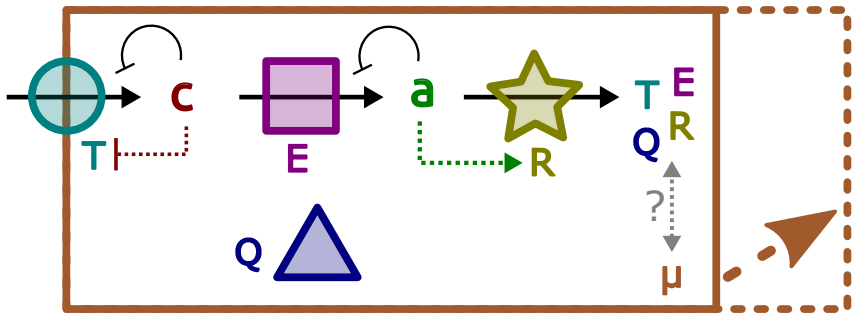
The interplay between stochasticity and regulation in a coarse-grained model of gene expression, metabolism, and growth

Istvan T. Kleijn

15th December 2015

Master's thesis in Theoretical Physics



Department of Physics and Astronomy
Faculty of Science
Utrecht University

Supervisor:
Dr Rutger Hermsen
Department of Biology

Universiteit Utrecht

Abstract

Recent time-lapse microscopy experiments on bacterial growth have shown large cell-to-cell variations in growth rate and protein expression levels. Earlier experiments at the population level have shown that the expression of different classes of protein is tightly regulated to achieve fast growth in various conditions. We have built a coarse-grained model of bacterial metabolism that incorporates both the stochasticity of protein production and division, and the regulation that optimises growth.

We introduce novel variables that quantify the coupling from gene expression to growth, which we call growth control coefficients. Analysis of the system in the optimal state shows that stochasticity in the growth rate has its main cause in the stochasticity of frequently occurring proteins, even though these fluctuate relatively little.

The global regularity of protein expression is explained by optimising the system for growth rate. We include a regulatory mechanism in our model that achieves this in close approximation. The regulation also counteracts the growth-inhibiting effects of stochasticity, by lowering the amplitude and decorrelation time of fluctuations in the protein concentrations. We use the model to compute crosscorrelations between protein expression and growth rate. The resulting graphs reproduce several features present their experimental counterparts.

Contents

1 Introduction

Growing bacteria show a fascinating contrast: the macroscopic scale of growing cultures shows striking regularity, whereas the behaviour of individual cells is highly stochastic[1]. Data on cultures in balanced exponential growth in varying conditions show that the average composition of cells is tightly regulated, so the behaviour of bacterial cultures is predictable and, usually, deterministic. However, individual bacteria are small and some of their key constituents occur only in small copy numbers. With the advent of more precise measuring techniques, research on cell-to-cell variability has increased in prominence[2]. Experiments show that the small copy numbers lead to significant amounts of noise in the biochemical networks that govern the cells. Interestingly, noise properties of the cellular composition and growth rate are nontrivially related.

We are interested in understanding the contrast between culture-wide regulation and cell-to-cell stochasticity from a modelling perspective. We will focus on cultures in steady-state, balanced, exponential growth. As I stated in the previous paragraph, the growth rate and gene expression levels are coupled in both the population average and in noise properties. It is natural to suggest that this coupling is due to the same system[3]. This is the central thesis of my thesis. We try to answer question such as:

- Can a simple, coarse-grained, model reproduce the features of both global regulation and stochasticity?
- What is the influence of global gene regulation on stochastic expression and growth?
- How can we quantify the coupling between stochasticity in protein abundance and growth rate?

In this thesis I will present my work on the coupling between noise and macroscopic laws in the context of bacterial growth. I will present a coarse-grained model of the metabolism resulting in protein production. Noise is incorporated in the final step of the metabolism, wherein the proteins are produced stochastically. The metabolic intermediates regulate the production of proteins in a way that gives rise to the observed macroscopic laws. Furthermore, I will introduce an analytic tool to quantify the dependence of the cell's growth rate on its composition, called growth control coefficients.

