentration?

iii. now we look at the balance for *gap*: this now reads (with $v2$ and $v3$ removed): $\frac{dgap}{dt} = 2v1 - v4$. Now construct a rate characteristic around *gap* (at the steady state *atp* level from the previous question!): what is your conclusion?

12. **Can gene expression optimise the flux through a metabolic network under a constraint of fixed maximal enzyme available for investment in the metabolic network?** As biologists we are all familiar with gene activity regulation leading to changes in protein concentrations as function of conditions. We accept this as a fact. Why the gene regulation functions as it does is however generally not clear. Some would argue that the gene regulation mechanisms have evolved to maximise fitness of the bacterium. Here we explore the consequences of this hypothesis and ask whether gene regulation can maximise the fitness of a metabolic pathway. We define the fitness of metabolic pathway as its steady state flux divided by the total enzyme amount invested in this pathway. The gene regulation mechanism should maximise the fitness as function of external conditions. Does such a gene regulation mechanism exist? And how does its parameterisation depend on the kinetics of the metabolic enzymes? In this question you will answer those questions yourself, step by step. We start by introducing the metabolic network, it is the simplest that is still realistic and illustrates all the features of realistic networks that are clearly involve hundreds of enzymes,

$$\underline{S} \overset{1}{\rightleftharpoons} X \overset{2}{\rightleftharpoons} \underline{P} \tag{6.50}$$

$$v_1 = e_1 \underbrace{k_1 \frac{s}{1+s+x}}_{f_1(x)} = e_1 f_1(x) \tag{6.51}$$

$$v_2 = e_2 \underbrace{k_2 \frac{x}{1+x}}_{f_2(x)} = e_2 f_2(x) \tag{6.52}$$

$$e_T = e_1 + e_2 \tag{6.53}$$

$$\frac{dx}{dt} = v_1 - v_2 = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} \tag{6.54}$$

We denote the steady-state flux by J and it is defined by the steady-state relation,

$$J = v_1 = v_2 \tag{6.55}$$

What we want to maximise is the fitness of the metabolic pathway, F,

$$F = \frac{J}{e_T} = \frac{J}{e_1 + e_2} \tag{6.56}$$

So given the amount of enzyme available, e_T, we ask for the distribution of enzymes that maximises J, which is the same as saying that we maximise $\frac{J}{e_T}$. Maximising $\frac{J}{e_T}$ is the same as minimising $\frac{e_T}{J}$.

(a) Show that $\frac{e_T}{J}$ obeys,

$$\frac{e_T}{J} = \frac{1}{f_1} + \frac{1}{f_2} = \frac{1+s+x}{k_1 s} + \frac{1+x}{k_2 x} = \frac{1}{k_1 s} + \frac{1}{k_1} + \frac{x}{k_1 s} + \frac{1}{k_2 x} + \frac{1}{k_2} \tag{6.57}$$

(b) When $\frac{e_T}{J}$ is minimal then $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) = 0$. Why is this true?

(c) Determine $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right)$ and solve for x. Call this x, the optimal x, denoted by x_o.

(d) Show that the relation $\frac{v_1}{v_2} = 1$ leads to the relation $e_2 = e_T \frac{f_1}{f_1+f_2} = e_T \frac{1}{1+\frac{f_2}{f_1}}$.

(e) The problem now is that the previous relation is always true, also in non-optimal states. However, the requirement that $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) = 0$ leads to an expression for $\frac{f_2}{f_1}$ that is only true in the optimal state, because when $x = x_o$ we have,

$$0 = \frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) = \frac{d}{dx}\frac{1}{f_1} + \frac{d}{dx}\frac{1}{f_2} = \frac{\partial \frac{1}{f_1}}{\partial f_1}\frac{\partial f_1}{\partial x} + \frac{\partial \frac{1}{f_2}}{\partial f_2}\frac{\partial f_2}{\partial x} \tag{6.58}$$

$$= -\frac{1}{f_1^2}\frac{\partial f_1}{\partial x} - \frac{1}{f_2^2}\frac{\partial f_2}{\partial x} = -\frac{1}{f_1}\frac{\partial \ln f_1}{\partial x} - \frac{1}{f_2}\frac{\partial \ln f_2}{\partial x}$$

$$\Rightarrow -\frac{1}{f_1}\frac{\partial \ln f_1}{\partial x} = \frac{1}{f_2}\frac{\partial \ln f_2}{\partial x} \quad \Rightarrow \frac{f_2(x_o)}{f_1(x_o)} = -\frac{\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o}}{\left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}} \tag{6.59}$$

Such that at the optimal steady state we have the relation $e_2 = e_T \frac{1}{1+\frac{f_2}{f_1}} = e_T \frac{1}{1-\frac{\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o}}{\left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}}}$. Determine $-\frac{\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o}}{\left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}}$ and then the equation for e_2.

(f) The problem is that $-\frac{\frac{\partial \ln f_2}{\partial x}}{\frac{\partial \ln f_1}{\partial x}}$ depends still on s, such that we have relation that expresses e_2 in terms of s and x_o, and we would like to

use this relation to choose a gene regulation mechanism that relates e_2 to x only, and not to s, because the gene regulation circuit depend on x – for instance, because a transcription factor binds to x and this transcription determines the synthesis of e_2. Using the previous relation, $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$, which holds in the optimum, we can express s in terms of x_o. What do we obtain for s and for e_2?

(g) The relation that we have just found, i.e. $e_2 = e_T \frac{k_1 x_o(1+x_o)}{k_2 x_o^2 + k_1(1+x_o)^2}$, relates e_2 to x_o; it is therefore a relation for $e_{2,o}$! What should be the mass balance for e_2 to obey this relation at steady state?

(h) Remark, not a question: The statement is now that the following coupled metabolic-transcription-translation system,

$$\frac{dx}{dt} = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} \tag{6.60}$$

$$\frac{d}{dt}e_2 = e_T \frac{k_1 x(1+x)}{k_2 x^2 + k_1(1+x)^2} - e_2, \tag{6.61}$$

gives the optimal steady-state for every value of s! Because this system always has as steady state: $x = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$, which is the requirement for optimality! Note that the $\frac{k_1 x(1+x)}{k_2 x^2 + k_1(1+x)^2}$ is very similar to a hyperbolic equation of x (you can verify by plotting it for different values of k_1 and k_2), suggesting that a single transcription factor that binds to X and to the DNA suffices for optimal regulation. In fact you can rewrite this equation into $\frac{x+x^2}{1+2x+\left(\frac{k_2}{k_1}+1\right)x^2}$ which sort of corresponds to a model of cooperative binding of the transcription factor to the promoter with two binding sites.

(i) Another remark: Another way of solving the problem of finding an optimal gene regulation network is to find a e_2 function that makes sure that $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$. This e_2 function, in terms of x, is then the steady state input-output relation of the optimal gene regulation network. How can you do this? Set $\frac{dx}{dt} = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} = 0$ and solve for x. Set this equation equal to $\frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$ and solve for e_2. Then we have found an e_2 steady state that is optimal. The problem now is that this e_2 equation is still in terms of s and not in terms x – and we want the gene network to sense the metabolic state and not the environmental state. To solve this, solve s from $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$ and substitute it in the equation in the e_2 relation to arrive at one that is in terms of x, and no longer in terms of s. This is bit faster than what was suggested above.

(j) Another remark: note that since the objective is solely defined in terms of metabolic properties, the parameterisation of the optimal

gene network is ultimately completely set by metabolic properties alone. So, the gene regulation is a 'controller' of the metabolic system that has the right information of the metabolic system to be able to steer it always to the optimum, regardless of the environment. Isn't that amazing? I think that is the right way of thinking about gene regulation. However this shifts the problem to identifying what the control objective, the fitness objective, is. It is certainly not always J/e_T. How to do that is still not clear to me (= Frank).

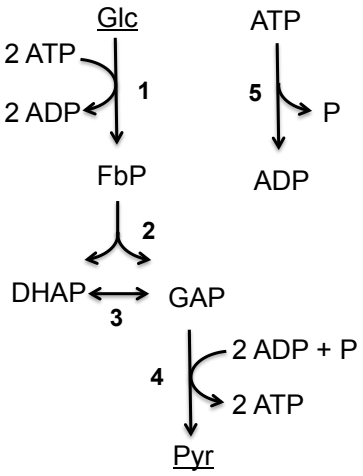


Figure 6.10: **Simplified glycolytic pathway.** The pathway starts with adding two phosphates on glucose to form fructose 1,6-bisphosphate (FbP), with ATP as phosphate donor. FbP is split into two molecules, dihydroxyacetone phosphate (DHAP) and glyceraldeyhde 3-phosphate (DHAP), which are interconverted very easily. GAP is subsequently converted into pyruvate in a number of steps, by which two ATP molecules are made for each GAP molecule. P is inorganic phosphate.

6.9 Key messages of this chapter

1. Metabolic activity is tuned towards the available nutrients as well as the cellular requirements.
2. During balanced exponential growth, all concentrations are constant.

3. The time before a newly synthesised protein reaches steady state during exponential growth depends on the degradation rate constant. Stable proteins require more than 4 generations to reach more than 95% of their steady state concentration.

4. Change in metabolic gene expression depends on the metabolic and environmental state, as shown by the *lac* operon.

5. Enzymes that are regulated in metabolism by metabolic intermediates are both sensitive to those the concentrations of those intermediates and they have an influence on pathway flux when changed in activity. Such rules-of-thumb follow immediately from the theory associated with metabolic control analysis.

Chapter 7

Answers to exercises

7.1 Answers to exercises of Chapter 2

7.2 Exercise Section 2.3

1. In this chapter, we will be doing calculations with units. To practise, answer the following exercises. Realise that this way of thinking is correct: $2\ m^3 = 2 \times (10\ dm)^3 = 2000\ dm^3 = 2000\ l$ (l is liter). **You may have to look up the meaning of f, μ, and p!.**

   (a) $1\ m = 100\ cm = 10^2\ cm$, $1\ m^2 = 1 \times (100\ cm)^2 = 100^2\ cm^2 = 10000\ cm^2 = 10^4\ cm^2$, and $1\ m^3 = 1\ (100\ cm)^3 = 100^3\ cm^3 = 10^6\ cm^3$

   (b) $1\ dm^3 = 1\ l$, $1\ cm^3 = 1\ (0.1\ dm^3) = 0.1^3\ dm^3 = 10^{-3}\ l$, and $1\ nm^3 = 1\ (10^-9m)^3 = 1\ (10^-9 \times 10dm)^3 = (10^-8)^3\ dm^3 = 10^{-24}\ dm^3 = 10^{-24}\ l$

   (c) $1\ fl = 10^{-15}\ l = 10^{-15}\ 10^6\ \mu l = 10^{-9}\ \mu l$, $1\ ml = 10^{-3}\ l$, and $1\ \mu l = 10^{-6}\ l = 10^{-6}10^{15}\ fl = 10^9\ fl$

   (d) $1\ mol = 6 \times 10^{23}\ molecules$ (Avogadro's number), $1\ \mu mol = 10^{-6} \times 6 \times 10^{23}\ molecules = 6 \times 10^{17}\ molecules$, and $1\ pmol = 10^{-12} \times 6 \times 10^{23}\ molecules = 6 \times 10^{11}\ molecules$

2. Assume that the average protein contains 300 amino acids.

   (a) How many amino acids should a cell make during its cell cycle when it contains 2×10^6 proteins at birth?
   Answer: a cell doubles itself from birth to division, so its needs to make $2 \times 10^6\ proteins$ and this corresponds to $2 \times 10^6 proteins \times 300 \frac{amino\ acids}{protein} = 600 \times 10^6 amino\ acids = 6 \times 10^8 amino\ acids$.

   (b) Calculate how many nitrogen atoms occur on average in an amino acid (use the first table on this Wikipedia page `https://en.wikipedia.`

`org/wiki/Amino_acid`).
Answer: See Figure 7.1 and $(4+3+2+1+1+1+1+2+2+1+1+1+1+1+1+1+2+1+1)/21 = 30/21 = 1.4\ \frac{nitrogen\ atom}{amino\ acid}$.

(c) Given the previous two answers, how many ammonium molecules (NH_4^+) should a cell consume during its cell cycle?
Answer: Assuming that all amino acids occur evenly across all proteins – they do not, but let's not dwell on that – the number of ammonium molecules requires equals:
$6 \times 10^8 amino\ acids \times 1.4\ \frac{nitrogen\ atom}{amino\ acid} \times 1 \frac{ammonium\ molecule}{nitrogen\ atom} = 6(1+0.4)10^8 ammonium\ molecules = (6+2.4)10^8 ammonium\ molecules = 8.4 \times 10^8 ammonium\ molecules.$

(d) How much does that amount of ammonium weigh in femtograms?
Answer: One mol of NH_4^+ weighs $14 + 4 \times 1 = 18$ gram, since the molar weights of nitrogen and hydrogen are respectively 4 and 1 gram/mol. An amount of $8.4 \times 10^8 ammonium\ molecules$ equals $8.4 \times 10^8 ammonium\ molecules \times \frac{1}{6 \times 10^{23}} \frac{mol}{molecules} = \frac{8.4}{6} 10^{-15}\ mol\ ammonium$ and this equals in femtograms $\frac{8.4}{6} 10^{-15}\ mol\ ammonium \times 18\ \frac{gram}{mol} = \frac{18 \times 8.4}{6} 10^{-15} gram = \frac{18 \times 8.4}{6}\ fgram = 25.2\ fgram.$

3. The length of the DNA of *E. coli*, a bacterium, is 1.5 mm long, consists of 4558953 bps, and its DNA polymerase runs at a speed of 800 bps/sec. How much time does it take for this enzyme to have replicated *E. coli*'s DNA by 50%?
Anwer: The time it takes equals $4558953\ bps \times \frac{1}{800}\ \frac{sec}{bps} \times 0.5 = \frac{4558953}{1600} = 2849\ sec = 2849\ sec \frac{1}{60} \frac{sec}{minute} = 47\ minutes.$

7.3 Exercise Section 2.7

1. We will consider exponential growth of cells and how this can become limited by food supply. Note by the way that $N(t)$ does not mean $N \times t$ but N at t!

(a) The meaning of $\frac{dN(t)}{dt}$ is that it equals the "rate of change" of $N(t)$ as function of time t. When you think about this for a moment you will probably realise that this equals the slope of a plot of $N(t)$ as function of t. Since it is the slope you can think of $\frac{dN(t)}{dt}$ as being equal to $\frac{N(t+dt)-N(t)}{t+dt-t}$ – which is after all the equation for the slope of the $N(t)$-vs-t plot. Given that $\frac{dN(t)}{dt} = \mu N(t)$ (with μ as the growth rate and positive), what happens to the slope in the $N(t)$-vs-t plot when t increases? (Does $N(t)$ increase with time? Why?)
Answer: The slope equals $\mu N(t)$ and $\mu > 0$ and $N(t) > 0$ (you cannot have a negative number of cells), thus $\frac{dN}{dt} = \mu N(t) > 0$; the slope rises, the number of cells increases.

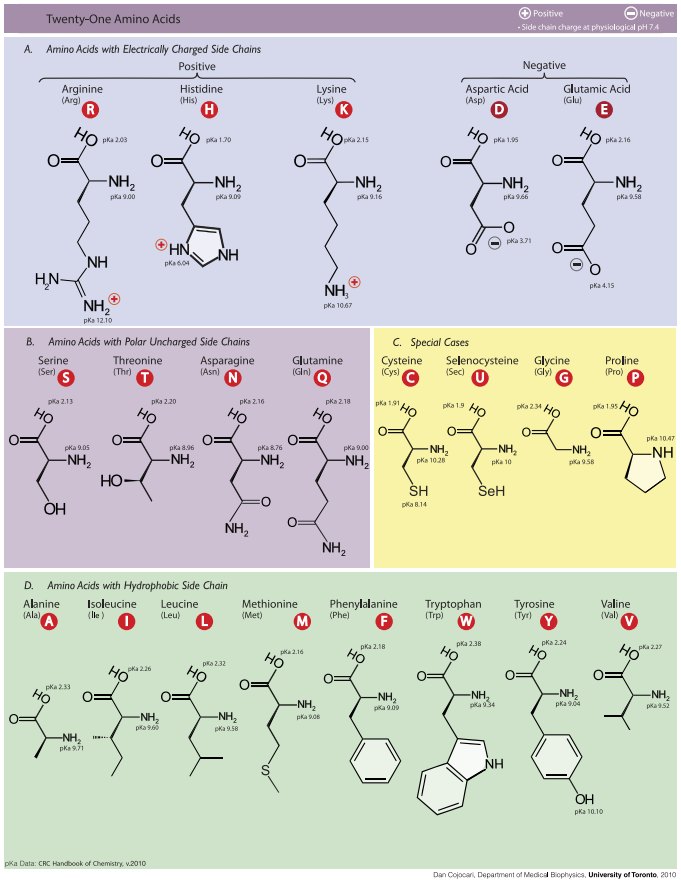


Figure 7.1: **Molecular structures of the 21 amino acids, the molecular building blocks of proteins.**

(b) When a population of cells grows according to $\frac{dN(t)}{dt} = \mu N(t)$ new cells are all the time being made by the existing cells. This requires nutrients. What happens to the nutrient concentration, denoted by $s(t)$, over time? How would you mathematically describe the slope of $s(t)$-vs-t? Would it have to depend on $s(t)$, $N(t)$ or both?
Answer: since each cells has the same nutrient requirement and the number of cells increases, the nutrient requirement increases too, so $s(t)$ will drop as function of time (and nutrients are not replenished, they will run out.) The slope of $s(t)$-vs-t, i.e. $\frac{ds(t)}{dt}$ therefore depends on $N(t)$. And on $s(t)$ (!), since $s(t)$ cannot become negative and no

consumption of nutrients can occurs when $s(t) = 0$.

(c) What is the problem with the following description $\frac{ds(t)}{dt} = -kN(t)$ with k as a nutrient consumption rate per cell? Why is it impossible?
Answer: $\frac{ds(t)}{dt} = -kN(t)$ indicates that the concentration of nutrient $s(t)$ always drops. So given a starting amount, this amount will decrease until it is finished. Then $s(t) = 0$, but according to $\frac{ds(t)}{dt} = -kN(t)$ consumption then still continues so that $s(t)$ becomes negative, and this is not possible. Thus, $\frac{ds(t)}{dt}$ needs to depend on $N(t)$ and $s(t)$!

(d) Eventually, when the nutrients run out, the growth of the population of cells stops. This can be described by the following differential equation $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t)$ with K as the so-called carrying capacity (or yield) of the environment.

i. What changes in the equation when we double the amount of nutrients?
Answer: The equation indicates that $N(t)$ becomes constant, i.e. $\frac{dN(t)}{dt} = 0$ when $N(t) = K$. Thus the number of cells that can be reached equals K and depends on the amount of nutrients supplied (of course). If twice as many nutrients are applied then K doubles in value, twice as many cells can be made after all.

ii. How would you characterise the growth when $N(t)$ is small relative to K?
Answer: then $\frac{dN(t)}{dt} \approx \mu N(t)$, the growth is exponential.

iii. What is growth rate of the culture when $N(t) = K$?
Answer: it is zero since then $\frac{dN(t)}{dt} = 0$

iv. Sketch the dependency of $N(t)$ on t. (If you want actual numbers set $N(t) = 1$, $K = 1000$ and $\mu = 1\ hr^{-1}$.)
Answer: we have just determined that for $N(t) << K$ the growth is exponential and for $N(t) \approx 0$, so we obtain the sketch shown in Figure 7.2.

v. What has happened when $N(t) = K$?
Answer: Growth has stopped because the nutrients have been depleted.