We will see that even large amounts of noise will leave the population laws intact, even though they may shift them a little. We will also see that the presence of noise in multiple enzymes can explain how, in a population, the most frequently occurring growth rate is not the optimal growth rate, or even the fastest growth rate that is present in the population. In addition, the model can be used to calculate cross-correlations between growth rate, protein production, and protein abundance. We will compare these functions to recently measured data on *E. coli*.

In the rest of this chapter, I will explain the experiments that lead up to our research more thoroughly. I will start by describing the work on the culture level, followed by the experiments on stochastic gene expression and growth. I will close the chapter with a short description of our modelling approach.

1.1 Macroscopic bacterial growth laws

The regularity of bacteria in balanced exponential growth has been quantified thoroughly. There are many empirical relationships that express physiological parameters as functions of the growth rate. These relationships are usually called *laws*.

The first example of an empirical law is the Monod law, named after its discoverer. Monod observed bacterial populations that were limited by the availability of a nutrient source. He found that the growth rate μ depended on the concentration $[N]$ of the nutrient[4]

according to a simple asymptotic relationship:

$$\mu \propto \frac{[N]}{K + [N]}$$

where K is a constant that depends on the type of bacteria and nutrient.

Another important observation was the dependence of cell size on growth[5]. When bacteria grow fast, they will be larger on average. The dependence is exponential:

$$\log \Omega \propto \mu$$

where Ω denotes the cell size and μ the growth rate.

More recent experiments have uncovered linear relations between the proteome composition of cells and their growth rate. At first, the analysis was limited to two types of protein: ribosomal proteins and catabolic proteins [6]. Upon limiting the catabolic proteins, by varying the source of carbon, the expression of catabolic proteins increased while the expression of ribosomal proteins decreased. Upon limiting the ribosomal proteins, by adding varying amounts of ribosome-inhibiting antibiotics, the expression of catabolic proteins decreased while the expression of ribosomal proteins increased. Expressed as a function of the growth rate, the relative expression levels were linear. We will refer to these linear relationships as the growth laws.

These two protein sectors were not enough to cover the entire proteome. About half of the cell consists of proteins that do not change in expression level with growth rate. In follow-up experiments the analysis was first expanded to include a separate growth law for anabolic proteins[7]. The most recent experiments used comprehensive mass spectrometry to classify hundreds of proteins in different classes, according to the trends in their expression under different types of limitation[8]. All of the relative expression levels are linear functions of growth rate.

The growth laws relating cellular composition and growth rate are understood in terms of optimal resource allocation. For example, if the cell is limited by its nutrient supply, it needs to produce more catabolic proteins. This comes at the cost of producing ribosomes, so the growth rate decreases.

This discussion on macroscopic laws highlights that physiological state variables and gene expression levels are linked, at least in the population-averaged sense. In the next section, we will see that the same general statement holds for single cells.

1.2 Single-cell stochastic gene expression

Advances in fluorescence microscopy have made it possible to observe intracellular stochasticity in real-time. Cell strains can be engineered to express small fluorescent proteins (usually GFP). Fluorescent proteins can be expressed on their own (to study expression noise in general or in engineered networks[9, 10]), at the same time as another protein (to study the expression noise of that protein[11]), or they can be fused to non-fluorescent proteins (to study protein dynamics[12]).

Recently, Kiviet *et al.* used the second of these techniques in combination with optical microscopy to obtain stochastic data of both cell size (and thence growth rate) and *lac* protein abundance (and thence production rate)[13]. The fluctuations in these quantities are very large and correlated among themselves in interesting ways. The experimenters can control the growth rate by varying the average expression of *lac* proteins, through inducing production by the chemical IPTG[1].