(e) Calculate the generation time of a cell in minutes that grows at a rate of 1 hr^{-1}.
Answer: In the main text you can find that the generation time, t_{gen} – the time elapsed from cell birth to division – relates to the growth rate, μ, as $t_{gen} = \frac{\ln 2}{\mu}$; thus, $t_{gen} = \frac{\ln 2}{1\ hr^{-1}} = \ln 2\ hr = 0.67\ hr$.

2. "Numerically solving a differential equation such as $\frac{dN(t)}{dt} = \mu N(t)$" means that the values of $N(t)$ are calculated at different times given a starting value of N at time zero, denoted by $N(0)$. From the definition of $\frac{dN(t)}{dt}$ it becomes clear how this can be done.

Figure 7.2: **Sketch of the dynamics of $N(t)$, so its value as function of time, for $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t)$.**

(a) Write down the definition of $\frac{dN(t)}{dt}$ as the slope of a $N(t)$-vs-t curve.
Answer: The slope at time t equals $\frac{dN(t)}{dt} = \frac{N(t+dt)-N(t)}{dt} \approx \frac{N(t+\Delta t)-N(t)}{\Delta t}$ (Note that $\Delta t >> dt$). So given a value for t you determine the associated $N(t)$ value, by using the $N(t) - vs - t$ plot, choose a $\Delta t > 0$ and determine $N(t + \Delta)$ and then you can determine the slope.

(b) Express the value of N at time $t + dt$ in terms of N at t, $\frac{dN}{dt}$ at t and dt. (We assume that dt has a constant value.)
Answer: We set $dt = \Delta t$, we do this because mathematically dt is defined as infinitesimally small and Δt as any value. Thus, $\frac{N(t+\Delta t)-N(t)}{\Delta t} = \mu N(t)$ and therefore $N(t + \Delta t) - N(t) = \mu N(t)\Delta t$ and $N(t + \Delta t) = N(t) + \mu N(t)\Delta t$; or, equivalently $N(t+\Delta t) = N(t) + \frac{N(t+\Delta t)-N(t)}{\Delta t}\Delta t$ (new value = old value + change). Note that $N(t + \Delta t) = N(t) + \mu N(t)\Delta t$ is useful relation, given the number of cells at time t you can determine the new number of cells at $t + \Delta t$!

(c) Calculate $N(dt)$ when $\mu = 1\ hr^{-1}$, $dt = 0.1$ and $N(0) = 1$.
Answer: We set $dt = \Delta t$. Let's exploit: $N(t+\Delta t) = N(t)+\mu N(t)\Delta t$. So, $N(\Delta t) = N(0) + \mu N(0)\Delta t = 1 + 1 \times 1 \times 0.1 = 1.1$.

(d) Calculate $N(2dt)$ given $N(dt)$.
Answer: $N(2\Delta t) = N(\Delta t) + \mu(N(\Delta t))\Delta t = 1.1 + 1 \times 1.1 \times 0.1 = 1.1 + 0.11 = 1.21$.

(e) Now use Excel to calculate $N(t)$ as function of time t from 0 to 10. If you are not familiar with Excel do this with a fellow student or consult `http://bmi.bmt.tue.nl/sysbio/Education/Excel_Euler_simulation`.

`pdf`.
Answer: See the Excel file *ode exponential growth.xlsx*.

(f) Check that the resulting curve obeys $N(t) = N(0)e^{\mu t}$.
Answer: See the Excel file *ode exponential growth.xlsx*. The red dots are the correct answer $N(t) = N(0)e^{\mu t}$ and the blue dots are the approximation $N(t+\Delta t) = N(t)+\mu N(t)\Delta t$. If you make Δt smaller the approximation is beter and the red and blue data points are closer to each other.

(g) Solve also $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t)$ numerically, using Excel, when $\mu = 1\ hr^{-1}$, $K = 10$ and $N(0) = 1$. Why is this curve and the previous one you have made the same for small times?
Answer: See the Excel file *ode carrying capacity.xlsx*. They are the same when because $N(t)/K \approx 0$ such that $\frac{dN(t)}{dt} = \mu\left(1 - \frac{N(t)}{K}\right)N(t) \approx \mu N(t)$, the exponential growth regime.

7.4 Exercise Section 2.9

Diffusion lies at the basis of life. Without it, no movement of molecules would occur and reactions cannot take place. It also limits life, since molecules cannot move faster than by diffusion. The speed of reactions involving two (or more) substrates is therefore limited by diffusion rates.

1. Read the abstract of Klumpp et al. [13]. What limits growth rate in *E. coli* according to them?
   Answer: The key statement is *"Here, we show that the slow diffusion of the bulky tRNA complexes in the crowded cytoplasm imposes a physical limit on the speed of translation, which ultimately limits the rate of cell growth."*. Thus, supply of amino acids to the ribosome is limiting the synthesis of proteins, which are essentially chains of amino acids. Those amino acids are bound to tRNA, which are large and slowly diffusing, prior to their transfer to ribosomes which add those amino acids to the growing amino acid chain. Thus diffusion plays a key role in living processes.

2. The time for two molecules to find each other in a cell volume is given by equation 2.8. What do you expect happens to this time – the time when the first collision occurs – when you have N copies of each molecule instead of 1 of each? Why does the time decrease when one of the molecules is bigger?
   Answer: Imagine that you have to find a friend when you are both blindfolded. You will both be random walking until you bump into each other. Suppose we N clones of you and of your friend. Now, clearly, you bump into each other a lot sooner. (In fact the time is shortened by the following factor $1/N^2$.) Say now that you have a very fat friend so that the volume he/she occupies is much greater than of a slim friend. You find the fat friend sooner because he occupies more space.

3. Why is the collision time of two molecules (equation 2.8) lower to the waiting time for them to form a complex?
   Answer: In order for two molecules to form a complex in addition to the collision also chemical processes need to occur like the binding itself, this also takes time.

4. Calculate how many minutes that it takes for one copy of a transcription factor with radius 5 nm to find a promoter of a gene. The diffusion coefficient of the transcription factor is 5 $\mu m^2/s$. Assume a spherical cell with a radius of 1.5 μm. Why can you assume that the promoter does not move?
   Answer: The equation is $\tau = \frac{V}{4\pi(D_{TF}+D_{PROM})(r_{TF}+r_{PROM})}$, let's substitute the known parameters in the right units:

$$\tau = \frac{\frac{4}{3}\pi(1.5\ \mu m)^3}{4\pi(5\ \frac{\mu m^2}{s} + 0)(5\ nm \times 10^{-3}\frac{\mu m}{nm} + 5\ nm \times 10^{-3}\frac{\mu m}{nm})} = \frac{14.13\ \mu m^3}{0.2\pi\frac{\mu m^3}{s}} = 22.5\ s$$

$$= 22.5\ s \times \frac{1}{60}\ \frac{min}{s} = 0.38\ min$$

5. How much slower does a molecular complex move than any of its components? (With which factor is the diffusion coefficient reduced?)
   Answer: The diffusion coefficient of a molecule equals $D = \frac{kT}{6\pi\eta a}$ with a as its radius. Say we have a complex that consists of two of those molecules. This complex will be larger than the two molecules so its radius a_c is larger than a. The ratio of the diffusion coefficient equals $D_c/D = a/a_c$, so the $D_c = \frac{a}{a_c}D$ the complex moves slower with the following factor $\frac{a}{a_c}$.

6. Why is it advantageous for a cell to construct large complexes on DNA, rather than forming them first in the cytoplasm and after that having them bind to the DNA?
   Answer: One reason might be that those large complexes move very slowly.

7.5 Exercise Section 2.11

1. Consider the following reaction

$$F + P \rightleftharpoons FP$$

   with F denoting a transcription factor and P the promoter of a gene. We will assume that the concentration of the transcription factor greatly exceeds that of the promoter, which occurs at a single copy per cell. Accordingly, we assume that the concentration of the unbound, "free" transcription factor is constant. We denote it by f_T. The total concentration of the promoter equals the sum of the concentrations of the free and occupied promoter, i.e. $p_T = p + fp$. We also assume that p_T is constant. All assumptions made until now are realistic.

(a) Show that the rate of the association (complex formation) reaction can be written in terms of only one unknown promoter concentration, choose fp.
Answer: $v^+ = k^+ \times f \times p \approx k^+ f_T(p_T - fp)$

(b) Show that the rate of the dissociation reaction can written in terms of only one unknown promoter concentration, choose fp.
Answer: $v^- = k^- fp$

(c) Which rate difference equals $\frac{dfp}{dt}$?
Answer: $\frac{dfp}{dt} = v^+ - v^- = k^+ f_T(p_T - fp) - k^- fp$, note that k^+, f_T, p_T k^- are all constants (they are the parameters). This $\frac{dfp}{dt}$ only depends on one variable: fp, the concentration of FP.

(d) Sketch the association rate and the dissociation rate as function of the concentration of fp. (Make a plot of the two rates as function of fp, since you do not know the parameter values you have think carefully about this.)
Answer: See the Figure 7.3.

(e) Can those two rates become equal? What happens then?
Answer: yes they can (and they will always do so). When they do so $v^+ = v^-$ and $\frac{dfp}{dt} = v^+ - v^- = 0$ such that the concentration of fp remains constant. This state is called an equilibrium state.

(f) Calculate the concentration of fp when those two rates are equal. Sketch this concentration as function of f_T. Does the outcome make sense? Is it what you would expect?
Answer: At equilibrium, we denote the concentration of FP by fp_e, and the following relation holds $k^+ f_T(p_T - fp_e) = k^- fp_e$. Solving for fp_e gives $k^+ f_T p_T - k^+ f_T fp_e = k^- fp_e \Rightarrow k^+ f_T p_T = k^+ f_T fp_e + k^- fp_e \Rightarrow k^+ f_T p_T = (k^+ f_T + k^-) fp_e$ such that

$$fp_e = \frac{k^+ f_T p_T}{k^+ f_T + k^-} = p_T \frac{f_T}{f_T + \frac{k^-}{k^+}}.$$

Next we need to sketch the dependency of fp_e on f_T. When $f_T = 0$ then $fp_e = 0$, when $f_T \to \infty$ then $fp_e \to p_t$, when $f_T = \frac{k^-}{k^+}$ then $fp_e \to p_T/2$. Give the dependency we are after is shown in Figure

7.6 Exercise section 2.14

1. Calculate how many proteins fit in *E. coli*'s cell membrane assuming that their radius is 5 nm. Assume a radius of an *E. coli* cell of 1 μm. How many proteins fit in its periplasm if this compartment is 15 nm thick? What is the ratio of the protein numbers in the membrane and periplasm over the number of proteins in its cytoplasm?
We assume *E. coli* to be spherical. Then it has as area $A_{cell} = 4 \times \pi \times$

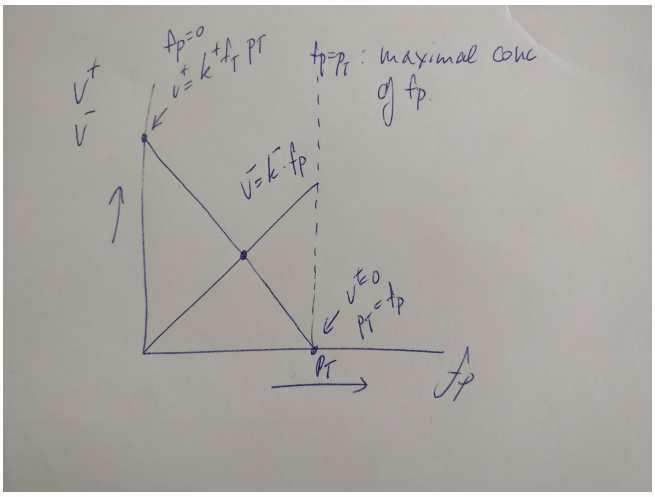


Figure 7.3: **Sketch of the association and dissociation rate as function of fp.**

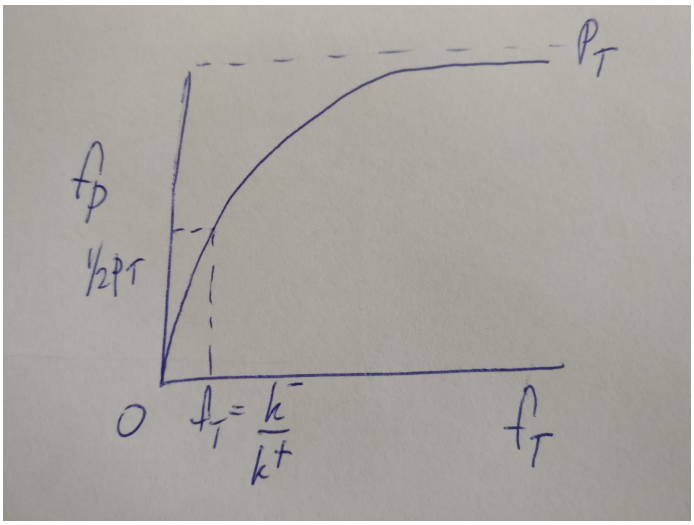


Figure 7.4: **Sketch of the dependency of the concentration of the promoter-TF complex on the total concentration of transcription factor.**

r_{cell}^2 with r_{cell} as the cell radius. The area of a disk equals $A_{protein} = \pi \times r_{protein}^2$. So the number of proteins in the membrane of *E. coli* equals maximally; $\frac{4\times\pi\times r_{cell}^2 \ \frac{\text{m}^2}{\text{cell membrane}}}{\pi\times r_{protein}^2 \ \frac{\text{m}^2}{\text{protein}}} = \frac{4\times\pi\times(10^{-6}\ \text{m})^2}{\pi\times(5\times10^{-9}\ \text{m})^2} \ \frac{\text{protein}}{\text{cell membrane}} =$

160000 $\frac{\text{protein}}{\text{cell membrane}}$. Next, we calculate the volume of the periplasm of an *E. coli* cell using the equation of the volume of a sphere $V = \frac{4}{3} \times \pi \times r^3$: $V_{periplasm} = \frac{4}{3} \times \pi \times (10^{-6}m + 15 \times 10^{-9}m)^3 - \frac{4}{3} \times \pi \times (10^{-6}m)^3 = 1.9 \times 10^{-19}m^3$. The volume of a protein equals $V_{protein} = \frac{4}{3} \times \pi \times (5 \times 10^{-9}m)^3 = 5.2 \times 10^{-25}m^3$. So, maximally $\frac{1.9 \times 10^{-19}}{5.2 \times 10^{-25}} = 0.4$ million protein fit in the periplasmic space. Which is an appreciable amount of the entire protein content of a cell. All those proteins need to be transported over the plasma membrane during cell growth! This requires an enormous transport capacity of a cell especially at maximal growth rates.

2. Find the area of the earth and the distance of the moon to earth on Google. Calculate the time that it takes for a bacterium to cover the area of planet earth if the bacterium has an area of 1 μm^2 and grows exponentially at a rate of 1 hr^{-1}. How much time does it takes for this bacterium to fill the distance between the earth and the moon if it is 1 μm thick if all cells are stacked on top of each other?
$510 \times 10^6\ km^2 = 510 \times 10^6 \times (1000 \times m)^2 = 5.1 \times 10^{14} m^2$ is the area of planet earth. $\frac{5.1 \times 10^{14}}{1 \times (10^{-6}m)^2} = 5.1 \times 10^{26}$ bacteria fit on the surface of planet earth. Exponential growth obeys $N(t) = N(0)e^{\mu \times t}$; so, we have $5.1 \times 10^{26} = e^t$ and $t = \ln(5.1 \times 10^{26}) = 61.5$ hour! Wow, this is fast! The distance to the earth and the moon is $384 \times 10^3\ km = 384 \times 10^6 m$. The number of bacteria that can be stacked on top of each other to pass this distance equals $\frac{384 \times 10^6 m}{10^{-6}m} = 3.8 \times 10^{14}$. This takes $\ln(3.8 \times 10^{14}) = 33.6$ hours!

3. **Stoichiometry of nutrient uptake fluxes during steady-state growth.** The elemental composition of an *E. coli* cell equals $CH_{1.77}O_{0.49}N_{0.24}$. These elements are components of the molecules making up biomass such as DNA, RNA, lipids and proteins.

   (a) Calculate the weight of mol of $CH_{1.77}O_{0.49}N_{0.24}$.
   Answer: Molar weight of $CH_{1.77}O_{0.49}N_{0.24}$ equals $12 + 1.77 + 0.49 \times 16 + 0.24 \times 14 = 25\ \frac{gram}{mol}$. So 25 grams of cells contains one mol of C.

   (b) A realistic value for the weight of *E. coli* cell 0.95 $pg = 0.95 \times 10^{-12} gram$. How many carbon and nitrogen atoms does a single *E. coli* cell contain?
   Answer: $\frac{0.95 \times 10^{-12}\ gram}{25\ \frac{gram}{mol}} = 3.8 \times 10^{-14} mol = 3.8\ fmol$. (f=femto=$10^{-15}$.) So one cell therefore contains 3.8 $fmol$ C and $0.24 \times 3.8\ fmol = 0.95\ fmol$ N.

   (c) How many glucose and ammonium molecules are minimally required to make a single *E. coli* cell?
   Answer: Glucose is $C_6H_{12}O_6$: $3.8 \times 10^{-15} mol \times 6 \times 10^{23} \frac{C-atoms}{mol} = 22.8 \times 10^8 C - atoms = \frac{1}{6} \times 22.8 \times 10^8\ glucose\ molecules = 3.8 \times 10^8\ glucose\ molecules$. Ammonium is NH_4: $0.95 \times 10^{-15} mol \times 6 \times 10^{23} \frac{N-atoms}{mol} = 5.7 \times 10^8 N - atoms = 5.7 \times 10^8\ ammonium\ molecules$.

This is the minimal requirement because more glucose is required to make energy and sometimes part of the C in glucose is wasted in the form of fermentation products. Wasting of NH_4 does not happen generally.

(d) When growing on glucose in mineral medium *E. coli* requires 5.9×10^9 ATP molecules to synthesise one cell. How many glucose molecules are required to attain this amount of ATP via respiration? How many via fermentation?
Answer:

4. **The number of proteins made per mRNA.** Thinking about translation is a bit similar to thinking about a conveyor belt that breaks down quickly. So during the life time of the conveyor belt, during which it carries boxes from the left to the right, the conveyor belt manages to transport a number of boxes. Multiple boxes are one the belt, and the distance between the boxes on the belt is determined by the speed of the belt and the times between consecutive placements of boxes onto the belt. If boxes are placed on the belt at a higher rate then boxes are closer to one another and if the belt runs slower the boxes are also closer to one another. A box resembles a ribosome and the moving belt corresponds to a ribosome walking over mRNA. The distance between ribosomes is determined by the translation initiation rate and the moving rate of ribosomes. The life time of the mRNA correspond to the operating time of the conveyor belt before it breaks down and stops working.

   (a) Say a ribosome produces peptide chains of a length of 20 amino acids per second. How many mRNA nucleotides does it pass in a second?
   Answer: $20 \ \frac{aa}{s} \times 3 \ \frac{nts}{aa} = 60 \ \frac{nts}{s}$.

   (b) What is the distance in nucleotides between neighbouring ribosomes on the mRNA if every two seconds a new ribosome hops into the mRNA?
   Answer: $60 \ \frac{nts}{s} \times 2 \ s = 120 \ nts$.

   (c) If every two seconds a ribosome hops on to start translation at steady state, what is time period between ribosomes leaving the mRNA transcript?
   Answer: again two seconds.

   (d) What is then the protein synthesis rate?
   Answer: one protein per 2 seconds per mRNA

   (e) What is the protein synthesis rate if two such mRNAs occur?
   Answer: 2 proteins per 2 seconds

   (f) If an mRNA lives 16 seconds how many proteins are made from it?
   Answer: 8.

   (g) Is the rate set by the elongation rate or the initiation rate in this exercise?