On average, it is observed that upward fluctuations in the growth rate follow production of *lac* enzymes. The time scale, relative to the cell division time, of the production versus

[1]This induction usually occurs by way of lactose, when cells grow on this sugar. In the experiments, the cells grow on lactulose. This lactose derivative can be catabolised by the *lac* enzymes, but it does not induce their production.

growth rate cross-correlation decreases with growth rate. The cross-correlation becomes less skewed as well. The cross-correlation between the concentration of *lac* enzymes and growth rate is symmetrical at low growth rate. At high growth rate, upward fluctations in the growth rate are preceded by high *lac* enzyme concentrations in some systems, and by low *lac* enzyme concentrations in other system. The authors explain this by the interplay between "common noise" that couples to both *lac* enzyme production and growth, and "dilution" decreasing all enzyme concentrations when the cell grows fast.

1.3 Approach

To understand the interplay between the growth laws and the stochastic gene expression, we will make a coarse-grained model, inspired by metabolic regulation at the transcription/translation level. Protein sectors will be represented by single representative enzymes, one type per sector. We keep the regulation mostly phenomenological. Stochasticity is included in the processes of cell division and protein production.

In chapter 2, I will describe the model in full detail. Chapter 3 contains an analysis on the system when it is optimised for growth rate. Chapter 4 covers the results of simulations where we regulated the system manually. Chapter 5 finally displays the results on simulations on the fully regulated system. I will close this thesis with a discussion on the results of my research, and suggest avenues of further research.

2 Description of the model

In this chapter I will describe our model in terms of its separate components. In section 2.1 I will explain the metabolism. Section 2.2 covers the relations between growth and production, and section 2.3 treats the process of cell division. Section 2.4 introduces noise in protein production. I will finally introduce the system's regulation in section 2.5. I will start, however, with some preliminary remarks.

As noted before in section 1.1, the macroscopic growth laws are formulated in terms of a coarse-grained model that divides the bacterial proteome into a small number of protein sectors. Inspired by this, we will investigate four classes of protein: "housekeeping" proteins, catabolic proteins, anabolic proteins and ribosomal proteins. We represent each class as a single enzyme in a minimal metabolic model.

We know that the fractional abundance of about half of the proteome does not change with growth rate. In the model, we represent this "housekeeping" sector by a protein, called "Q". This protein does not function as an enzyme in metabolism.

The first metabolically active protein is an enzyme that transports nutrient molecules from the environment of the cell to the inside. It represents the class of catabolic enzymes and is denoted "T", for transporter. The product of its reaction is a molecule that we call the catabolite, denoted with the letter "c".

The second metabolic protein is an enzyme that transforms the catabolite into amino acids. This protein represents the class of anabolic enzymes and is denoted by the letter "E". We denote the amino acids with the letter "a".

The third metabolic protein represents the ribosomes, that is, ribosome are also modelled as single proteins. The letter "R" denotes the ribosomal protein. This enzyme consumes amino acids and directly produces all proteins T, E, R, and Q.

A cartoon of the metabolism is pictured in Figure 2.1. We use the following notation convention: capitals in italic (T, E, R, and Q) refer to the number of proteins of their respective protein class in a cell.

The (buoyant) density of bacterial cells is remarkably constant between genetically identical cells in the same and even different growth conditions[14]. This suggests, that it is justifyable to equate the total protein abundance in a cell with cell size. We denote the cell size as $\Omega = T + E + R + Q$. For protein class X, the protein fraction $\phi_X = \frac{X}{\Omega}$ can then be interpreted as the concentration.

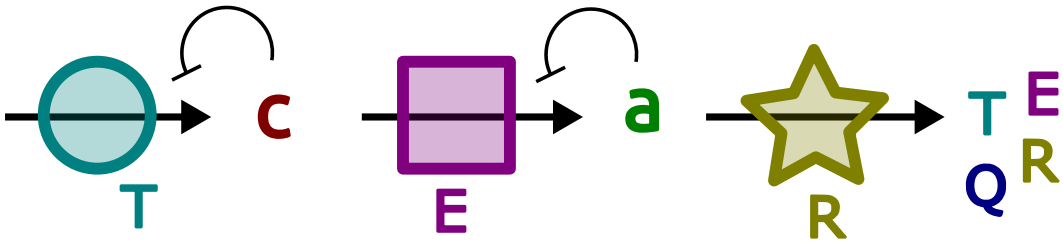


Figure 2.1: Cartoon of the metabolic model. Thick, black arrows indicate chemical reactions; thin, black lines with barred ends denote product inhibition. Differently coloured letters indicate different molecular species: capitals denote proteins, lower case letters denote metabolic intermediates. Food from outside the cell is imported by a transporter protein "T" and converted into the first metabolic intermediate "c". This is converted by the enzyme "E" into the second intermediate "a", that is in turn converted by the ribosome "R" into the four proteins themselves (including the housekeeping proteins "Q", that do not feature in the metabolism). The colour coding from this figure is used throughout this thesis: cyan for transporters, purple for enzymes, yellow for ribosomes, navy for housekeeping proteins; red for catabolic intermediates, and green for amino acids.

2.1 Metabolism

We assume that, for a given protein composition, the cellular metabolism reaches its equilibrium fast. This allows us to model it deterministically, using ordinary differential equations. The metabolism is assumed to be in quasi-steady state throughout protein production.

Given a cell with a certain composition (in terms of the protein numbers T, E, R, and Q), we will calculate the steady-state concentrations of catabolic intermediate molecules and amino acids in the system. These concentrations will determine the rates at which the ribosome produces new proteins.

Box 2.1: Michaelis-Menten kinetics

The most basic enzymatic reaction mechanism is called the Michaelis-Menten mechanism. This mechanism considers three molecules: E, S, and P. The molecule E functions as an enzyme, transforming the substrate S into the product P. When E binds S, they form a complex C. This complex can disintegrate back into E and S, or the reaction can progress to form E and P. The product cannot bind to the complex again. In a reaction scheme:

$$E + S \rightleftharpoons C \rightarrow E + P \tag{2.1}$$

Using this enzymatic reaction instead of linear kinetics means that the enzyme limits the reaction. The substrate can saturate the enzyme, so the reaction rate will remain bounded even if the substrate concentration is very large.
Defining the total turnover rate v of the reaction as the rate at which molecules of P are produced, the following equation can be derived[15]:

$$v = \frac{\mathrm{d}P}{\mathrm{d}t} = -\frac{\mathrm{d}S}{\mathrm{d}t} = k(E + C)\frac{[S]/M}{1 + [S]/M} \equiv kE_{\mathrm{tot}}\frac{s}{1 + s} \tag{2.2}$$

where $[S]$ denotes the concentration of the substrate and P, S, E, and C the number of molecules in the system of product, free substrate, free enzyme and complex, respectively. The total number of enzymes is $E_{\mathrm{tot}} = E + C$. The system has two parameters: the Michaelis constant M and the rate constant k. We can rescale the substrate concentration to become dimensionless by dividing by the Michaelis constant: $s = \frac{[S]}{M}$. The production rate of P molecules then depends only on the total number of enzymes and the dimensionless substrate concentration.

The three metabolically active proteins function as enzymes. A straightforward way of modelling enzymatic reactions is by assuming Michaelis-Menten kinetics, described in Box 2.1. This approach was used in [16] as part of a deterministic model. However, when we applied this mechanism to stochastic systems, some problems emerged, which we will describe below. We will start by studying the Michaelis-Menten approach anyway, because it is instructive to see the analysis proceed for this relatively simple scheme. In our more complicated model, the steps are not easily written down.

We will calculate the metabolic state of the system when the protein numbers T, E, and R are given. Note that the Q proteins do not contribute to metabolism, so we do not need to specify Q. Let us first assume that the proteins T, E, and R function as Michaelis-Menten enzymes, with rate constants k_{T}, k_{E}, and k_{R}, respectively. The total turnover rates of transporters, enzymes, and ribosomes are

$$v_{\mathrm{T}} = k_{\mathrm{T}}T\frac{\sigma}{1 + \sigma} \tag{2.3}$$

$$v_{\mathrm{E}} = k_{\mathrm{E}}E\frac{c}{1 + c} \tag{2.4}$$

$$v_{\mathrm{R}} = k_{\mathrm{R}}R\frac{a}{1 + a} \tag{2.5}$$

We consider the metabolic intermediates in terms of their rescaled concentrations $c = \frac{[c]}{M_{\mathrm{E}}}$

and $a = \frac{[a]}{M_\mathrm{R}}$; we want to express these variables as a function of the protein copy numbers. The external nutrient concentration is a parameter. Just as the concentrations of the metabolic intermediates, we divide by the Michaelis constant of the reaction that consumes it. σ denotes the rescaled nutrient concentration.