Answer: Initiation rate and this is likely the most realistic except for mRNA that have very fast initiation rates then ribosome collisions can occur that disappear on the time scale of translation elongation.

7.7 Extra exercise section 2.15

1. **Transcription and translation data in human cells.**

   (a) For the mRNA's: $17 \frac{\text{mRNA}}{\text{cell}} \frac{1}{6\times10^{23}} \frac{\text{mol}}{\text{mRNA}} \frac{1}{1800} \frac{\text{cell}}{\mu m^3} = \frac{17}{6\times10^{23}\times1800} \frac{\text{mol}}{(10^{-6}10\ \text{dm})^3} = 16\ \frac{\text{pmol}}{\text{l}}$. For the proteins: $50000 \frac{\text{proteins}}{\text{cell}} \frac{1}{6\times10^{23}} \frac{\text{mol}}{\text{proteins}} \frac{1}{1800} \frac{\text{cell}}{\mu m^3} = \frac{50000}{6\times10^{23}\times1800} \frac{\text{mol}}{(10^{-6}10\ \text{dm})^3} = 46.3\ \frac{\text{nmol}}{\text{l}}$

   (b) The steady state protein concentration equals,

   $$0 = \frac{d}{dt}p = k_t^p \times m_s - k_d^p \times p_s \tag{7.1}$$

   with k_t^p as the translation rate constant, k_d^p as the degradation rate constant, and m_s and p_s as the steady state mRNA and protein concentration. So,

   $$k_d^p = \frac{k_t \times m_s}{p_s} = \frac{180\ \frac{\text{proteins}}{\text{mRNA}\times\text{hour}} \times 17\ \frac{\text{mRNA}}{\text{cell}}}{50000\ \frac{\text{proteins}}{\text{cell}}} = 0.0612\ \frac{1}{\text{hour}} \tag{7.2}$$

   So, a protein lives on average 16 hours ($= \frac{1}{k_d^p}$).

   (c) So, the half life equals 9 hour $= \frac{\ln 2}{k_d^m}$; such that $k_d^m = 0.08\ \frac{1}{hour}$. At steady state, $0 = \frac{d}{dt}m = k_t^m - k_d^m m_s$ with the k_t^m as the transcription rate (constant) and k_d^m as the mRNA degradation rate constant; so, $k_t^m = k_d^m m_s = 0.08\ \frac{1}{\text{hour}} \times 17\ \frac{\text{transcripts}}{\text{cell}} = 1.36\ \frac{\text{transcripts}}{\text{hour}\times\text{cell}}$.

   (d) Proteins involved in those cellular functions live long, because they tend to have the highest concentration in a cell. A high turnover of those proteins would therefore be very costly. Long life times of those proteins is therefore an energy saving mechanism. In addition, cellular effects of metabolic enzymes are typically noticeable only after the doubling time of a human cell which varies from 17-32 hours; so having them life much shorter than this time window is not very productive.

2. **Transcription and translation data in *E. coli.***

   (a) This has two interpretations: i. if one cell out of 20 cells has on average 1 mRNA molecule then 0.05 mRNA molecule occurs per cell, and ii. if one cell only has 1 mRNA molecule for 5% of the time (or 2 for 2.5% of the time, etc) then it has on average 0.05 mRNA molecule.

(b) We know from the main text that 1 molecule in 1 fl corresponds to 1 nM in *E. coli*. So for mRNA: 0.05-5 nM and for protein 0.1-5000 nM.

(c) Higher mean concentrations of essential proteins is advantageous because this reduces the probability that a single cell lacks those proteins and would therefore be unviable.

(d) No, you cannot divide 9 molecules evenly. This example therefore illustrates that daughter cells, deriving from the same mother, can have different numbers of molecules. Those 5 identical mothers with each 10 molecules will not partition the molecules evenly over all daughters; so, not all daughters will receive 5 molecules each. This you can easily imagine; equal partitioning will only occur if at the time of division each cell half has 5 molecules; but molecules move randomly by diffusion through the cell. So, one cell half could have 6 molecules such that the other half has 4; or 3 and 7 or 2 and 8, etc. One might expect that a partitioning of 1 and 9 is perhaps less likely than 4 and 6. How should we describe the statistics of this process? It is the same as coin flipping stats; say we have 10 coins (molecules) and role (partition) each of them once to give rise to heads (right daughter) of tails (left daughter). This is described by a binomial distribution: the probability p_k for a daughter to obtain k molecules out of 10 equals $p_k = \left(\begin{array}{c} 10 \\ k \end{array} \right) \frac{1}{2}^k \frac{1}{2}^{10-k}$. This distribution is shown in figure 7.5.

3. **Spontaneous genetic variation and competing bacteria.** Genetic variation occurs spontaneously in bacterial populations because of copying errors are always made during replication. This leads to the spontaneous generation of mutants in bacterial populations that have slightly different properties. Some of those mutants are better adapted to the current environment, purely by chance. Those mutants will outgrow the resident population and can overtake the entire population such that their genotype becomes the dominant one at the expense of the resident genotype.

   We define the mutation probability as,

$$p \quad = \quad \text{probability for a single base pair change in DNA} \qquad (7.3)$$

   Let's apply some elementary probability reasoning to make some inferences about mutation and selection.

   (a) What is expected genome size with a single base pair change?
   Answer:

$$\frac{1}{p} = N_1 = \text{average genome size that has a single base pair change} \qquad (7.4)$$

(b) What is the expected number of mutations in a genome of length N?
Answer:

$$n_N = \frac{N}{N_1} = \text{expected number of mutations in a genome of length } N \tag{7.5}$$

(c) What should be the population size of bacteria with genome length N that contains all single mutations?
Answer: Number of possible single mutations equals N. The number of mutations in a genome of length N equals n_N. The population size that is required to contain all single mutations equals therefore $\frac{N}{n_N} = \frac{N}{\frac{N}{N_1}} = N_1 = \frac{1}{p}$. Which of course makes sense.

(d) Bionumbers tells us that a *E. coli* culture at $OD_{600} = 0.1$ contains 10^8 $\frac{cells}{ml}$. Take as the mutation rate $10^{-10}\frac{mutations}{nucleotide \times generation}$ (Lee, ..., Foster, PNAS, 2012). What is the expected OD_{600} of a 1 ml culture to contain all single mutations?
Answer: The number of cells required for all single mutations equals 10^{10}. So 100 ml is required of $OD_{600} = 0.1$ and 10 ml of a $OD_{600} = 1$ culture.

Given these numbers, we conclude that during serial-batch evolution experiments should therefore quickly sample all single mutations, as they involve growth experiments in shake flasks of 100 ml that are grown to $OD_{600} = 1$, diluted by a factor of 10, regrown in a new flask to $OD_{600} = 1$, which repeated for 100s of times.

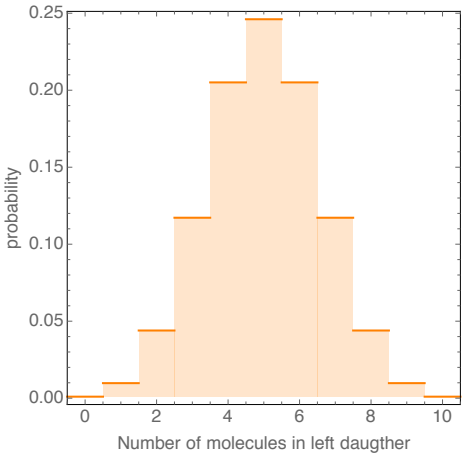


Figure 7.5: **Probability distribution for the number of molecules in a daughter cell when the mother contained 10 molecules**

7.8 Answers to exercises of Chapter 3

Exercise section 3.2.3

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

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

$$\begin{aligned} \frac{ds}{dt} &= -v_1 \\ \frac{dx}{dt} &= v_1 - v_2 \\ \frac{dp}{dt} &= v_2 \\ v_1 &= k_1^+ s - k_1^- x \\ v_2 &= k_2^+ x - k_2^- p \end{aligned}$$

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

$$\begin{aligned} \frac{dx}{dt} &= v_1 - v_2 \\ \frac{dp}{dt} &= v_2 \\ v_1 &= k_1^+ s - k_1^- x \\ v_2 &= k_2^+ x - k_2^- p \end{aligned}$$

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

$$\begin{aligned} \frac{da}{dt} &= -3v_1 \\ \frac{db}{dt} &= 2v_1 - v_2 \\ \frac{dc}{dt} &= v_1 - 2v_3 \\ \frac{dd}{dt} &= 2v_2 \\ \frac{de}{dt} &= 3v_3 \\ v_1 &= k_1^+ a^3 - k_1^- b^2 \cdot c \\ v_2 &= k_2^+ b - k_2^- d^2 \\ v_3 &= k_3^+ c^2 - k_3^- e^3 \end{aligned} \tag{7.6}$$

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

$$\begin{aligned}
\frac{dxy}{dt} &= -v_1 \\
\frac{dz}{dt} &= -v_1 + v_3 \\
\frac{dxyz}{dt} &= v_1 - v_2 \\
\frac{dx}{dt} &= v_2 \\
\frac{dyz}{dt} &= v_2 - v_3 \\
\frac{dy}{dt} &= v_3 \\
v_1 &= k_1^+ xy \cdot z - k_1^- xyz \\
v_2 &= k_2^+ xyz - k_2^- x \cdot yz \\
v_3 &= k_3^+ yz - k_3^- y \cdot z
\end{aligned} \tag{7.7}$$

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:
$E + S \rightleftharpoons ES$, $ES \rightleftharpoons E + P$

(b) $\frac{dx}{dt} = k_1^+ a \cdot x^2 - k_1^- x^3 - k_2^+ x + k_2^- b$
Answer:
$A + 2X \rightleftharpoons 3X$, $X \rightleftharpoons B$

(c) $\frac{dx}{dt} = k_1^+ a - k_1^- x + k_3 x^2 \cdot y$, $\frac{dy}{dt} = k_2 b - k_3 x^2 \cdot y$
Answer:
$A \rightleftharpoons X$, $B \rightarrow Y$, $2X + Y \rightarrow 3X$

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

Exercise section 3.6.1

Consider the following system,

$$\underline{S} \underset{k_1^- y}{\overset{k_1^+ s}{\rightleftharpoons}} Y \underset{k_2^- p}{\overset{k_2^+ y}{\rightleftharpoons}} \underline{P} \tag{7.8}$$

with the concentration of S and P fixed (hence, the underline).

a. 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.
Answer:

$$\frac{d}{dt}y = k_1^+ s - k_1^- y - k_2^+ y + k_2^- p = \underbrace{k_1^+ s + k_2^- p}_{k_1} - \underbrace{(k_1^- + k_2^+)}_{k_2} y \tag{7.9}$$

So, we can define new 'lumped' constants $k_1 = k_1^+ s + k_2^- p$ and $k_2 = k_1^- + k_2^+$ such that the differential equation for y because the same as used in the previous section in the main text. So, understanding the dynamics is the same problem for the two systems.

b. 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.
Answer: The steady state concentration of Y equals,

$$\frac{d}{dt}y = 0 \quad \Rightarrow y_s = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} \tag{7.10}$$

When the system is at thermodynamic equilibrium, i.e. when $v_1 = v_2 = 0$ then, the concentration of Y equals,

$$v_1 = 0 \Rightarrow \frac{y_e}{s_e} = \frac{k_1^+}{k_1^-} \equiv K_{eq,1} \tag{7.11}$$

$$v_2 = 0 \Rightarrow \frac{p_e}{y_e} = \frac{k_2^+}{k_2^-} \equiv K_{eq,2} \tag{7.12}$$

$$\Rightarrow \quad y_e = \frac{k_1^+ s_e}{k_1^-} = \frac{k_2^- p_e}{k_2^+} \tag{7.13}$$

So $y_s \neq y_e$, only for a particular choice of the parameters s and p.

c. Does mass always flow in the same direction?
Answer: No, the reactions are reversible; so one would expect if s is chosen as 0, and held fixed, that mass flows from right to left, and when p is fixed at 0 that mass flows from left to right. Below this will be rigorously shown.

d. 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?
Answer: $y_e = \frac{k_1^+ s_e}{k_1^-} = \frac{k_2^- p_e}{k_2^+}$

e. 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.
Answer: $\frac{p_e}{s_e} = \frac{y_e}{s_e}\frac{p_e}{y_e} = K_{eq,1}K_{eq,2} = \frac{k_1^+ k_2^+}{k_1^- k_2^-} = 40$.

f. What happens to the steady-state mass flow when you decrease this P/S ratio and when you increase P/S?
Answer: Let's first derive the steady state reaction rates,

$$
\begin{aligned}
v_1(y_s) &= v_2(y_s) \equiv J && (7.14)\\
J &= k_1^+ s - k_1^- \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} && (7.15)\\
&= \frac{k_1^+ s(k_1^- + k_2^+) - k_1^-(k_1^+ s + k_2^- p)}{k_1^- + k_2^+} && (7.16)\\
&= \frac{k_1^+ k_1^- s + k_1^+ k_2^+ s - k_1^+ k_1^- s - k_1^- k_2^- p}{k_1^- + k_2^+} && (7.17)\\
&= \frac{k_1^+ k_2^+ s - k_1^- k_2^- p}{k_1^- + k_2^+} && (7.18)\\
&= \frac{k_1^+ k_2^+ s}{k_1^- + k_2^+}\left(1 - \frac{k_1^- k_2^- p}{k_1^+ k_2^+ s}\right) && (7.19)\\
&= \frac{k_1^+ k_2^+ s}{k_1^- + k_2^+}\left(1 - \frac{p}{K_{eq,1}K_{eq,2}s}\right) && (7.20)
\end{aligned}
$$

(Note that $\frac{p_e}{s_e} = K_{eq,1}K_{eq,2}$.) So, if

i. $\frac{p}{s} < K_{eq,1}K_{eq,2}$ then the flux is positive, a steady state occurs, and mass flows from $S \to P$.

ii. $\frac{p}{s} = K_{eq,1}K_{eq,2}$ then the flux is zero and no net mass flow but thermodynamic equilibrium.

iii. $\frac{p}{s} > K_{eq,1}K_{eq,2}$ then the flux is negative, a steady state occurs, and mass flows from $P \to S$.

g. Try to write the steady-state flux in terms of P/S and the remaining parameters of the system. Figure 3.2 should be helpful while doing this exercise.
Answer: See previous equation.

Exercise section 3.7.1

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 $\underline{s} \rightleftharpoons x$. Show that the time to reach half the steady-state concentration is halved when all rate constants are doubled in value.
   Answer:
   See the Excel file: *equilibriumrelaxation.xlsx*.

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,

$$\frac{d}{a} = \frac{k_1^+ k_2^+ k_3^+}{k_1^- k_2^- k_3^-} \tag{7.21}$$

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,

$$\frac{d}{a} = \frac{k_1^+ k_3^+}{k_1^- k_3^-} \tag{7.22}$$

   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 section 3.9.1

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,

$$ab = a_T \frac{b/K_D}{1 + b/K_D} \tag{7.23}$$

   as

$$\frac{b}{K_D} = \frac{ab}{a_T - ab} = \frac{ab/a_T}{1 - ab/a_T} \tag{7.24}$$

   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.

$$\frac{TF}{K_D} = \frac{TFDNA}{TF_T - TFDNA} = \frac{TFDNA/TF_T}{1 - TFDNA/TFDNA_T} \tag{7.25}$$

$\frac{TF}{K_D}$ 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:

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

Define a K_D for the first reaction and the second reaction. Do you understand that those can indeed be different?
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,

$$\begin{aligned} b_T &= b + 2ab + a_2b = b + 2\frac{a \cdot b}{K_{D1}} + \frac{a \cdot ab}{K_{D2}} = b + 2\frac{a \cdot b}{K_{D1}} + \frac{a^2 \cdot b}{K_{D1}K_{D2}} \\ \Rightarrow b &= \frac{b_T}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{K_{D1}K_{D2}}} \end{aligned} \tag{7.26}$$

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,

$$a_2b = b_T \frac{\frac{a^2}{K_{D1}K_{D2}}}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{K_{D1}K_{D2}}} \tag{7.27}$$

And if A also interacts with A in AB besides its interaction with B then

$$a_2b = b_T \frac{\frac{a^2}{\alpha K_{D1}K_{D2}}}{1 + 2\frac{a}{K_{D1}} + \frac{a^2}{\alpha K_{D1}K_{D2}}} \tag{7.28}$$

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

$$\begin{aligned} A + B &\rightleftharpoons AB \\ A + AB &\rightleftharpoons A_2B \\ A + A_2B &\rightleftharpoons A_3B \end{aligned} \tag{7.29}$$

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),

$$b = \frac{b_T}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.30}$$

And we obtain

$$ab = b_T \frac{3\frac{a}{K_{D1}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.31}$$

$$a_2b = b_T \frac{3\frac{a^2}{\alpha K_{D1}K_{D2}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.32}$$

$$a_3b = b_T \frac{\frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.33}$$

(a) At what concentration of A is 50% of B in the A_3B complex?
Answer:
This requires solving

$$\frac{a_3b}{b_T} = \frac{\frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} = 0.5 \tag{7.34}$$

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

$$\frac{a_2b}{b_T} = \frac{\frac{a^2}{\alpha K_{D1}K_{D2}}}{1 + 3\frac{a}{K_{D1}} + 3\frac{a^2}{\alpha K_{D1}K_{D2}} + \frac{a^3}{\alpha\beta K_{D1}K_{D2}K_{D3}}} = 0.5 \tag{7.35}$$

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?

$$\frac{ab}{b_T} = \frac{3\frac{a^*}{K_{D1}}}{1 + 3\frac{a^*}{K_{D1}} + 3\frac{a^{*2}}{\alpha K_{D1}K_{D2}} + \frac{a^{*3}}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.36}$$

$$\frac{a_3b}{b_T} = \frac{\frac{a^{*3}}{\alpha\beta K_{D1}K_{D2}K_{D3}}}{1 + 3\frac{a^*}{K_{D1}} + 3\frac{a^{*2}}{\alpha K_{D1}K_{D2}} + \frac{a^{*3}}{\alpha\beta K_{D1}K_{D2}K_{D3}}} \tag{7.37}$$

Exercise section 3.9.2

1. **Steady state versus equilibrium state** Many biological systems attain steady states. The principles of steady states are therefore very important and you will study them in this exercise. Consider the following reactions and assume them to follow mass action kinetics

$$S \overset{1}{\rightleftharpoons} X \tag{7.38}$$

$$X \overset{2}{\rightleftharpoons} P \tag{7.39}$$

We will focus on X. The rate of reactions 1 and 2 are denoted respectively by v_1 and v_2. Here we use the convention that the concentration of molecules are written in small font and the name of the molecule in capitol font.

(a) Give the mass balance for the concentration of X.
Answer: $\frac{dx}{dt} = v_1 - v_2$

(b) When has X attained a steady state?
Answer: When $v_1 = v_2 \neq 0$.

(c) Can this happen when the concentrations of S and P are not fixed?
Answer: No, since the rates of the reactions are depend on the concentration of S and P, which will change when they are not held fixed, such that the rates becomes time dependent until $\frac{ds}{dt}$, $\frac{dx}{dt}$ and $\frac{dp}{dt}$ become zero. When this happens $v_1 = v_2 = 0$ because $\frac{ds}{dt} = -v_1$ and $\frac{dp}{dt} = v_2$, and this is not a steady state.

(d) What is the name of the state that the system attains when S and P are not fixed?
Answer: Thermodynamic equilibrium.