In metabolic steady state the turnover rates at all enzymes must be equal. The functional form of the metabolite concentrations $c(T, E, R)$ and $a(T, E, R)$ must reflect that $v_\mathrm{T} = v_\mathrm{E} = v_\mathrm{R}$. Solving these equations for c and a yields them as functions of the protein numbers:

$$c = \frac{k_\mathrm{T} T \frac{\sigma}{1+\sigma}}{k_\mathrm{E} E - k_\mathrm{T} T \frac{\sigma}{1+\sigma}} \tag{2.6}$$

$$a = \frac{k_\mathrm{T} T \frac{\sigma}{1+\sigma}}{k_\mathrm{E} R - k_\mathrm{T} T \frac{\sigma}{1+\sigma}} \tag{2.7}$$

Notably, c and a have finite positive values only when $k_\mathrm{E} E > k_\mathrm{T} T \frac{\sigma}{1+\sigma}$ and $k_\mathrm{R} R > k_\mathrm{T} T \frac{\sigma}{1+\sigma}$. In the case where the inequalities do not hold, no metabolic steady state exists. In other words, steady state exists only if

$$T < \frac{k_\mathrm{E}}{k_\mathrm{T}} \frac{1+\sigma}{\sigma} E \tag{2.8}$$

$$\text{and} \quad T < \frac{k_\mathrm{R}}{k_\mathrm{T}} \frac{1+\sigma}{\sigma} R \tag{2.9}$$

are both true. We can interpret this limitation in the following way, beginning with the first one. Suppose that $k_\mathrm{T} T \frac{\sigma}{1+\sigma} > k_\mathrm{E} E$. This means that the transporter T is pumping catabolite molecules into the cell faster than the enzymes E can handle even when they are functioning at maximum efficiency. The concentration of catabolites will thus diverge and no steady state can exist. The turnover rate of the transporter will always be larger than the turnover rate of the enzyme.

In equations: because $\frac{c}{1+c} < 1$ and $k_\mathrm{E} E < k_\mathrm{T} T \frac{\sigma}{1+\sigma}$, $v_\mathrm{E} < v_\mathrm{T}$ for all $(T, E) \in \mathbb{N}^2$. Likewise, when $k_\mathrm{R} R < k_\mathrm{T} T \frac{\sigma}{1+\sigma}$, $v_\mathrm{R} < v_\mathrm{T}$ for all $(T, R) \in \mathbb{N}^2$ because $\frac{a}{1+a} < 1$.

Leaving these considerations aside for a moment, we will now consider the way in which this metabolism occurs in real life. Real cells are constantly competing with each other for environmental resources. A faster growing cell has an advantage over its competitors, because it will claim resources more quickly than them. Over eons of natural selection, bacteria have become optimised for growth rate and any model of real cells must be very close to the optimally growing state. In our model, we would like to consider this ideal state as the baseline, and consider stochastic fluctuations around the optimum.

It is obvious that cells with a faster metabolism (higher turnover rates) will grow faster, because they produce their constituents at a higher rate. In fact, in our model, the two are directly related (see section 2.2). According to equation (2.3), the turnover rate in the Michaelis-Menten model is directly proportional to T—an optimal cell will have as many transporters as possible.

When we return to the inequalities (2.8) and (2.9), we see that the number of transporters is limited. This conflicts with the desire to maximise T. The best we can do, while keeping the Michaelis-Menten kinetics intact, is choosing the system suboptimally with high, but finite, metabolite concentrations. This, however, leads to its own problems when we introduce stochasticity. Due to noise in protein production, the protein concentrations will not always have their average values. Because of the fluctuations in protein abundances, it is possible that the inequalities (2.8) and (2.9) are violated sometimes. This becomes more frequent the closer we choose the average composition to the optimum.

All this suggests that we have to modify the metabolic kinetics of the model to incorporate these points. An analysis of real metabolic networks suggests that the aforementioned problems are encountered by real cells as well as our simple model. These networks contain several types of metabolic feedback loops that keep metabolite pools finite. Examples

include product inhibition, where an end product inhibits the first enzyme of the pathway that produces it, leakage of metabolites from the cytoplasm to the cell's environment, and reactions that can be catalysed by enzymes working in both directions. An overview of the metabolism of *E. coli* can be found on the web[17].

A general theory of metabolic feedback has been developed by Savageau[18]. Unfortunately, this theory is not fully applicable to our work, because it does not account for the allocation of finite enzyme production resources. As the most important feature of the feedback mechanism is the reduction of metabolite pools, we choose an extension of the Michaelis-Menten model for which we do not need to introduce any new proteins to the model: product inhibition. The differences between the simple ichaelis-Menten model and the one including product inhibition are explained in Box 2.2.

Box 2.2: Product inhibition

This extension of the Michaelis-Menten kinetics ensures that the end product of an enzymatic reaction inhibits the reaction that produces it directly. It does so by binding to the enzyme. The Michaelis-Menten reaction scheme (2.1) is extended to[15]

$$\begin{aligned} E + S &\rightleftharpoons C_1 \rightarrow E + P \\ E + P &\rightleftharpoons C_2 \end{aligned}$$

Note that there are two possible enzyme complexes. C_1 forms when E binds S, and C_2 when E binds P. Once a product molecule P is formed, it cannot be turned back into a substrate molecule S. In this reaction scheme, the turnover rate becomes

$$v = k(E + C_1 + C_2)\frac{[S]/M_\mathrm{S}}{1 + [S]/M_\mathrm{S} + [P]/M_\mathrm{P}} \equiv kE_\mathrm{tot}\frac{s}{1 + s + p} \tag{2.10}$$

The only difference with the Michaelis-Menten turnover rate (2.2) is the presence of the suitably rescaled product concentration in the denominator. There are three parameters per reaction, namely the reaction rate k and two constants M_S and M_P functioning in the same way as the Michaelis constant.

The reactions that produce c and a are inhibited by their respective products. This is not true for the reaction that produces proteins. In this model, the turnover rates of T, E, and R are modified from equation (2.3) to

$$v_\mathrm{T} = k_\mathrm{T} T \frac{\sigma}{1 + \sigma + \beta_\mathrm{c} c} \tag{2.11}$$

$$v_\mathrm{E} = k_\mathrm{E} E \frac{c}{1 + c + \beta_\mathrm{a} a} \tag{2.12}$$

$$v_\mathrm{R} = k_\mathrm{R} R \frac{a}{1 + a} \tag{2.13}$$

where the parameters $\beta_\mathrm{c} = \frac{M_\mathrm{E,c}}{M_\mathrm{T,c}}$ and $\beta_\mathrm{a} = \frac{M_\mathrm{R,a}}{M_\mathrm{E,a}}$ are the fractions between the concentration constants associated with each of the two metabolites.

As before, the equations $v_\mathrm{T} = v_\mathrm{E} = v_\mathrm{R}$ can be solved for c and a. Although this amounts to a third-order linear equation, the closed form is convoluted and uninformative. However, we can readily see that the turnover rate in the product-inhibited scheme depends on all enzymes, not just the transporter. This is due to the presence of at least one of $c(T, E, R)$ and $a(T, E, R)$ in each of the three expressions (2.11), (2.12), and (2.13). The dependence of the turnover rate (and hence the growth rate) on the different protein abundances will be quantified in Section 3.3.