(e) Calculate the concentration of X in the final state when S and P are not fixed, assuming reversible mass-action kinetics.
Answer: In the final state $v_1 = v_2 = 0$ and $v_1 = k_1^+ s - k_1^- x$ and $v_2 = k_2^+ x - k_2^- p$. To solve for x we have to solve the following system of equations,

$$\begin{aligned} k_1^+ s_e - k_1^- x_e &= 0 \\ k_2^+ x_e - k_2^- p_e &= 0 \\ T = s_e + x_e + p_e & \end{aligned} \tag{7.40}$$

The last condition states that the total concentration of molecules remains fixed, the subscript indicates that we are considering the

equilibrium concentrations. Solving this gives,

$$\begin{aligned} s_e &= \frac{k_1^- k_2^-}{k_1^- k_2^- + k_1^+ k_2^- + k_1^+ k_2^+} T \\ x_e &= \frac{k_1^+ k_2^-}{k_1^- k_2^- + k_1^+ k_2^- + k_1^+ k_2^+} T \\ p_e &= \frac{k_1^+ k_2^+}{k_1^- k_2^- + k_1^+ k_2^- + k_1^+ k_2^+} T. \end{aligned} \tag{7.41}$$

(f) When S and P are not fixed, the concentration X can only become constant if the rates of reactions 1 and 2 are each equal to zero. Derive the two equations that relate the concentration ratio of $\frac{x}{s}$ of $\frac{p}{x}$ in this equilibrium state. Define those ratio's as equilibrium constants K_1 and K_2

Answer: $v_1 = 0 \;\Rightarrow\; \frac{x}{s} = \frac{k_1^+}{k_1^-} = K_1$ and $v_2 = 0 \;\Rightarrow\; \frac{p}{x} = \frac{k_2^+}{k_2^-} = K_2$.

(g) Rewrite the rate of reaction 1 and 2 in terms of the constant K_1 and K_2 and show that the rates are negative when $\frac{x}{s} > K_1$ and $\frac{p}{x} < K_2$ and positive when $\frac{x}{s} < K_1$ and $\frac{p}{x} < K_2$. What happens if a reaction rate changes sign?

Answer:

$$\begin{aligned} v_1 &= k_1^+ s - k_1^- x = k_1^+ s \left(1 - \frac{x}{s \times K_1}\right) \\ v_2 &= k_2^+ x - k_2^- p = k_2^+ x \left(1 - \frac{p}{x \times K_2}\right) \end{aligned} \tag{7.42}$$

Clearly these equations agree with the greater and smaller conditions. If a reaction rate changes change then the current product becomes the substrate and the current substrate becomes product, so the reaction start to work in the opposite direction.

(h) Fix S and P and solve the differential equation with $x(0) = x_0$ as initial state.

Answer: The following differential equation needs to be solved,

$$\frac{dx}{dt} = \underbrace{k_1^+ s + k_2^- p}_{a} - \underbrace{(k_1^- + k_2^+)}_{b} x, \quad x(0) = x_0 \tag{7.43}$$

which leads to $x(t) = x(0)e^{-bt} + \frac{a}{b}\left(1 - e^{-bt}\right)$ with $\frac{a}{b} = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+}$ as the steady state concentration of x.

(i) What is the concentration of X when time become very large?

Answer: It becomes equal the steady state concentration defined by $\frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+}$.

(j) Are the rates of the reactions necessarily unequal to zero in the final state or does it depend on the choice of the value of the fixed concentrations of S and P?
Answer: It depends on the choice of the concentrations of S and P. If those are chosen as the equilibrium concentration (question e) then the final state will be an equilibrium state.

(k) Express the concentration of X in the final state in terms of the system parameters.
Answer: $x_s = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+}$.

(l) Show that at the steady state, when S and P are fixed and the rate of X synthesis and degradation are equal, that the direction of mass flow – so from S to P or vice versa – depends on the ratio of $\frac{p}{s}$.
Answer: At the steady state, $v_1 = v_2 = J$ with J therefore as $J = k_1^+ s - k_1^- x_s = k_1^+ s - k_1^- \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} = \frac{k_1^+ k_2^+ s - k_1^- k_2^- p}{k_1^- + k_2^+} = \frac{k_1^+ k_2^+}{k_1^- + k_2^+} s \left(1 - \frac{k_1^- k_2^- p}{k_1^+ k_2^+ s}\right) = \frac{k_1^+ k_2^+}{k_1^- + k_2^+} s \left(1 - \frac{p}{s K_1 K_2}\right)$. This indicates that the flux is positive, when P is made out of S, when $\frac{p}{s} < K_1 K_2$ and if $\frac{p}{s} > K_1 K_2$ it is negative such that S is made out of P.

Exercise section 3.10

1. **Enzyme kinetics.** Enzymes are the workhorses of a cell. Essentially all reactions are catalysed by them. They speed up reactions by offering a favourable physicochemical environment in their catalytic site for the reaction to occur. Without the enzyme the reaction would also take place, as enzymes cannot change the equilibrium constant of a reaction, but the reaction rate would be orders of magnitude slower. So one way to envision cellular metabolism is that a cell selects reactions that are favourable for its fitness, by expressing the associated enzymes that can catalyse those reactions, out of all possible reactions. In this exercise, we will think about the principles of enzyme catalyses. Consider the following enzyme-catalysed conversion,

$$\underline{S} + E \underset{v_1^-}{\overset{v_1^+}{\rightleftharpoons}} ES \xrightarrow{v_2} E + \underline{P} \tag{7.44}$$

The underlines indicate that the concentrations of S and P are held fixed.

(a) Give the mass balances for the concentrations of the variable species in the model in terms of rates of the reactions.
Answer:

$$\frac{de}{dt} = -v_1^+ + v_1^- + v_2 \tag{7.45}$$

$$\frac{des}{dt} = v_1^+ - v_1^- - v_2 \tag{7.46}$$

(b) An enzyme is a catalyst that is not spent during the reaction. This has one particular consequence for the concentration of enzyme in the system. What is this consequence?
Answer: The total enzyme concentration remains fixed, $e_T = e + es$.

(c) Why is $\frac{de}{dt} + \frac{des}{dt} = 0$?
Answer: Because the total enzyme concentration remains fixed. $0 = \frac{de_T}{dt} = \frac{d(e+es)}{dt} = \frac{de}{dt} + \frac{des}{dt}$.

(d) Express the rate of the reactions in terms of mass action kinetics.
Answer:

$$\frac{de}{dt} = -v_1^+ + v_1^- + v_2 = -k_1^+ \cdot e \cdot s + k_1^- es + k_2 es \quad (7.47)$$

$$\frac{des}{dt} = v_1^+ - v_1^- - v_2 = k_1^+ \cdot e \cdot s - k_1^- es - k_2 es \quad (7.48)$$

(e) Give the units of all the terms appearing in the mass balance equation with mass action kinetics.
Answer: k_1^+ has as unit $\frac{1}{\text{time} \times \text{concentration}}$ and k_1^- and k_2 have each as unit $\frac{1}{\text{time}}$.

(f) Solve for es at steady state.
Answer:

$$0 = \frac{des}{dt} = k_1^+ \cdot (e_T - es) \cdot s - k_1^- es - k_2 es = k_1^+ e_T s + (-k_2 - k_1^+ s - k_1^-) es \quad (7.49)$$

$$\Rightarrow es = \frac{k_1^+ e_T s}{k_2 + k_1^- + k_1^+ s} \quad (7.50)$$

(g) The steady-state rate of the enzyme is defined as $v = k_2 es$. Express this enzyme rate in terms of s and identify the combination of constants that you have to make in order to write this enzyme rate equation into its more familiar form,

$$v = V_{max} \frac{s}{K_M + s} \quad (7.51)$$

Answer:

$$v = k_2 es = k_2 e_T \frac{k_1^+ s}{k_2 + k_1^- + k_1^+ s} = \underbrace{k_2 e_T}_{V_{max}} \frac{s}{\underbrace{\frac{k_2 + k_1^-}{k_1^+}}_{K_M} + s} = V_{max} \frac{s}{K_M + s} \quad (7.52)$$

(h) Studying $\frac{1}{v}$ leads to an intuitive understanding of how an enzyme works (we set e_T to 1),

$$\frac{1}{v} = \frac{k_2}{k_2 k_1^+ s} + \frac{k_1^-}{k_2 k_1^+ s} + \frac{k_1^+ s}{k_2 k_1^+ s} = \frac{1}{k_1^+ s}\left(1 + \frac{k_1^-}{k_2}\right) + \frac{1}{k_2} \quad (7.53)$$

$\frac{1}{v}$ now corresponds to the waiting time for 1 enzyme to convert one molecule of S into P. If $k_2 >> k_1^-$ then $ES \rightarrow E + P$ nearly always occurs, rather than $ES \rightarrow E+S$, and the waiting time for the reaction becomes,

$$\frac{1}{v} = \underbrace{\frac{1}{k_1^+ s}}_{\substack{\text{time to} \\ \text{bind}}} + \underbrace{\frac{1}{k_2}}_{\substack{\text{catalysis} \\ \text{time}}} \tag{7.54}$$

Which makes intuitive sense, both reactions have to occur before a P molecule appears. Why does the $\left(1 + \frac{k_1^-}{k_2}\right)$ factor appear when k_2 is not much larger than k_1^-?

Answer: The factor $\left(1 + \frac{k_1^-}{k_2}\right)$ equals the number of re-bindings of S to E before the catalysis reaction $ES \rightarrow E + P$ occurs. The binding time $\frac{1}{k_1^+ s}$ therefore need to be multiplied by this factor to get the waiting time for the $ES \rightarrow E + P$ to occur which takes itself $\frac{1}{k_2}$ time to complete.

2. **The chemostat for culturing of cells.** The chemostat is a bioreactor set up that allows for the continuous steady-state cultivation of cells, it keeps the cells at a steady-state growth rate. The chemostat state is sometimes called a continuous culture. The concept is that medium flows into the reactor from a medium vessel at a fixed flow rate F, expressed in $\frac{liter}{hr}$. The volume of the culture V is kept fixed by flowing out medium, including cells, from the bioreactor into an exhaust vessel at the same rate. The dilution rate D is now defined as $D = \frac{F}{V}$. Medium leaves and enters the vessel at this rate. You can therefore think of D as a rate constant. The medium vessel contains the growth substrate, which limits growth, e.g. glucose, at a concentration s_m. Since the cells in the bioreactor consume this substrate the concentration in the bioreactor, s will be smaller than the concentration in the medium vessel: $s < s_m$. The cells have a specific growth rate μ that depends on the concentration of the growth-limiting substrate in the reactor. We model it as,

$$\mu = \mu_{max} \frac{s}{K_s + s} \tag{7.55}$$

with $1/K_s$ as the affinity of the organism for the growth-limiting substrate. This equation is often called the Monod equation with μ_{max} as the maximal growth rate and K_s as the Monod constant. A model of the chemostat contains minimally two variables: the concentration of substrate (in mol/l) and cells (in $gram/liter$) in the bioreactor. For those

variables we can write the following mass balances,

$$\frac{ds}{dt} = D(s_m - s) - \frac{1}{Y_{x/s}}\mu x \tag{7.56}$$

$$\frac{dx}{dt} = (\mu - D)x \tag{7.57}$$

with $Y_{x/s}$ as the yield of biomass on substrate in $\frac{gram\ biomass}{mol\ substrate}$.

(a) Explain the meaning of all the terms in the balances: Ds_m, $-Ds$, $-\frac{1}{Y_{xs}}\mu x$, μx and of $-Dx$.
Answer:

$$\begin{aligned} Ds_m &= \text{Inflow rate of substrate from the medium vessel into the bioreactor in } \frac{M}{hr} \\ -Ds &= \text{Outflow rate of substrate from bioreactor into the exhaust vessel in } \frac{M}{hr} \\ -\frac{1}{Y_{x/s}}\mu x &= \text{Substrate consumption rate in } \frac{M}{hr} \\ \mu x &= \text{Growth rate of the culture in } \frac{gram}{liter \times hr} \\ -Dx &= \text{Outflow rate of cells out of the bioreactior in } \frac{gram}{liter \times hr} \end{aligned}$$

(b) At steady state the specific growth rate μ equals the dilution rate D. Why does this make sense?
Answer: Since cells do not enter the bioreactor, only growth can compensate for cell loss by outflow. Only if the growth rate equals the outflow rate will the concentration of cells remain fixed, which is the requirement for steady state.

(c) Express the steady state concentration of biomass x_s and of growth substrates$_s$ in terms of model parameters and plot their dependences on the dilution rate.
Answer:

$$D = \mu = \mu_{max}\frac{s_s}{K_s + s_s} \quad \text{Valid at steady state} \tag{7.58}$$

$$\Rightarrow s_s = \frac{K_s}{\mu_{max} - D} \tag{7.59}$$

$$0 = D(s_m - s_s) - \frac{1}{Y_{x/s}}Dx_s \tag{7.60}$$

$$\Rightarrow x_s = Y_{x/s}(s_m - s_s) = Y_{x/s}(s_m - \frac{K_s}{\mu_{max} - D}) \tag{7.61}$$

(d) What is the maximal possible value of D at which cells still occur in the chemostat?
Answer: $D = \mu_{max}$

(e) What is concentration of growth-limiting substrate in the bioreactor when the maximal D is reached?
Answer: $s_m \approx s_s$.
Add experimental figure of s_s and x_s as function of D.

3. **Persister cells in bacterial populations** In the last few years, it has become clear that populations of many bacterial species consist of two subpopulations. One that is growing and another that is non-growing – 'dormant'. The non-growing cells are more stress resistant than the growing cells. Antibiotic resistant bacteria are often dormant cells and since they persist antibiotic conditions, or stress conditions, they are generally referred to as 'persister cells'. A single growing cell can switch to become persister and a persister can switch to become a growing cell. Clearly, the persister cells do not grow. We therefore have the following three processes: i. growth of a cell in the growing state, ii. a cell in the growing states that switches to the persister state, and iii. a persister state cell that turns into a growing cell state. We use the following notation: $\mu =$ growth rate, $k_p =$ the rate constant for switching from the growing to the persister state, $k_g =$ the rate constant for switching from the persister state to the growing state, the concentration of growing and persister cells equals g and p, respectively.

(a) What are the mass balances for the concentrations of growing and persister cells?
Answer:

$$\frac{d}{dt}g = \mu g - k_p g + k_g p \tag{7.62}$$

$$\frac{d}{dt}p = k_p g - k_g p \tag{7.63}$$

(b) Why does this system never settle to a steady-state concentration of the cell states?
Answer: Because the growing cells continue to grow and do not settle to a steady state at which cell death or cell loss balances with cell growth.

(c) Do you think the fraction of persister and growing cells becomes fixed over time?
Answer: Yes, I expect this to happen. Why will become clear in the next questions.

(d) The balance for the fraction of persister cells, ϕ, is given by,

$$\frac{d\phi}{dt} = \frac{d}{dt}\underbrace{\left(\frac{p(t)}{g(t)+p(t)}\right)}_{\phi} = \underbrace{k_p + \mu\phi^2}_{\text{Synthesis rate}} - \underbrace{(k_p + k_g + \mu)\phi}_{\text{Degradation rate}} \tag{7.64}$$

The steady state fraction equals,

$$\phi_s = \frac{k_d + k_g + \mu - \sqrt{(-k_d - k_g - \mu)^2 - 4k_d\mu}}{2\mu} \tag{7.65}$$

This fraction was obtained by setting the previous equation to zero and solving for ϕ. Derive this equation yourself.
Answer:

$$\frac{d\phi}{dt} = 0 = \underbrace{\mu}_{a}\phi_s^2 \underbrace{-(k_p + k_g + \mu)}_{b}\phi_s + \underbrace{k_p}_{c} \tag{7.66}$$

$$\phi_{s,1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{7.67}$$

Only one of those two solutions, the with the '-' in front of the square root, gives ϕ value between 0 and 1.

(e) Calculate the steady state fraction. Use realistic parameters: $\mu = 1\ hr^{-1}$, $k_g = \frac{1}{10}\ hr^{-1}$, and $k_d = \frac{1}{100}\ hr^{-1}$.
Answer: This fraction equals 0.9%.

(f) Show that the persister fraction indeed moves towards a stable steady state by plotting the synthesis rate and degradation rate as function of ϕ. Use the same parameters as in the previous question.
Answer: You indeed find one intersection of the synthesis and degradation rate curves at $\phi_s = 0.009$. Left from this steady state point the synthesis exceeds the degradation rate, so ϕ increases to ϕ_s. Right from this steady state point the degradation rate exceeds the synthesis rate, so ϕ drops to ϕ_s.

7.9 Answers to exercises of Chapter 4

7.10 Answer for exercise section 4.2

1. In signal transduction, proteins are often activated by phosphorylation, this causes them to change shape and have different binding affinities and enzymatic activities. They are then inactivated by dephosphorylation. Kinase phosphorylate proteins and phosphatase dephosphorylate them.

   (a) What is the reaction catalysed by a kinase?
   Answer: $X + ATP \xrightarrow{kinase} XP + ADP$ with X as the target protein that modified by the kinase and phosphatase. This reaction occurs at rate v_{kinase}

   (b) What is the reaction catalysed by a phosphatase?
   Answer: $XP \xrightarrow{phosphatase} X + P$ with as inorganic phosphate, PO_4^{3-} and occurring at rate $v_{phosphatase}$.

(c) If we keep the concentrations of ATP, ADP, Pi, kinase and phosphatase constant which differential equations would you need to model the activation of protein by a kinase and its inactivation by a phosphatase?
Answer: then we would end up with the following differential equations:

$$\begin{aligned} \frac{dx}{dt} &= v_{phosphatase} - v_{kinase} \\ \frac{dxp}{dt} &= v_{kinase} - v_{phosphatase} \end{aligned} \tag{7.68}$$

Note that $\frac{dx}{dt} + \frac{dxp}{dt} = 0$, which means that total concentration of X, so $x + xp$ stays constant!

(d) What would be suitable enzyme kinetics for the kinase and the phosphatase?
Answer: many options exist – which you may know when you are familiar with enzyme kinetics of multiple substrate and product reactions – but the simplest ones would incorporate the concentrations of molecules that are held fixed into enzyme parameters and then one would, for instance, obtain,

$$v_{kinase} = V_k \frac{x}{K_k + x} \quad \text{and} \quad v_{phosphatase} = V_p \frac{xp}{K_p + xp}.$$

See, for instance, the reference Goldbeter & Koshland [14].

(e) Explain why one differential equation is sufficient for the dynamic description of this process.
Answer: Since $x + xp$ is constant, say equal to x_{tot}, such as that $x = x_{tot} - xp$ then

$$v_{kinase} = V_k \frac{x_{tot} - xp}{K_k + x_{tot} - xp} \quad \text{and} \quad v_{phosphatase} = V_p \frac{xp}{K_p + xp},$$

such that both rates only depend on one variable, the concentration of XP. Therefore, we only need one differential equation,

$$\frac{dxp}{dt} = v_{kinase} - v_{phosphatase} = V_k \frac{x_{tot} - xp}{K_k + x_{tot} - xp} - V_p \frac{xp}{K_p + xp}.$$

(f) Do you think that this system reaches a steady state or an equilibrium state?
Answer: The rates of phosphatase and kinase are assumed irreversible (these cannot become negative after all) and therefore they cannot equal 0 when for nonzero concentrations of their reactants. So an equilibrium state where both $v_{kinase} = 0$ and $v_{phosphatase} = 0$ is impossible. Thus only a steady state is possible.

(g) Is this state stable? What if is not?
Answer: At the steady state, the concentration of XP is such that $v_{kinase} = v_{phosphatase} \neq 0$, lets call this concentration, xp_s. The steady state with concentration xp is stable: i. if for $xp < xp_s$, $\frac{dxp}{dt} > 0$ such that the concentration of xp rises with time as long as $xp < xp_s$ and ii. if for $xp > xp_s$, $\frac{dxp}{dt} < 0$ such that the concentration of xp decreases with time as long as if $xp > xp_s$. Otherwise it is not and the concentration of xp moves away from xp_s. When you sketch v_{kinase} and $v_{phophatase}$ as function of xp you can deduce that the steady state is stable.

7.11 Answers for exercise section 4.6

1. **Dimerisation kinetics of a membrane receptor.**

   (a) Since every dimer is composed out of two monomers and a monomer equals one receptor protein, the total number of receptor proteins equals the number of dimers times 2 plus the number of monomers.

   (b) The mass balance for the dimers equals $\frac{d}{dt}m_2 = v_a - v_d$. We have the following relation $m_T = 2 \times m_2 + m$ and therefore $0 = 2\frac{d}{dt}m_2 + \frac{d}{dt}m$ such that $\frac{d}{dt}m_2 = -\frac{1}{2}\frac{d}{dt}m$.

   (c) Apparently the life time of the dimer is quite short; so, their interaction is rather weak such that it quickly falls apart.

2. **Kinase and phosphatase kinetics and the steepness of the input/output relationship of a kinase-phosphatase pair.** Play with the file "kinase and phosphatase.cdf".

7.12 Answers for exercise section 4.7

1. **How negative feedback in a signaling cascade causes insensitivity to cancer drugs.**

   (a) See figure 7.6 for a visualisation of the MAPK network. Note that the paths in the network scheme from EGFR to ERK and the circular path from ERK to ERK appear in the scheme as well as in the response equation.

   (b) The response equation without the feedback corresponds to the strength of the path from EGFR to ERK; so, $d\ln ERK = r^{ERK}_{MEK} \times r^{MEK}_{RAF} \times r^{RAF}_{EGFR} \times d\ln EGFR$. Since the feedback is negative the influence of the feedback is that it reduces the response of ERK with respect to a change in EGFR. In terms of y and u, we obtain in absence of the feedback: $y = A \times u$

(c) If the feedback is strong, the response of ERK (y) becomes independent of the cascade response $r^{ERK}_{MEK} \times r^{MEK}_{RAF} \times r^{RAF}_{EGFR}$ (A) because then $f \times A \gg 1$ and $y \approx \frac{1}{f} \times u$. So, the feedback makes the system in sensitive to changes in A; hence, drugs intervening with processes between RAF and ERK have no longer an effect on the response of ERK with respect to $EGFR$.

(d) No, the feedback does not protect the system against drugs intervening above RAF, so to those affecting u, because the response equation still depends on u when the feedback is strong; because then $y = \frac{1}{f} \times u$.

(e) The thermostat works properly when y (room temperature disturbances) is zero for most values of u (outside temperature disturbances). So, when $u \approx f \times y$ then $y \approx 0$; see equation $y = A \times (u - f \times y)$. Now note that when $u \approx f \times y$, $y \approx \frac{1}{f} \times u$. So, this is exactly the equation that holds when f is large such that $f \times A \gg 1$! Hence, robustness occurs when the feedback is strong; as it dampens the effect of an outside disturbance u by a factor $\frac{1}{f}$. So, feedback causes robustness and then the thermostat works.

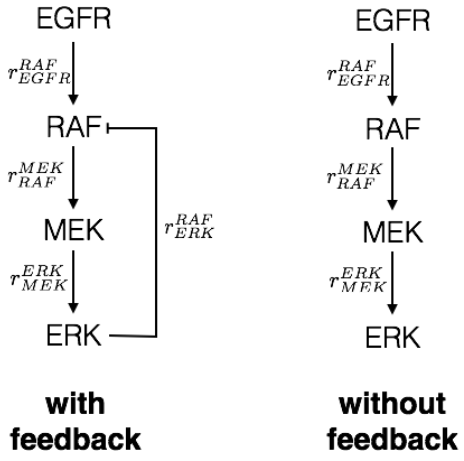


Figure 7.6: **Depiction of the MAPK interaction network with and without the feedback of ERK onto RAF.**

2. **Bacteria communicating with each other and deciding together.** Bacteria are much smarter than we often think. For instance, they use a mechanism called quorum sensing to keep track of the number of bacteria in the population. In this way, bacteria can 'measure' their population size and initiate a response when this number passes a threshold. Some bacteria even use this mechanism to measure the abundance of competing or harmful bacteria. The principle of quorum sensing is shown in figure

4.7. (We note that not all bacteria use sensors to measure quorum-sensing compounds, you can also think of the sensor as a transcription factor inside the bacterium, and the green compound as freely diffusing over membranes.)

Say that the production rate of green compound per cell equals, k_p, in $\frac{\text{nM}}{\text{minute}\times\text{cell}}$, that we have N cells, and that the diffusion rate equals k_d in $\frac{1}{\text{minute}}$.

(a) Write down the mass balance of the extracellular concentration of the green compound, denoted by g, and express this equation in terms of k_p, N, and k_d.
Answer: $\frac{d}{dt}g = k_p \times N - k_d \times g$

(b) Express the steady state concentration of g, called g_s, in terms of k_p, N, and k_d.
Answer: $g_s = \frac{k_p}{k_d} \times N$

(c) Suppose that the bacteria grow exponentially, such that their number increases as $N(t) = N(0) \times e^{\mu \times t}$, with μ as the growth rate and t as time. We also know that when g_s exceeds the threshold concentration g^* that the cells respond with gene expression. Express the time at which the cells initiate a response in terms of k_p, k_d, $N(0)$, and μ. Assume that at every moment in time $g \approx g_s$.
Answer:

$$\begin{aligned} g^* &= \frac{k_p}{k_d} \times N(0) e^{\mu \times t^*} \quad \Rightarrow \\ e^{\mu \times t^*} &= \frac{g^* \times k_d}{k_p \times N(0)} \quad \Rightarrow \\ t^* &= \frac{1}{\mu} \times \ln\left(\frac{g^* \times k_d}{k_p \times N(0)}\right) \end{aligned} \tag{7.69}$$

(d) Calculate the time that it takes before the threshold concentration is reached when we start with 1 cell, $k_p = 5\ \frac{\text{nM}}{\text{minute}\times\text{cell}}$, $k_d = 100\ \frac{1}{\text{minute}}$, $\mu = 1\ \text{hour}^{-1}$ and $g^* = 100$ nM.
Answer:

$$t^* = \frac{1}{\mu} \times \ln\left(\frac{g^* \times k_d}{k_p \times N(0)}\right) = \frac{1}{1} \times \ln\left(\frac{100 \times 100}{5 \times 1}\right) = 6.9 \text{ hour} \tag{7.70}$$

(e) What is the threshold number of cells?
Answer: $N^* = N(t^*) = 1 \times e^{1 \times 6.9} = 2000$ cells

3. **Does a cell need to be covered completely with sensors for maximal sensing or witg transporters for a maximal uptake rate of nutrients?** We consider a cell with radius a that has N transporters or sensors on its surface, each with radius s. The transporters transport nutrient molecules that find the cell by diffusion whereas the sensors bind

molecules and initiate signaling. The nutrient or signal concentration far away from the cell is c_∞. The nutrients and signals diffuse with a diffusion coefficient D. The relation between the uptake flux, J, (or sensor binding rate) and those parameters is (Berg & Purcell, Biophysical Journal, 1977),

$$J = \underbrace{4\pi D c_\infty a}_{J_{max}} \frac{Ns}{Ns + a\pi} \quad \Rightarrow \quad \frac{J}{J_{max}} = \frac{Ns}{Ns + a\pi} \tag{7.71}$$

with J_{max} as the maximal uptake rate. N can be considered as the only variable in this equation. Consider a transporter of 5 nm and the cell of 1 μm in radius. Those are realistic numbers for bacteria.

(a) How many receptor are required for half maximal uptake?
Answer: $1/2 = \frac{N\times 5\times 10^{-9}}{N\times 5\times 10^{-9}+10^{-6}\pi}$ which leads to a requirement of 628 transporters.

(b) Consider the transporters as disks and the cell as a sphere. What is percentage of area covered by transporters for half-maximal uptake? Write down your expectation first.
Answer: Area cell: $A_c = 4\pi a^2$ and of the receptor $A_r = \pi s^2$. Percentage area covered $\frac{628\times\pi(5\times 10^{-9})^2}{4\pi(10^{-6})^2} \times 100\% = 0.4\%$ Wow! This is very small! Very far from the expected 50%.

(c) Plot the $\frac{J}{J_{max}}$ as function of the cell-surface area that is occupied by transporters.
Answer: Fraction of covered cell-surface area with receptors: $\phi = \frac{N\pi s^2}{4\pi a^2} = \frac{Ns^2}{4a^2} \Rightarrow N = \frac{4\phi a^2}{s^2}$; such that $\frac{J}{J_{max}} = \frac{Ns}{Ns+a\pi} = \frac{\frac{4\phi a^2}{s^2}s}{\frac{4\phi a^2}{s^2}s+a\pi} = \frac{4a\phi}{4a\phi+\pi s} = \frac{\phi}{\phi+\frac{\pi s}{4a}} \approx \frac{\phi}{\phi+\frac{s}{a}}$. With $s = 5\times 10^{-9}m$ and $a = 10^{-6}$, 96% of maximal uptake is reached when 10% of the cell surface is covered with transporter. The explanation is that diffusion resembles a random walk and a molecule that collides with the cell will do so many times, effectively 'scanning' the surface for transporters. The uptake 'zone' of a transporter is therefore much great than its own dimensions.

These were very surprising results and also exact results, no approximations were made in the derivation of the equations mentioned above that are not biologically realistic.

7.13 Answers to exercises of Chapter 5

1. **Induction and repression of gene activity.**

   (a) See figure 7.7 for a plot of the experimental data.

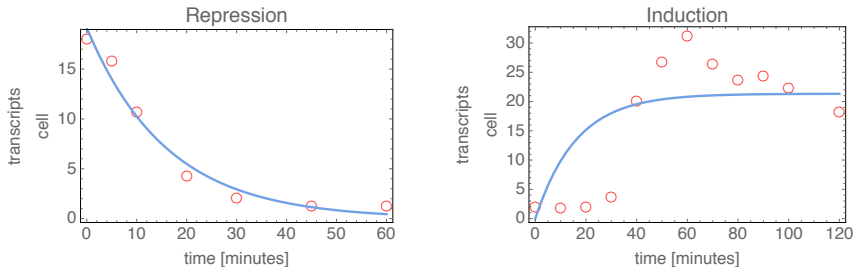


Figure 7.7: **Fitting induction and repression models to experimental data on single-cell transcripts numbers.** Left: Repression data fitted to $x(t) = x(0) \times e^{-k_d \times t}$ which gives as best fit values $x(0) = 19.1 \ \frac{\text{transcripts}}{\text{cell}}$ and $k_d = 0.06 \ \frac{1}{\text{min}}$. Right: Induction data fitted to $x(t) = x_s(1 - e^{-k_d \times t})$ with $x_s = \frac{k_s}{k_d}$ given rise to $k_s = 1.32 \ \frac{\text{transcripts}}{\text{min}}$. Clearly, the hypothesis underlying the induction model are insufficient for a satisfactory explanation of the experimental data. The induction model should incorporate an transduction induction delay and an overshoot.

(b) Since the transcription being induced causes the production of a metabolic enzyme, it is likely that the transcription activity is inhibited when sufficient metabolic enzyme has been synthesised. However, at the time of transcription inhibition new protein is still being made of the existing mRNA and RNA polymerases are still on the transcribing gene; so, the effect of the inhibition on transcript levels is delayed.

(c) The fitted value of the k_d equals $0.06 \ \frac{1}{min}$. The life time of the mRNA equals $\frac{1}{k_d} = 16 \ min$. So, the induction delay discussed previously is expected to be roughly 16 min, which is indeed fairly reasonable given the data.

It is hard to decide from the data, but a small repression delay could be visible from the data; the fit of the exponential curve to the data is also better when the first data point is removed. The fit is also slight improved when a background level of mRNA is assumed; so, $x(t) = x_b + x_0 e^{-k_d \times t}$.

(d) The transcription rate constant k_s has as unit $\frac{\text{transcript concentration}}{\text{min}}$ (note that if the volume of the cell remains fixed, the transcription rate constant can also be expressed in $\frac{\text{transcript}}{\text{min}}$). The assumed model is insufficient for explaining the data; a model that incorporates a transcription delay and also for an overshoot would fit the data much better.

2. **Dynamics of feedforward loop gene networks.**

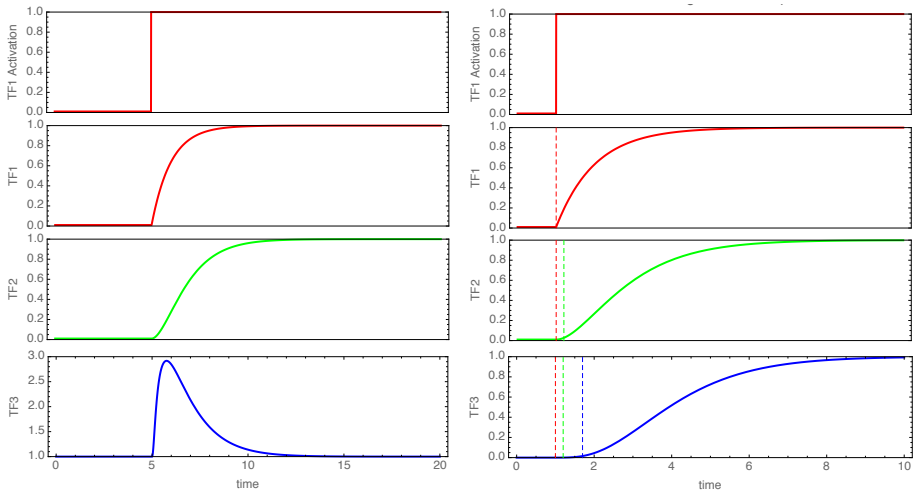


Figure 7.8: **Dynamics of FFL systems with either an activating or inhibiting feedforward loop** Left: An inhibiting FFL system gives rise to a peaked activation of TF_3 Right: An activating FFL system gives rise to a delated activation of TF_3. The dotted lines indicate the response times of the different TF's.

- The FFL system with the inhibiting influence of TF_2 on the synthesis of TF_3 shows the peaked activation of TF_3; because the TF_1-activation of TF_3 synthesis is immediate, it acts sooner than the inhibition of TF_3-synthesis by TF_2, which is delayed as it depends first on the increase in the concentration of TF_1 and only then TF_2 can increase. So, TF_2 and TF_3 rise together due to the rise in TF_1 and when TF_2 is high enough it reduces TF_3.
- The FFL system with the activating influence of TF_2 on the synthesis of TF_3 shows the delayed activation of TF_3; TF_3 only rises when TF_1 and TF_2 are both high, because the synthesis of TF_3 depends on the product of their concentrations.

3. **The *lac* operon of *Escherichia coli*.** The *lac* operon is active when glucose is low and lactose is high. Therefore, in the presence of glucose and lactose, the consumption of lactose starts when glucose has been consumed. So, *E. coli* prefers glucose over lactose.

4. **Transcription factor binding to DNA** Transcription factors bind to DNA to influence the rate of mRNA synthesis. We consider the following reaction,

$$F + D \underset{\text{dissociation}}{\overset{\text{association}}{\rightleftharpoons}} DF, \tag{7.72}$$

with F as the transcription factor and D as the DNA binding site. Concentrations are in small font and names of molecules in capitol font.

(a) Why are the total concentrations of F, which is equal to $f_T = f + df$, with subscript 'T' for total, and of D, equal to $d_T = d + df$, constant when the binding reaction occurs in a test tube?
Answer: Since no new molecules are added and removed from the test tube, those total concentrations remain fixed.

(b) If this reaction can be described in terms of mass action kinetics then what would be the rate equation for this reaction and what be the units of all the terms occurring in it?
Answer: The rate equation would be $v = k_a \cdot f \cdot d - k_d \cdot df$ with k_a as the association rate constant, with unit $\frac{1}{\text{concentration}\cdot\text{time}}$ and the k_d as the dissociation rate constant with unit $\frac{1}{\text{time}}$.

(c) Consider the situation where we keep the concentration of the transcription factor fixed. Give the mass balances for d and df.
Answer:

$$\begin{aligned} \frac{d}{dt}d &= -v \\ \frac{d}{dt}df &= v \\ v &= k_a \cdot f_T \cdot d - k_d \cdot df \end{aligned} \tag{7.73}$$

Since $d_T = d + df$ we can eliminate one of those mass balances and, for instance, focus only on,

$$\frac{d}{dt}df = k_a \cdot f_T \cdot (d_T - df) - k_d \cdot df \tag{7.74}$$

(d) Explain why the following relation holds: $\frac{d}{dt}d = -\frac{d}{dt}df$.
Answer: Since the sum $d + df$ remains fixed, any change in d is also accompanied by an opposing change in df.

(e) Show that at thermodynamic equilibrium the following relation holds,

$$df = d_T \frac{f}{\frac{k_d}{k_a} + f} \tag{7.75}$$

Answer: At thermodynamic equilibrium: $v = k_a \cdot f_T \cdot (d_T - df) - k_d \cdot df = 0$, solving for df gives the required equation.

(f) The ratio $\frac{k_d}{k_a}$ is called the dissociation constant and often written as K_D. What is its unit?
Answer: Its unit is concentration since k_a is in $\frac{1}{\text{concentration}\times\text{time}}$ and k_d in $\frac{1}{\text{time}}$.

(g) What is the unit of $\frac{f}{K_D+f}$?
Answer: It is dimensionless, all its terms have concentration as unit.

(h) Say the volume of a bacterial cell is 1 fl and 15 molecules of f occur in this cell. What is the concentration of f in nM?
Answer: $\frac{15 \text{ molecules}}{10^{-15}\text{l}} \frac{1}{6\times10^{23}} \frac{\text{mol}}{\text{molecules}} = \frac{15 \text{ mol}}{6\times10^{8} \text{ l}} = 25\ nM.$

(i) If one DNA site exists in this bacterial cell then what would be its concentration in nM.
Answer: $0.25/15\ nM = 1.67\ nM$

(j) Plot the occupancy fraction of the DNA binding site, defined as $\frac{df}{d_T}$, as function of f. Set K_D=2 nM. When is $df = \frac{d_T}{2}$?
Answer: This is a hyperbolic curve that ranges on the y-axis from 0 to 10 and equals 1/2 when $df = \frac{d_T}{2}$.

(k) If f is maximally 15 nM then what is the maximal occupancy fraction?
Answer: $\frac{15}{2+15} = 0.88$.

(l) If the transcription rate equals $v = k_t df$ and k_t equals $5 \frac{\text{nM mRNA}}{\text{nM DNA site}\times\text{min}}$ then what is the rate of mRNA synthesis when f equals 0.1 nM, 1 nM, or 10 nM. What is the maximal rate?
Answer: $v = 5 d_T \frac{f}{2+f} = 5 \times 1.67 \times \frac{f}{2+f}$ in $\frac{\text{nM mRNA}}{\text{min}}$ the rates becomes 0.4, 2.8, and 7.0 $\frac{\text{nM mRNA}}{\text{min}}$.

(m) If the lifetime of mRNA equals 10 min, which is a realistic number for bacteria, what is the steady state concentration of mRNA when $f = 1\ nM$?
Answer: We have the following mass balance with m as the concentration of mRNA and k_{deg} as the degradation rate constant of mRNA.

$$\frac{dm}{dt} = v - k_{deg} m = k_t \times d_T \times \frac{f}{K_D + f} - \frac{1}{\tau} \times m \tag{7.76}$$

with the life time of the mRNA as τ. The steady state concentration mRNA equals,

$$m = \tau \times k_t \times d_T \times \frac{f}{K_D + f} = 10 \times 5 \times 1.67 \times \frac{1}{2+1} = 27.8\ nM. \tag{7.77}$$

5. **Sliding of transcription factors along DNA shortens the promoter search time.** In bacteria, transcription factors find their target DNA sites, from which they regulate transcription, via shortly sliding in 1-D along the DNA, starting from a random DNA site that they encountered after a 3-D diffusive search for DNA in the cytoplasm. If during a single slide the target site is not found the transcription factor falls off and the process starts again. The search time for the target DNA site by a single transcription factor molecule is given by the following equation

$$\tau_s = \left(\frac{V}{4\pi D L} + \frac{l_s^2}{2D_1} \right) \frac{L}{l_s}, \tag{7.78}$$

with V as the cell volume, D as the cytosolic diffusion coefficient, L as the DNA length, l_s as the sliding length and D_1 as the sliding diffusion coefficient. The search time for the target site without sliding equals

$$\tau_{ws} = \frac{V}{4\pi Da}, \tag{7.79}$$

with a as the reaction radius of the target site on the DNA. In this question consider the following parameters: $D = D_1 = 5\ \mu m^2/s$, $a = 5\ nm$, $L = 1.5 \times 10^3\ \mu m$, $l_s = 30\ nm$ and $V = (2\ \mu m)^3$.

(a) What is the length of DNA in mm? Note that this length is a realistic value for *E. coli*). What is the length of an *E. coli* cell? How many times should the DNA be folded to place it in a cytosolic compartment of half the length of an *E. coli* cell? (Can you imagine now how weird it is that during DNA replication the two genome copies have to unwind and separate? This is a longstanding issue in *E. coli* and currently believed to occur spontaneously due to entropic forces.)
Answer: The length of DNA is 1.5 mm and the length of an *E. coli* is roughly 1 μm. So DNA needs to be packages 3000 fold to end in a compartment of 0.5 μm in length.

(b) Calculate the search time, τ_{ws}, for the DNA target site without sliding.
Answer:

$$\tau_{ws} = \frac{V}{4\pi Da} = \frac{8\ \mu m^3}{4 \times \pi \times 5\ \frac{\mu m^2}{s} \times 5 \times 10^{-3} \mu m} = 25.5\ s \tag{7.80}$$

(c) Calculate the search time, τ_s, for the DNA target site with sliding.

$$\begin{aligned} \tau_s &= \left(\frac{V}{4\pi DL} + \frac{l_s^2}{2D_1}\right)\frac{L}{l_s} \\ &= \left(\frac{8\ \mu m^3}{4 \times \pi \times 5\ \frac{\mu m^2}{s} \times 1.5 \times 10^3\ \mu m} + \frac{(30 \times 10^{-3}\ \mu m)^2}{2 \times 5\ \frac{\mu m^2}{s}}\right)\frac{1.5 \times 10^3\ \mu m}{30 \times 10^{-3}\ \mu m} \\ &= 8.7\ s \end{aligned} \tag{7.81}$$

(d) Explain the meaning of $\frac{V}{4\pi DL}$, $\frac{l_s^2}{2D_1}$, and $\frac{L}{l_s}$ in equation 7.78.
Answer: $\frac{V}{4\pi DL}$ equals the (average) 3-D search time for DNA, $\frac{l_s^2}{2D_1}$ equals the (average) 1-D search time along the DNA, and $\frac{L}{l_s}$ equals the (average) number of searches before the entire DNA has been scanned for targets.

(e) When is sliding along DNA by transcription factors advantageous?
Answer: When $\tau_{ws} > t_s$ which is for instance the case when the cell volume is large and the target size is small.

7.14 Answers to exercises of Chapter 6

1. **Synthesis of cellular components by metabolism and adaptations of metabolism upon a nutrient change.**

   (a) Cyanobacteria generate energy obtained from sunlight during photosynthesis. This gives rise to ATP and NAD(P)H. This energy is used to make cellular components and to convert CO_2 into carbon intermediates of central metabolism. So, CO_2-fixation and photosynthesis makes them special and useful, because those sunlight and CO_2 are cheap.

   (b) i. The generation time equals $\frac{\ln 2}{5.2\times10^{-2}\ hr^{-1}} = 13.3\ hr$.

   ii. When we have 1 gram cyanobacterium then after 1 generation time we have 2 grams cyanobacterium. So, the mmol amount of DNA, RNA, and lipid per gram cyanobacterium needs to be produced within one generation time. This amounts to 0.24 $\frac{mmol}{gram\times hr} \times 13.3\ hr$ for protein, 0.0053 $\frac{mmol}{gram\times hr} \times 13.3\ hr$ for DNA, and 0.0082 $\frac{mmol}{gram\times hr} \times 13.3\ hr$ for lipid. These compounds are made by metabolism.

   iii. The precursors of DNA are: dATP, dGTP, dTTP, and dCTP. And for RNA we have: GTP, CTP, UTP, and ATP. DNA and RNA are both composed out of nucleotides but DNA out of the "deoxy" form; DNA deoxynucleotides contain one oxygen less than RNA nucleotides.

   iv. (Acetyl-CoA derived) Fatty acids and glycerol derived compounds.

   (c) It grows slower on glycerol; $7.9\times10^{-3}\ \frac{1}{hr}$ on glycerol versus $5.2\times10^{-2}\ hr^{-1}$ on CO_2. The largest changes occur in pentose phosphate pathway, TCA-cycle and glycolysis metabolism when metabolism shifts from glycerol to CO_2 metabolism (or vice versa).

2. **Measurement of the promoter activity of genes during steady-state cell growth (Advanced question).**

   (a) Let's first simplify the denominator of the promoter activity equation (with μ as the exponential growth rate),

$$\begin{aligned}\int_{t_1}^{t_2} OD(t)dt &= \int_{t_1}^{t_2} OD(0)e^{\mu t}dt = \frac{OD(0)(e^{\mu t_2} - e^{\mu t_1})}{\mu_I} = \frac{OD(t_2) - OD(t_1)}{\mu} \\ &= \frac{\Delta OD}{\mu} = \frac{OD(t_1)(e^{\mu(t_2-t_1)} - 1)}{\mu} \end{aligned} \tag{7.82}$$

   Next, we simplify the numerator,

$$f(t_2) - f(t_1) = f(t_1)(e^{\mu(t_2-t_1)} - 1) \tag{7.83}$$

Finally, we can simplify the promoter activity to (with τ as the generation time $\frac{\ln 2}{\mu}$),

$$\rho = \frac{f(t_2)-f(t_1)}{\int_{t_1}^{t_2} OD(t)dt} = \frac{f(t_1)(e^{\mu(t_2-t_1)}-1)}{\frac{OD(t_1)(e^{\mu(t_2-t_1)}-1)}{\mu}} = \frac{f(t_1)}{OD(t_1)}\frac{\ln 2}{\tau} \quad (7.84)$$

$$\ln\rho = \ln\frac{f}{OD} + \ln\frac{\ln 2}{\tau} = \ln\frac{f}{OD} + \ln\ln 2 - \ln\tau \quad (7.85)$$

The factor $\frac{f(t_1)}{OD(t_1)}$ is a concentration measure and time-independent at balanced growth, only dependent on the growth condition; we can therefore drop the time reference: $\frac{f}{OD}$. The generation time at the considered growth condition equals τ. Now, note that ρ is related to the transcription rate constant k: the growth rate is defined as $\mu = \frac{\ln 2}{\tau}$ and the concentration as $\alpha \times c = \frac{f}{OD}$. So, $\rho = \alpha \times c \times \mu$. When you remember the balance $\dot{c} = k - \mu \times c$, such that we have at steady state: $k = \mu \times c$! Therefore, the authors have measured $\alpha \times c \times \mu$ at steady state and this equals $\alpha \times k$.

(b) All the class one genes fall on a line when the natural logarithm of the promoter activity for two conditions (A and B) are plotted as function of each other because,

$$\ln\rho_A = \ln\frac{f_A}{OD_A} + \ln\frac{\ln 2}{\tau_A} = \ln\frac{f_A}{OD_A} + \ln\ln 2 - \ln\tau_A$$

$$\ln\rho_B = \ln\frac{f_B}{OD_B} + \ln\frac{\ln 2}{\tau_B} = \ln\frac{f_B}{OD_B} + \ln\ln 2 - \ln\tau_B$$

Class 1 genes: $\ln\frac{f_A}{OD_A} = \ln\frac{f_B}{OD_B} \Rightarrow \ln\rho_B - \ln\mu_B = \ln\rho_A - \ln\mu_A$

Therefore: $\ln\rho_B = \ln\rho_A - (\ln\mu_A - \ln\mu_B)$

$\Rightarrow \ln(\alpha \times k_B) = \ln(\alpha \times k_A) - (\ln\mu_A - \ln\mu_B)$

$\Rightarrow \ln k_B = \ln k_A - (\ln\mu_A - \ln\mu_B)$

(c) Because they also display concentration differences between the two conditions.

(d) Class 1 genes are constitutively expressed genes (condition unspecific) under the two conditions whereas class 2 genes are condition specific.

3. **Metabolism and growth responses of single cells exposed to nutrient changes.** They tell you that not all cells are capable of making a nutrient shift. So, subpopulations can occur in populations of microorganisms of which some can adapt to new conditions and some not. The origin of these subpopulations emerges from surprising stochastic behaviour of single cells that typically has its basis in stochasticity of gene expression.

4. **What sets the maximal growth rate of a bacterium?** During steady-state exponential growth – 'balanced growth' – the total number of

cells, N, increases exponentially in time as $N(t) = N(0)e^{\mu t}$ with μ as the specific growth rate in hr^{-1}. The total culture volume and total cell mass increase exponentially as well. Since metabolism is operating at steady state during balanced growth, the concentrations of all molecular species remain fixed over time such that the total number of molecules of every molecular species increases equally fast as the total volume. If we focus on the concentration of ribosome then we have the following balance equation at balanced growth, with f_{ribo} as the ribosome synthesis rate per ribosome and ϕ_{ribo} as the fraction of ribosome that is synthesising ribosomes (i.e. $\phi_{ribo} = \frac{\text{number of ribosomes synthesising ribosomes}}{\text{number of ribosomes synthesising protein}}$),

$$\begin{aligned} \frac{d}{dt}c_{ribo} &= v_{synthesis} - \mu c_{ribo} = 0 && (7.86)\\ &= f_{ribo}\phi_{ribo}c_{ribo} - \mu c_{ribo} && (7.87)\\ \Rightarrow \mu &= f_{ribo}\phi_{ribo} && (7.88) \end{aligned}$$

The last equation is a definition of the growth rate in terms of biochemistry.

(a) Calculate f_{ribo} in ribosomes per hour, given that a single ribosome contains 7459 amino acids and that the translation rate is 20 amino acids per ribosome per second.
Answer:

$$f_{ribo} = \frac{20\ \frac{aa}{ribosome \times s}}{7459\ \frac{aa}{ribosome}} 3600\ \frac{s}{hr} \approx 10 \frac{ribosomes}{ribosome \times hr} \qquad (7.89)$$

(b) Calculate the maximal growth rate.
Answer: the maximal growth rate occurs with $\phi_{ribo} = 1$, so the cell only makes proteins that make new proteins themselves, such that $\mu_{max} = f_{ribo} = 10\ hr^{-1}$.

(c) How many doublings do you have per hour with this maximal growth rate?
Answer: The minimal generation time equals $\tau_{min} = \frac{\ln 2}{\mu_{max}} = \frac{\ln 2}{10} = 0.07\ hr = 4.5\ minutes$. So every 4.5 minutes the cell number doubles, therefore per hour about 13 doublings occur.

(d) Calculate the maximal doubling rate in $\frac{doublings}{hr}$.
Answer: The maximal doubling rate equals 4.5 $\frac{doublings}{hr}$.

(e) The experimentally-determined minimal doubling time (= minimal generation time) that the bacterium *Escherichia coli* can attain equals 20 $\frac{min}{doubling}$. Calculate the fraction of ribosome synthesising ribosomes at this growth rate.
Answer: $\mu = \frac{\ln 2}{\tau} = \frac{\ln 2}{20} = 0.034\ min^{-1}\ 60\ \frac{min}{hour} = 2\ hr^{-1} = f_{ribo}\phi_{ribo} = 10\ hr^{-1}\ \phi_{ribo}$. So $\phi_{ribo} = 0.2$, 20% of the ribosomes makes ribosomes. The remaining 80% of the ribosome make proteins that are not ribosomes.

(f) In reality, ϕ_{ribo} will always be smaller than 1 because the cell needs always other proteins than ribosomes. Examples of the other proteins are metabolic proteins that make amino acids out of nutrients, in order to allow ribosomes making proteins. Therefore if $\phi_{ribo} < 1$ other proteins are made in addition to ribosomes, such that ribosomal protein fraction of $\Phi_{ribo} = \frac{[ribosomes]}{[proteins]}$ is smaller than 1. Show that the fraction of ribosomes making ribosomes equals the ribosomal protein fraction: $\phi_{ribo} = \Phi_{ribo}$.
Answer: At steady state, the total protein and ribosome balance show this relation,

$$\frac{d}{dt}c_{ribo} = f_{ribo}\phi_{ribo}c_{ribo} - \mu c_{ribo} = 0 \quad \Rightarrow \mu = f_{ribo}\phi_{ribo} \tag{7.90}$$

$$\frac{d}{dt}c_{protein} = f_{ribo}\Phi_{ribo}c_{protein} - \mu c_{protein} = 0 \quad \Rightarrow \mu = f_{ribo}\Phi_{ribo} \tag{7.91}$$

$$\Rightarrow \quad \phi_{ribo} = \Phi_{ribo} \tag{7.92}$$

(g) Relate the ribosomal protein fraction to the growth rate.
Answer: $\mu = f_{ribo}\phi_{ribo} = f_{ribo}\Phi_{ribo}$.

(h) In reality some percentage of the ribosomes is always inactive – they are 'maturating' – such that the total ribosome concentration equal the sum of active and inactive ribosome: $c_{ribo} = c^A_{ribo} + c^I_{ribo}$ and therefore $\mu = f_{ribo}\frac{c^A_{ribo}}{c^A_{ribo}+c^I_{ribo}+c_m} = f_{ribo}\frac{c_{ribo}-c^I_{ribo}}{c^A_{ribo}+c^I_{ribo}+c_m} = f_r(\Phi_{ribo} - \Phi^I_{ribo})$. When we use experimental data for *E. coli* to study the relationship between the ribosomal protein fraction (y-axis) as function of the growth rate (x-axis), we find a linear relation (Figure 1A, in Scott et al, Science, 2010). Use the $\mu = f_{ribo}(\Phi_{ribo} - \Phi^I_{ribo})$ to identify what the slope and intercept of this linear relation is.
Answer: We plot Φ_{ribo} as function of μ which is given by $\Phi_{ribo} = \frac{\mu}{f_{ribo}} + \Phi^I_{ribo}$ since $\mu = f_{ribo}(\Phi_{ribo} - \Phi^I_{ribo})$. So, the slope equals $1/f_{ribo}$, which corresponds to 1 divided by the translation rate per ribosome. The intercept equals Φ^I_{ribo}, which equals the inactive ribosome fraction.

(i) How does the previous relation change when we inhibit translation using a translation inhibitor (chloramphenicol) that only affect f_{ribo} (Figure 1B, in Scott et al, Science, 2010)?
Answer: the slope increases because it equals $1/f_{ribo}$ and f_{ribo} is reduced.
Add experimental figure 1AB of Scott and Hwa paper.

5. **Mass flow through a steady state metabolic network.** Metabolic networks are responsible for the synthesis of energy and precursor molecules for the construction of cellular macromolecules such as proteins, DNA, RNA and membranes. Metabolic networks are highly branched and can contain upto thousands of reactions. Navigating through those networks

is therefore not straightforward and computational tools are often used to simplify this process. Since mass balance depend linearly on reaction rates, we can use linear algebra to study these huge networks. This linearity is reflected in the mass balance equations that depend on the reaction rates as a linear function: $\frac{dx}{dt} = \sum_{i=1}^{r} n_i v_i$, with the n_i as a (positive or negative) fixed stoichiometry coefficient, v_i as a concentrations-depend reaction rate and r as the number of the reactions in the network. The relation would be nonlinear when, for instance, $\frac{dx}{dt} = \sum_{i=1}^{r} n_i v_i^m$, with m as a number greater than 1 and then linear algebra would not be as useful. In this exercise you will realise the consequence of this linear relation, its use and how it relates to linear algebra. We study a simplified, toy network that has particular illustrative features shown in figure 6.7.

(a) Give the mass balances of the variable metabolite concentrations.
Answer:

$$\begin{aligned}
\frac{dx_1}{dt} &= v_1 - v_2 - v_7 \\
\frac{dx_2}{dt} &= v_2 - v_3 - v_5 \\
\frac{dx_3}{dt} &= v_3 - v_4 \\
\frac{dx_4}{dt} &= v_5 - v_6 \\
\frac{dx_5}{dt} &= v_7 - v_8 - v_9 \\
\frac{dx_6}{dt} &= v_8 + v_9 - v_1 0
\end{aligned} \tag{7.93}$$

(b) Consider the network at steady state with reaction rate 1 equal to 10 $\frac{mM}{min}$. Give the reaction rate values of the reactions when the effluxes are equal to $v_6 = 5\ \frac{mM}{min}$, $v_4 = 2\ \frac{mM}{min}$, and $v_{10} = 3\ \frac{mM}{min}$.
Answer: See figure 7.9, note that the fluxes values of v_8 and v_9 cannot be determined, only their sum, given the flux information of only the input and the output fluxes.

(c) Think a bit more about the cycle, containing reaction 8 and 9, why do you not know their rates at steady state when you only know $v_1,\ v_4,\ v_6$, and v_{10}?
Answer Because only their sum can be determined from this information. You only know that $v_8 + v_9 = 3$ and this can be achieved with many fluxes values of v_8 and v_9 since they can have any numbers. For instance $v_8 = -5$ and $v_9 = 8$ would work or $v_8 = 7$ and $v_9 = -4$; in fact an infinite number of combinations is possible.

(d) Say you would know reaction rates v_3 , v_5 , v_7 and v_9 would this allow you to unambiguously determine all flux values?
Answer: Yes, then we would get an unique answer because if we

know v_7 then we know $v_8 + v_9$ and v_{10}, by knowing v_9 we can determine v_8.

(e) Another way to look at mass flow through this network at steady state is as if it is the superposition of 4 subnetworks, which each can attain a steady state on their own. Which reactions would those subnetworks contain?
Answer: One possibility is, their exist several, shown in figure 7.10, each of those subnetworks can achieve a steady state on its own and the steady-state flux distribution of the entire network is the superposition – the addition – of the steady-state flux values through each of the subnetworks (because then the entire network is also at steady state).

(f) The flux vector $\mathbf{j}$ of this network is defined as the vector that contains all the values of the reaction rates, so

$$\mathbf{j} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \end{pmatrix} \tag{7.94}$$

Write down the flux vectors, $\mathbf{j}_i$ of the 4 subnetworks, each having 1 column and 10 rows, and put zeros at entries that correspond to reactions that are not used in this subnetwork.
Answer: Subnetwork 1:

$$\mathbf{j} = \begin{pmatrix} v_1 \\ v_2 \\ 0 \\ 0 \\ v_5 \\ v_6 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{7.95}$$

Subnetwork 2:

$$\mathbf{j} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{7.96}$$

Subnetwork 3:

$$\mathbf{j} = \begin{pmatrix} v_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ v_7 \\ v_8 \\ 0 \\ v_{10} \end{pmatrix} \tag{7.97}$$

Subnetwork 4:

$$\mathbf{j} = \begin{pmatrix} v_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ v_7 \\ 0 \\ v_9 \\ v_{10} \end{pmatrix} \tag{7.98}$$

(g) Now the statement is that the following relation holds,

$$\mathbf{j} = \sum_{i=1}^{5} \alpha_i \mathbf{j}_i \tag{7.99}$$

with α_i as the contribution of the subnetwork to the flux through the entire network. Use the flux vector that you have identified in (b) to find the values of α_i when you set all the rates in the subnetworks to

1.
Answer: Subnetwork 2:

$$\mathbf{j} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 10 \\ 7 \\ 2 \\ 2 \\ 5 \\ 5 \\ 3 \\ 3 - v_9 \\ 3 - v_8 \\ 3 \end{pmatrix} \quad (7.100)$$

The fifth and sixth row tells you that $\alpha_1 = 5$, the third row tells you that $\alpha_2 = 2$, the 8th row tells you that $\alpha_3 = 3 - v_9$ and the 9th row tells you that $\alpha_4 = 3 - v_8$.

(h) Do the α_i values change when the effluxes change in value?
Answer: Yes, if the effluxes change the fluxes in the network change such that α's should change.

6. **Operon organization in bacterial genomes.** Genes are often contained in groups, called 'operons', in bacterial genomes. Consider again figure 6.7. We do not really understand why operon structure is as it is, one hypothesis worked out in this exercise. We know however that this cannot be the whole story because of all kinds of processes that distort the gene order on bacterial genomes such as horizontal gene transfer and because of gene shuffling due recombination events to transposons.

(a) Why would it make sense to have enzyme 5 and 6 in one operon and 3 and 4 in another? Why are those two operons expected not to have enzyme 2 in it?
Answer: It would make sense to have reaction pairs $\{5, 6\}$ and $\{3, 4\}$ always in one operon because those reaction pairs always carry the same flux value. Since $v_2 = v_5 + v_3$ it should not be in each of those operons but in separate operon.

(b) Why are enzyme 1 and 2 most likely not in the same operon?
Answer: Because their flux values are generally not the same, since $v_1 = v_2 + v_7$.

(c) If the cycle has an additional function, not related to this pathway, then it is expected be part of a different operon. Why could reaction 7 and 10 then still be part of the same operon?
Answer: Because reaction 7 and 10 always carry the same flux.

(d) What is the assignment of metabolic-enzyme genes into operons leading to the smallest number of operons (and the smallest number of

genes; so do not use the same gene twice) that still allows for complete flux flexibility through the network?
Answer:

- operon 1: enzyme 1
- operon 2: enzyme 2
- operon 3: enzyme 5 and 6
- operon 4: enzyme 3 and 4
- operon 5: enzyme 7 and 10
- operon 6: enzyme 8 and 9

(e) Why do *you* think – because I (Frank) do not know the answer – that genes do not occur in multiple copies such that they can participate in different operons?
Answer: One of the reasons, which is sometimes proposed, is that when a gene is duplicated – and carries out the same function – one copy would deteriorate by random mutations and the other would remain functional. This holds when the duplicate copy is truly redundant and carries out the same function. This would not be the case in the scenario that we are considering now; the duplicate copy would be in a different operon and likely controlled by a different set of transcription factors. Nonetheless we rarely find duplicate gene copies across bacterial genomes.

7. **Why some enzymes couple ATP hydrolysis to biochemical conversions.** In metabolism, many reactions are coupled to the hydrolysis of ATP into ADP and Pi. In this question you will study why this is so common. Consider the following reactions,

$$S \overset{1}{\rightleftharpoons} P, \quad v_1 = V_{1,MAX} \frac{s \cdot \left(1 - \frac{p}{s \cdot K_{eq,1}}\right)}{1 + s + p}$$

$$S + ATP \overset{2}{\rightleftharpoons} P + ADP + Pi, \quad v_2 = V_{2,MAX} \frac{s \cdot atp \cdot \left(1 - \frac{p \cdot adp \cdot pi}{s \cdot atp \cdot K_{eq,2}}\right)}{(1 + s + p)\,(1 + pi)\,(1 + atp + adp)}$$

The difference between reaction 1 and 2 is that enzyme 2 couples the hydrolysis of ATP to the formation of P out of S and enzyme 1 does not do this. The following constraints apply in the cell: i. P should be formed out S to allow for growth, ii. in the cell, the concentrations of P, S, ATP, ADP, and Pi fall within strict physiological bounds, they all vary between 0.1 and 10 mM, iii. the equilibrium constant of reaction 1 equals $K_{eq,1} = 10^{-3}$ and for reaction 2 it equals $K_{eq,2} = 10^3\ mM$. (In this question, do not consider that at the level of this reaction the adenosine and phosphate in ATP, ADP and Pi remain fixed.) Show that reaction 2 can form P out of S inside the cell and that reaction 1 cannot achieve this.

Answer: The maximal rate of reaction 1 equals,

$$v_1 = V_{1,MAX}\frac{10(1-\frac{0.1}{10\times10^{-3}})}{1+10+0.1} = -8.1V_{1,MAX}$$

$$v_2 = V_{2,MAX}\frac{10\times10\left(1-\frac{0.1\times0.1\times0.1}{10\times10\times10^3}\right)}{(1+10+0.1)(1+0.1)(1+10+0.1)} = 0.74V_{2,MAX} \quad (7.101)$$

By coupling the reaction to ATP hydrolysis, P is made out of S which allows the cell to grow.

8. **Regulation of flux at branch points in metabolism.** In metabolism it occurs very often that a metabolic pathways branches into two directions. The balance for the metabolite concentration at this junction equals, the difference between the synthesis rate v and the two consumption rates v_1 and v_2 corresponding to the branch rates.

$$\frac{dx}{dt} = v_1 - v_2 - v_3 \quad (7.102)$$

$$v_1 = V_1\frac{\frac{s}{K_s}}{1+\frac{s}{K_s}+\frac{x}{K_1}} \quad (7.103)$$

$$v_2 = V_2\frac{x}{K_2+x} \quad (7.104)$$

$$v_3 = V_3\frac{x}{K_3+x} \quad (7.105)$$

At steady state we denotes rates as fluxes and we get,

$$0 = J_1 - J_2 - J_3 \quad (7.106)$$

$$v_1 = J_1 \quad (7.107)$$

$$v_2 = J_2 \quad (7.108)$$

$$v_3 = J_3 \quad (7.109)$$

$$x = x_s \quad (7.110)$$

$$\Rightarrow V_1\frac{\frac{s}{K_s}}{1+\frac{s}{K_s}+\frac{x_s}{K_1}} = V_2\frac{x_s}{K_2+x_s} + V_3\frac{x_s}{K_3+x_s} \quad (7.111)$$

Set $s = 10,\ K_s = 1,\ V_1 = 10,\ K_1 = 10,\ V_2 = 10,\ K_2 = 1, V_3 = 10$ and $K_3 = 10$.

(a) Use a plotting program, e.g. Excel, to plot the rates of the reaction as function of x. Determine the steady state value of x.
Answer: See figure 7.11. Note that the steady state concentration is determined by the intersection of v_1 with v_1+v_2. This concentration equals 2.3.

(b) Plot the ratio v_2/v_3 as function of x. Why is reaction 2 more active than 3 at low concentrations of X? Explain your result.
Answer: Reaction 2 has a lower K_M for X and is therefore more active at low X than reaction 3 is.

(c) Double the maximal rate of third reaction. What happens to the steady state flux and the concentration of x? Explain you result.
Answer: The concentration of X drops to 1.6, the flux through reaction 3 increases, through reaction 2 it decreases, and through reaction 1 it increases.

(d) Half the maximal rate of third reaction. What happens to the steady state flux and the concentration of x? Explain you result.
Answer: The concentration of X rises to 3.2, the flux through reaction 3 decreases, through reaction 2 it increases, and through reaction 1 it decreases.

(e) What happens when you reduce K_3 by a factor of 2? Explain your result.
Answer: The concentration of X drops to 1.7, the flux through reaction 3 decreases, through reaction 2 it increases, and through reaction 1 it increases.

9. **Supply and demand analysis of metabolic pathways.** Negative feedback occurs often in metabolic pathways. The negative feedback metabolite cuts the pathway into blocks or systems: a supply and a demand system (figure 6.9). In this exercise you will study the basic consequences of negative feedback in metabolic pathways, using a simplified model that captures the main effects of negative feedback,

$$\frac{dx}{dt} = \underbrace{V_s \frac{1}{1+\left(\frac{x}{K}\right)^n}}_{\text{supply rate}} - \underbrace{V_d \frac{x}{1+x}}_{\text{demand rate}} \tag{7.112}$$

The basic parameter setting is that all parameters equal 1. In all plots below, plot always the supply rate and the demand rate.

(a) Investigate the influence of n. Use Excel to plot the supply rate and degradation rate as function of x. Vary n: set it to 1, 2, 4, 6, and 8.
Answer: See figure 7.12.

(b) Investigate the influence of K. Set n to 4 and vary K: set it to 0.1, 0.5, 1, 2, 4.

(c) Investigate the influence of V_1. Set n to 1 and to 6. Vary for those two cases V_1: set it to 0.5, 1, and 2.
Answer: See figure 7.12.

(d) Investigate the influence of V_2. Set n to 1 and to 6. Vary for those two cases V_2: set it to 0.5, 1, and 2.
Answer: See figure 7.12.

(e) Is it true that the effect of a strong feedback is that the concentration of x hardly changes when changes are made to either V_1 or V_2 (metabolite homeostasis)? And that the influence of the maximal

rate of the demand system on the steady-state flux can greatly exceed that of the supply system if metabolite homeostasis occurs (flux control by demand)?
Answer: Yes, this is indeed the effect of a strong feedback, which makes v_1 to steeply decline as function of x. So, feedback leads to homeostasis of x and the flux is set by the demand process, not by the supply process.

Metabolite homeostasis and flux control by demand have been identified as the main functional consequences of negative feedback in metabolic pathways.

10. **Steady-state responses of metabolic pathways to changes in metabolic enzyme concentration through gene expression.** When the environment changes many metabolic pathways are adapted to the new condition via increased or reduced expression of genes coding for metabolic enzymes.

    (a) Say the steady-state flux in the pathway increases because the first enzyme is increased in concentration. What happens to the steady state concentrations of X_1, X_2, X_3 and X_4?
    Answer: All concentration will rise. If the steady-state flux increased then the substrate concentration of the last enzyme, X_4, must have increased. This leads to a higher flux through enzyme 5 but also more inhibition of enzyme 4 which should be compensated for by an increase in x_3. This inhibits enzyme 3, x_2 should therefore rise to compensate. And finally x_1 should rise to compensate for the rate inhibition of enzyme 2 by X_2. x_1 can rise because the enzyme 1 was increased in concentration which leads to a rate enhanced of reaction 1.

    (b) Consider the change in the steady state of the metabolic pathway when enzyme 2 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?
    Answer: The flux will generally increase, the concentration of X_1 will drop and those of X_2 to X_4 will all rise. If the flux increases then enzyme 1 can only run faster if product inhibition is relieved via a lower concentration of X_1. Enzyme 5 can only run faster if its substrate is increased in concentration, x_4, which inhibits enzyme 4 more and should be compensated by a rise in x_3, which inhibits enzyme 3 more and should be compensated by a rise in x_2. This rise in enzyme 2 still leads to an increase in flux through enzyme 2 because more enzyme was added.

    (c) Consider the change in the steady state of the metabolic pathway when enzyme 3 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?

Answer: The flux will generally increase, the concentration of X_1 and X_2 will drop and those of X_3 to X_4 will all rise.

(d) Consider the change in the steady state of the metabolic pathway when enzyme 4 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?
Answer: The flux will generally increase, the concentration of X_1 and X_2 and X_3 will drop and that of X_4 will rise.

(e) Consider the change in the steady state of the metabolic pathway when enzyme 5 is increased in concentration. What would you predict for the changes in the steady state flux and the metabolite concentrations, i.e. X_1 and X_4?
Answer: The flux will generally increase, all concentrations will drop.

11. **A different look at a classical metabolic pathway: glycolysis in yeast** Central to most biochemical pathways is glycolysis, the breakdown of glucose to pyruvate in 10 enzymatic steps. The details of the pathway can be looked up in any textbook of biochemistry, or on wikipedia. Here we simplify and look at some interesting aspects from a systems biology point of view. The simplified pathway is depicted in figure 6.10

(a) make a stoichiometry matrix N of this pathway: take glucose and pyruvate fixed, so they do not need to be balanced and are therefore not part of the matrix. Each row represent a balance for a metabolite, each column correpsonds to a reaction, such that $N \cdot v$ form the set of balances for this system.
Answer:

$$\begin{pmatrix} & v1 & v2 & v3 & v4 & v5 \\ \dot{fbp} & 1 & -1 & 0 & 0 & 0 \\ \dot{dhap} & 0 & 1 & -1 & 0 & 0 \\ \dot{gap} & 0 & 1 & 1 & -1 & 0 \\ \dot{atp} & -2 & 0 & 0 & 2 & -1 \\ \dot{adp} & 2 & 0 & 0 & -2 & 1 \\ \dot{p} & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

(b) apply Gaussian elimination on this matrix to transform it into its reduced row echelon form and decide on the number of row dependencies in this pathway. Think about what such a row dependency means in biochemical terms. For example, you should readily see that the rows for ATP and ADP sum up to 0 (and so these two rows are dependent). What does this mean? Provide a similar explanation for the other dependency.

Answer:

$$\begin{pmatrix} & v1 & v2 & v3 & v4 & v5 \\ \dot{fbp} & 1 & -1 & 0 & 0 & 0 \\ \dot{dhap} & 0 & 1 & -1 & 0 & 0 \\ \dot{gap} & 0 & 0 & 2 & -1 & 0 \\ \dot{atp} & 0 & 0 & 0 & 1 & -1 \\ \dot{adp} & 0 & 0 & 0 & 0 & 0 \\ \dot{p} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

There are two rows with only zero's: these are dependent rows. It means that something is conserved. For example, $\dot{adp} + \dot{atp} = 0$, which means that $adp + atp = C$. This is conservation of the "A", adenosine. Only a phosphate is removed from ATP and put back on, the A moiety remains. The other row of zero's reflect the conservation of phosphate: to get the $\dot{p}$ row zero, we in the end had to do the following: add 2 times row 1, add row 2, add row 3 and add row 4. Hence: $2\dot{fbp} + \dot{dhap} + \dot{gap} + \dot{atp} + \dot{p} = 0$. These are all the phosphate-containing species in the model! You should figure out why $f\mathbf{b}p$ counts double...

(c) in the reduced row echelon form you can read off the solution to the equation: $N \cdot v = 0$. This will give you dependencies between the steady state rates, which we call fluxes.

i. Show that indeed, in steady state, glycolysis yields 2 ATP per glucose
Answer: Reading off the RRE form of N gives:

$$\begin{aligned} v2 &= v1 \\ v3 &= v2 & &= v1 \\ v4 &= 2v3 & &= 2v1 \\ v5 &= v4 & &= 2v1 \end{aligned}$$

So we can conclude that $v5$, the rate of ATP expenditure, is two times the rate of glucose conversion, $v1$. So 2 ATP was made per Glc.

ii. Sketch the solution space (or null space) of this system in the $v1$,$v5$ plane under the constraints that $v1 > 0$ and $v5 > 0$.
Answer: See figure 7.14.

(d) inspect in the reduced row echelon form of N what would happen if you would not have reaction 5 in, i.e. no reaction that consumes ATP? Do this by simply removing the last column of N. Link the mathematical consequence to biochemical reasoning: can ATP be in steady state, then, and why (not)?

Answer: removing $v5$ gives:

$$\begin{pmatrix} & v1 & v2 & v3 & v4 \\ \dot{fbp} & 1 & -1 & 0 & 0 \\ \dot{dhap} & 0 & 1 & -1 & 0 \\ \dot{gap} & 0 & 0 & 2 & -1 \\ \dot{atp} & 0 & 0 & 0 & 1 \\ \dot{adp} & 0 & 0 & 0 & 0 \\ \dot{p} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Now you can read that for ATP to be balanced: $\dot{atp} = 0 \Rightarrow v4 = 0$, and hence all fluxes are 0! Simply put: if there is no consumption of ATP, glycolysis cannot run because it produces ATP and it needs a sink for it.

(e) now we will inspect a peculiar state that glycolysis get in. We put $v5$ back in, and let's assume that $v2$ and $v3$ act really fast and FbP, DHAP and GAP form one pool, and effectively $v1$ produces 2 GAP molecules. $v2$ and $v3$ are then simply thrown out of the model. We remain with 3 reactions with the following kinetics:

$$v1 = V_{m1} \cdot Glc \frac{atp}{atp + 0.5} \quad (7.113)$$

$$v4 = V_{m4} \frac{gap}{gap + 1} \quad (7.114)$$

$$v5 = k_5 \cdot atp \quad (7.115)$$

Suppose now that V_{m4} is really low, say 0.1 (in reality this is caused by a low inorganic phosphate concentration whose action we model through V_{m4}). V_{m1} and k_5 are set to 1.

i. in this state gap will accumulate to very high levels because of the low activity of V_{m4}. What will then be the rate of $v4$?
Answer: If gap is much higher than 1 (the K_m of the enzyme catalyzing reaction 4), $\frac{gap}{gap+1} \to 1$ and hence, $v4 = V_{m4} = 0.1$. This rate is now independent of other metabolite concentrations.

ii. construct a *rate characteristic* by plotting ATP production ($v4$) and ATP consumption ($v1 + v5$) as a function of atp. Where is the steady state, at what flux and atp concentration?
Answer:
The two lines in figure 7.14 cross at $atp \approx 0.035$ at a rate of 0.1. Here ATP production rate equals ATP consumption rate.

iii. now we look at the balance for gap: this now reads (with $v2$ and $v3$ removed): $\frac{dgap}{dt} = 2v1 - v4$. Now construct a rate characteristic around gap (at the steady state atp level from the previous question!): what is your conclusion?
Answer:

This plot, figure 7.14, is constructed with a rate of $v1$ that is set by $atp = 0.035$. Only $v4$ depends on *gap*. Since for a balance of GAP $v4$ needs to be twice as fast as $v1$, you can see that there is no *gap* where a steady state is possible: gap will accumulate forever. This is actually observed experimentally in some yeast mutants!

12. **Can gene expression optimise the flux through a metabolic network under a constraint of fixed maximal enzyme available for investment in the metabolic network?** As biologists we are all familiar with gene activity regulation leading to changes in protein concentrations as function of conditions. We accept this as a fact. Why the gene regulation functions as it does is however generally not clear. Some would argue that the gene regulation mechanisms have evolved to maximise fitness of the bacterium. Here we explore the consequences of this hypothesis and ask whether gene regulation can maximise the fitness of a metabolic pathway. We define the fitness of metabolic pathway as its steady state flux divided by the total enzyme amount invested in this pathway. The gene regulation mechanism should maximise the fitness as function of external conditions. Does such a gene regulation mechanism exist? And how does its parameterisation depend on the kinetics of the metabolic enzymes? In this question you will answer those questions yourself, step by step. We start by introducing the metabolic network, it is the simplest that is still realistic and illustrates all the features of realistic networks that are clearly involve hundreds of enzymes,

$$\underline{S} \overset{1}{\rightleftharpoons} X \overset{2}{\rightleftharpoons} \underline{P} \tag{7.116}$$

$$v_1 = e_1 \underbrace{k_1 \frac{s}{1+s+x}}_{f_1(x)} = e_1 f_1(x) \tag{7.117}$$

$$v_2 = e_2 \underbrace{k_2 \frac{x}{1+x}}_{f_2(x)} = e_2 f_2(x) \tag{7.118}$$

$$e_T = e_1 + e_2 \tag{7.119}$$

$$\frac{dx}{dt} = v_1 - v_2 = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} \tag{7.120}$$

We denote the steady-state flux by J and it is defined by the steady-state relation,

$$J = v_1 = v_2 \tag{7.121}$$

What we want to maximise is the fitness of the metabolic pathway, F,

$$F = \frac{J}{e_T} = \frac{J}{e_1 + e_2} \tag{7.122}$$

So given the amount of enzyme available, e_T, we ask for the distribution of enzymes that maximises J, which is the same as saying that we maximise $\frac{J}{e_T}$. Maximising $\frac{J}{e_T}$ is the same as minimising $\frac{e_T}{J}$.

(a) Show that $\frac{e_T}{J}$ obeys,

$$\frac{e_T}{J} = \frac{1}{f_1} + \frac{1}{f_2} = \frac{1+s+x}{k_1 s} + \frac{1+x}{k_2 x} = \frac{1}{k_1 s} + \frac{1}{k_1} + \frac{x}{k_1 s} + \frac{1}{k_2 x} + \frac{1}{k_2} \tag{7.123}$$

(b) When $\frac{e_T}{J}$ is minimal then $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) = 0$. Why is this true?

(c) Determine $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right)$ and solve for x. Call this x, the optimal x, denoted by x_o.
Answer: $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$

(d) Show that the relation $\frac{v_1}{v_2} = 1$ leads to the relation $e_2 = e_T \frac{f_1}{f_1+f_2} = e_T \frac{1}{1+\frac{f_2}{f_1}}$.

(e) The problem now is that the previous relation is always true, also in non-optimal states. However, the requirement that $\frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) = 0$ leads to an expression for $\frac{f_2}{f_1}$ that is only true in the optimal state, because when $x = x_o$ we have,

$$\begin{aligned} 0 = \frac{d}{dx}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) &= \frac{d}{dx}\frac{1}{f_1} + \frac{d}{dx}\frac{1}{f_2} = \frac{\partial \frac{1}{f_1}}{\partial f_1}\frac{\partial f_1}{\partial x} + \frac{\partial \frac{1}{f_2}}{\partial f_2}\frac{\partial f_2}{\partial x} \quad (7.124) \\ &= -\frac{1}{f_1^2}\frac{\partial f_1}{\partial x} - \frac{1}{f_2^2}\frac{\partial f_2}{\partial x} = -\frac{1}{f_1}\frac{\partial \ln f_1}{\partial x} - \frac{1}{f_2}\frac{\partial \ln f_2}{\partial x} \\ \Rightarrow -\frac{1}{f_1}\frac{\partial \ln f_1}{\partial x} &= \frac{1}{f_2}\frac{\partial \ln f_2}{\partial x} \quad \Rightarrow \frac{f_2(x_o)}{f_1(x_o)} = -\frac{\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o}}{\left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}} \quad (7.125) \end{aligned}$$

Such that at the optimal steady state we have the relation $e_2 = e_T \frac{1}{1+\frac{f_2}{f_1}} = e_T \frac{1}{1-\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o} \Big/ \left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}}$. Determine $-\frac{\left.\frac{\partial \ln f_2}{\partial x}\right|_{x=x_o}}{\left.\frac{\partial \ln f_1}{\partial x}\right|_{x=x_o}}$ and then the equation for e_2.
Answer: $e_2 = e_T \frac{x_o(1+x_o)}{s+(1+x_o)^2}$

(f) The problem is that $-\frac{\frac{\partial \ln f_2}{\partial x}}{\frac{\partial \ln f_1}{\partial x}}$ depends still on s, such that we have relation that expresses e_2 in terms of s and x_o, and we would like to use this relation to choose a gene regulation mechanism that relates e_2 to x only, and not to s, because the gene regulation circuit depend

on x – for instance, because a transcription factor binds to x and this transcription determines the synthesis of e_2. Using the previous relation, $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$, which holds in the optimum, we can express s in terms of x_o. What do we obtain for s and for e_2?
Answer: $s = \frac{k_2 x_o^2}{k_1}$ and $e_2 = e_T \frac{k_1 x_o(1+x_o)}{k_2 x_o^2 + k_1(1+x_o)^2}$.

(g) The relation that we have just found, i.e. $e_2 = e_T \frac{k_1 x_o(1+x_o)}{k_2 x_o^2 + k_1(1+x_o)^2}$, relates e_2 to x_o; it is therefore a relation for $e_{2,o}$! What should be the mass balance for e_2 to obey this relation at steady state?
Answer: For instance, $\frac{d}{dt}e_2 = e_T \frac{k_1 x(1+x)}{k_2 x^2 + k_1(1+x)^2} - e_2$

(h) Remark, not a question: The statement is now that the following coupled metabolic-transcription-translation system,

$$\frac{dx}{dt} = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} \quad (7.126)$$

$$\frac{d}{dt}e_2 = e_T \frac{k_1 x(1+x)}{k_2 x^2 + k_1(1+x)^2} - e_2, \quad (7.127)$$

gives the optimal steady-state for every value of s! Because this system always has as steady state: $x = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$, which is the requirement for optimality! Note that the $\frac{k_1 x(1+x)}{k_2 x^2 + k_1(1+x)^2}$ is very similar to a hyperbolic equation of x (you can verify by plotting it for different values of k_1 and k_2), suggesting that a single transcription factor that binds to X and to the DNA suffices for optimal regulation. In fact you can rewrite this equation into $\frac{x+x^2}{1+2x+\left(\frac{k_2}{k_1}+1\right)x^2}$ which sort of corresponds to a model of cooperative binding of the transcription factor to the promoter with two binding sites.

(i) Another remark: Another way of solving the problem of finding an optimal gene regulation network is to find a e_2 function that makes sure that $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$. This e_2 function, in terms of x, is then the steady state input-output relation of the optimal gene regulation network. How can you do this? Set $\frac{dx}{dt} = k_1(e_T - e_2)\frac{s}{1+s+x} - k_2 e_2 \frac{x}{1+x} = 0$ and solve for x. Set this equation equal to $\frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$ and solve for e_2. Then we have found an e_2 steady state that is optimal. The problem now is that this e_2 equation is still in terms of s and not in terms x – and we want the gene network to sense the metabolic state and not the environmental state. To solve this, solve s from $x_o = \frac{\sqrt{k_1}\sqrt{s}}{\sqrt{k_2}}$ and substitute it in the equation in the e_2 relation to arrive at one that is in terms of x, and no longer in terms of s. This is bit faster than what was suggested above.

(j) Another remark: note that since the objective is solely defined in terms of metabolic properties, the parameterisation of the optimal gene network is ultimately completely set by metabolic properties

alone. So, the gene regulation is a 'controller' of the metabolic system that has the right information of the metabolic system to be able to steer it always to the optimum, regardless of the environment. Isn't that amazing? I think that is the right way of thinking about gene regulation. However this shifts the problem to identifying what the control objective, the fitness objective, is. It is certainly not always J/e_T. How to do that is still not clear to me (= Frank).

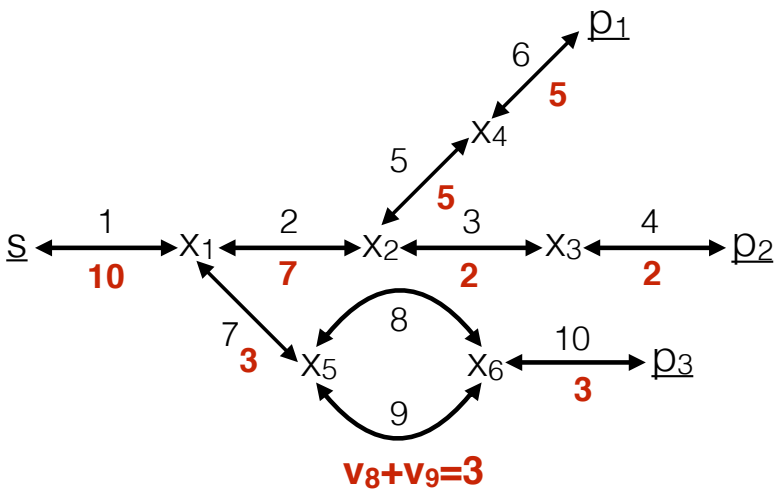


Figure 7.9: **Flux distribution answer.**

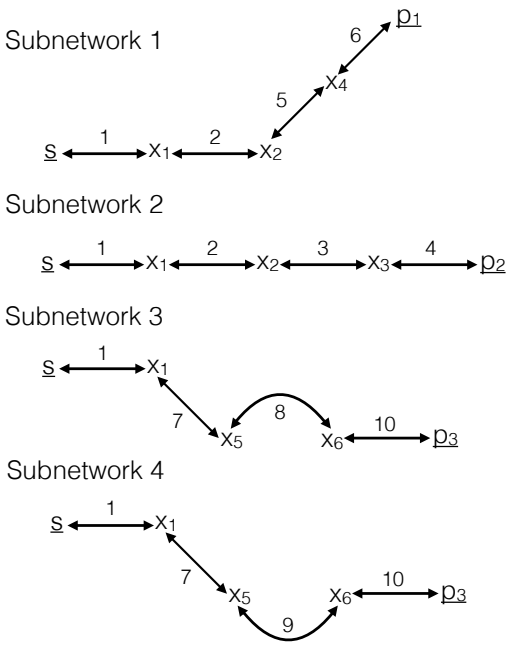


Figure 7.10: **Subnetworks answer.**

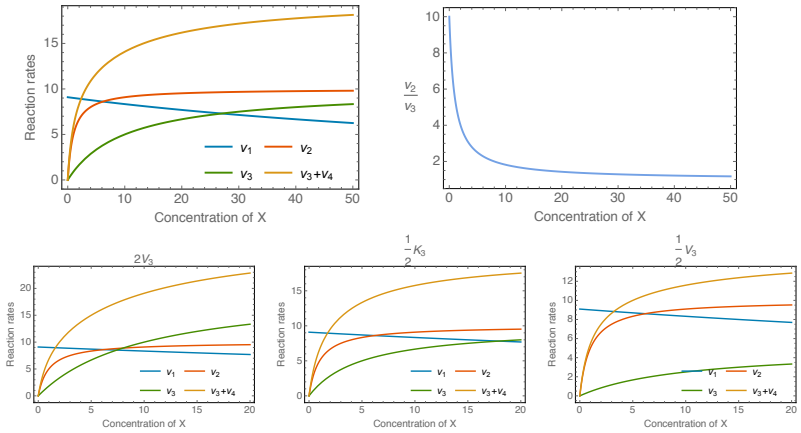


Figure 7.11: **A metabolism branch**

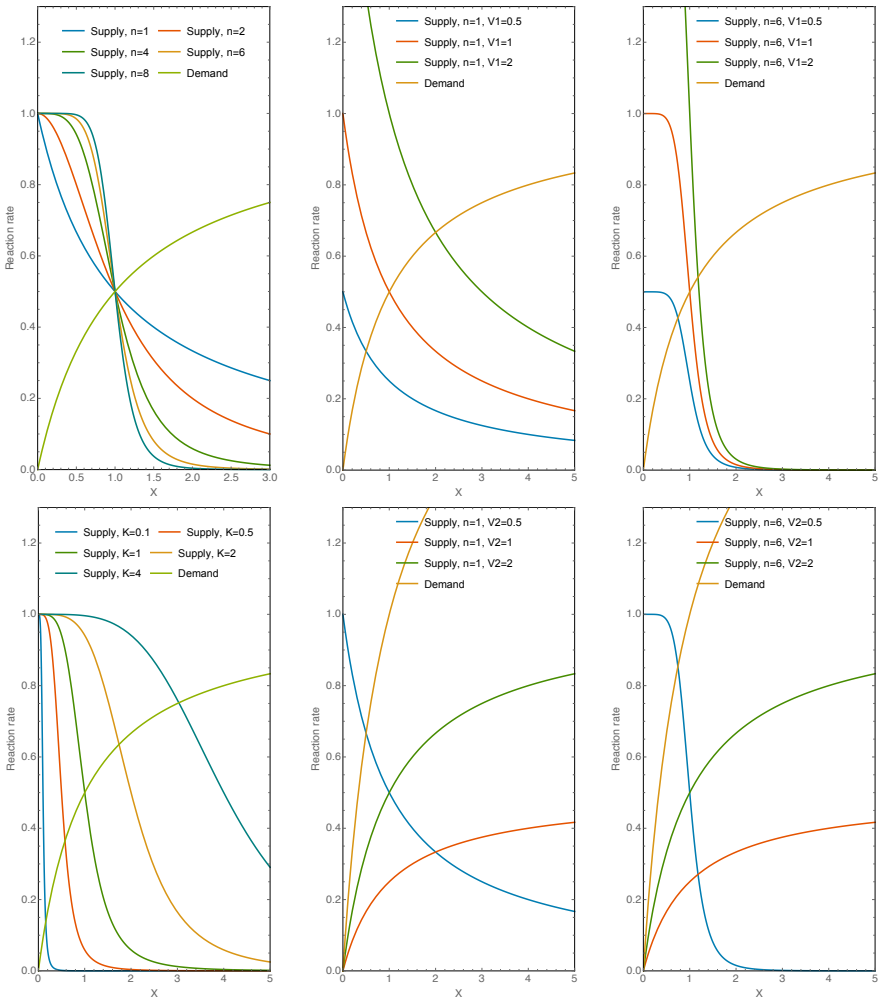
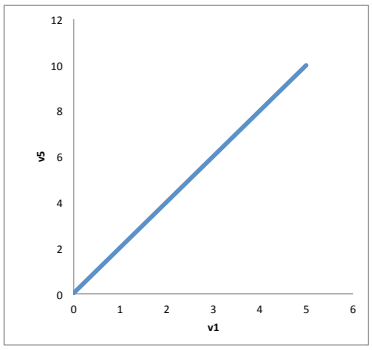
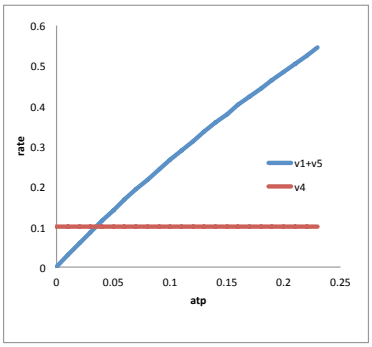
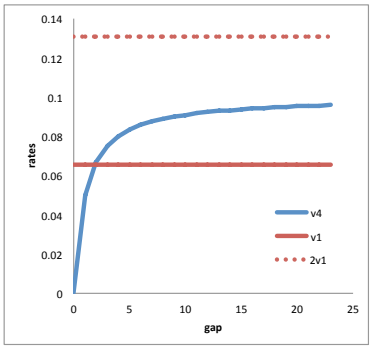


Figure 7.12: **Supply-demand model of metabolism with a negative feedback loop.**


12
10
8
6
4
2
0
v5
0
1
2
3
4
5
6
v1


0.6
0.5
0.4
0.3
0.2
0.1
0
rate
0
0.05
0.1
0.15
0.2
0.25
atp
v1+v5
v4


0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
rates
0
5
10
15
20
25
gap
v4
v1
2v1

Bibliography

[1] U Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–461, Jun 2007.

[2] E H Battley. Calculation of entropy change accompanying growth of escherichia coli k-12 on succinic acid. *Biotechnol Bioeng*, 41(4):422–428, Feb 1993.

[3] P J Choi, L Cai, K Frieda, and X S Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446, Oct 2008.

[4] R John Ellis. Macromolecular crowding: obvious but underappreciated. *Trends in biochemical sciences*, 26(10):597–604, 2001.

[5] R Fritsche-Guenther, F Witzel, A Sieber, R Herr, N Schmidt, S Braun, T Brummer, C Sers, and N Blüthgen. Strong negative feedback from erk to raf confers robustness to mapk signalling. *Mol Syst Biol*, 7:489–489, May 2011.

[6] J S Hofmeyr and A Cornish-Bowden. Regulating the cellular economy of supply and demand. *FEBS Lett*, 476(1-2):47–51, Jun 2000.

[7] H Kacser and J A Burns. The control of flux. *Symp Soc Exp Biol*, 27:65–104, 1973.

[8] H Kacser and J A Burns. The control of flux. *Biochem Soc Trans*, 23(2):341–366, May 1995.

[9] R S Kasai and A Kusumi. Single-molecule imaging revealed dynamic gpcr dimerization. *Curr Opin Cell Biol*, 27:78–86, Apr 2014.

[10] R S Kasai, K G Suzuki, E R Prossnitz, I Koyama-Honda, C Nakada, T K Fujiwara, and A Kusumi. Full characterization of gpcr monomer-dimer dynamic equilibrium by single molecule imaging. *J Cell Biol*, 192(3):463–480, Feb 2011.

[11] L Keren, O Zackay, M Lotan-Pompan, U Barenholz, E Dekel, V Sasson, G Aidelberg, A Bren, D Zeevi, A Weinberger, U Alon, R Milo, and E Segal.

Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol*, 9:701–701, 2013.

[12] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[13] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences*, 110(42):16754–16759, 2013.

[14] Daniel E Koshland, Albert Goldbeter, and Jeffry B Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, pages 220–225, 1982.

[15] O Kotte, B Volkmer, J L Radzikowski, and M Heinemann. Phenotypic bistability in escherichia coli's central carbon metabolism. *Mol Syst Biol*, 10(7):736–736, 2014.

[16] Yuri Lazebnik. Can a biologist fix a radio?—or, what i learned while studying apoptosis. *Cancer cell*, 2(3):179–182, 2002.

[17] T R Maarleveld, J Boele, F J Bruggeman, and B Teusink. A data integration and visualization resource for the metabolic network of synechocystis sp. pcc 6803. *Plant Physiol*, 164(3):1111–1121, Mar 2014.

[18] A Schwabe and F J Bruggeman. Single yeast cells vary in transcription activity not in delay time after a metabolic shift. *Nat Commun*, 5:4798–4798, 2014.

[19] B Schwanhäusser, D Busse, N Li, G Dittmar, J Schuchhardt, J Wolf, W Chen, and M Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, May 2011.

[20] B Schwanhäusser, D Busse, N Li, G Dittmar, J Schuchhardt, J Wolf, W Chen, and M Selbach. Corrigendum: Global quantification of mammalian gene expression control. *Nature*, 495(7439):126–127, Mar 2013.

[21] Y Setty, A E Mayo, M G Surette, and U Alon. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A*, 100(13):7702–7707, Jun 2003.

[22] A Solopova, J van Gestel, F J Weissing, H Bachmann, B Teusink, J Kok, and O P Kuipers. Bet-hedging during bacterial diauxic shift. *Proc Natl Acad Sci U S A*, 111(20):7427–7432, May 2014.

[23] O E Sturm, R Orton, J Grindlay, M Birtwistle, V Vyshemirsky, D Gilbert, M Calder, A Pitt, B Kholodenko, and W Kolch. The mammalian mapk/erk pathway exhibits properties of a negative feedback amplifier. *Sci Signal*, 3(153), 2010.

[24] Y Taniguchi, P J Choi, G W Li, H Chen, M Babu, J Hearn, A Emili, and X S Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, Jul 2010.

[25] J H van Heerden, M T Wortel, F J Bruggeman, J J Heijnen, Y J Bollen, R Planqué, J Hulshof, T G O'Toole, S A Wahl, and B Teusink. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114–1245114, Feb 2014.

[26] Ned S Wingreen and Kerwyn Casey Huang. Physics of intracellular organization in bacteria. *Annual review of microbiology*, 69:361–379, 2015.

[27] D Wootten, A Christopoulos, and P M Sexton. Emerging paradigms in gpcr allostery: implications for drug discovery. *Nat Rev Drug Discov*, 12(8):630–644, Aug 2013.

[28] Steven B Zimmerman and Allen P Minton. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annual review of biophysics and biomolecular structure*, 22(1):27–65, 1993.