2.2 Intensive and extensive variables

Up until now we have talked about the total turnover rate per reaction in the entire cell. The resulting turnover rate of the ribosome is the change of total protein number over time. We assume that the proteins in our model are stable for several cell cycles, so we ignore protein degradation. Recall that we the total protein number equates to the cell size: $\Omega = T + E + R + Q$. In the metabolic steady state then, we can define the metabolic flux $J = v_{\mathrm{T}} = v_{\mathrm{E}} = v_{\mathrm{R}}$, for which

$$J = \frac{\mathrm{d}\Omega}{\mathrm{d}t}$$

Note that the metabolic flux scales linearly with the size of the cell, because the total turnover rates are proportional to the enzyme numbers. Because J scales with the size of the system, we can call it an *extensive* quantity, in analogy to thermodynamics.

When a culture reaches balanced exponential growth, its composition in terms of concentrations does not change with the size of the population any more—a state of exponential growth calls for *intensive* variables, that are constant with varying system size. In balanced exponential growth, the logarithm of the cell size increases linearly with time. We can introduce the instantaneous growth rate μ as

$$\mu = \frac{\mathrm{d}\log\Omega}{\mathrm{d}t} = \frac{1}{\Omega}\frac{\mathrm{d}\Omega}{\mathrm{d}t} = \frac{J}{\Omega}$$

which is simply the flux divided by the cell size. As the flux is an extensive quantity, the growth rate must be an intensive quantity.

It is insightful to note here that the equations $J = v_{\mathrm{T}} = v_{\mathrm{E}} = v_{\mathrm{R}}$ can be rewritten to $\mu = \frac{v_{\mathrm{T}}}{\Omega} = \frac{v_{\mathrm{E}}}{\Omega} = \frac{v_{\mathrm{R}}}{\Omega}$. Dividing Equations (2.11), (2.12), and (2.13) by Ω amounts to changing X to ϕ_{X} for $X \in \{T, E, R\}$. Equating these expressions and solving them for a and c works exactly as before, so the functional forms of $(c, a)(T, E, R)$ and $(c, a)(\phi_{\mathrm{T}}, \phi_{\mathrm{E}}, \phi_{\mathrm{R}})$ must be exactly equal. In other words, the metabolite concentrations and the growth rate μ depend only on the ratios of the protein abundances.

The state of the system can be expressed in terms of the protein fractions ϕ_{X} alone, which are subject to the constraint that

$$\sum_{\mathrm{X}} \phi_{\mathrm{X}} = 1 \tag{2.14}$$

We will now relate the protein fractions and their constraint to the protein production rates in steady state.

When the system is at steady state, production rates of all proteins are constant. Recall that the total turnover rate of the ribosomes is $v_{\mathrm{R}} = k_{\mathrm{R}} R \frac{a}{1+a}$. We assume that this rate of protein production is split between the different proteins, such that the abundance of protein X changes over time as

$$\frac{\mathrm{d}X}{\mathrm{d}t} = r_{\mathrm{X}} R \frac{a}{1+a} \tag{2.15}$$

with constant

$$k_{\mathrm{R}} = \sum_{\mathrm{X}} r_{\mathrm{X}} \tag{2.16}$$

In steady-state growth, the amount of each protein type grows exponentially with the *same* growth rate μ and their fractional abundances remain constant. We may write

$$\frac{\mathrm{d}X}{\mathrm{d}t} = X_0 \mathrm{e}^{\mu t} = \phi_{\mathrm{X},0}\Omega_0 \mathrm{e}^{\mu t} = \phi_{\mathrm{X}}\Omega_0 \mathrm{e}^{\mu t} \tag{2.17}$$

with X_0 the number of X proteins, $\Omega_0 = \sum_{\mathrm{X}} X_0$ the cell size, and $\phi_{\mathrm{X},0} = \frac{X_0}{\Omega_0}$ the fractional abundance of X, all at time $t = 0$. In steady state, $\phi_{\mathrm{X},0} = \phi_{\mathrm{X}}$ at all times.

We can now consider two types of protein, X and Y. We can express the ratio of the time derivatives of their total abundances using both equation (2.15) and equation (2.17):

$$\frac{\frac{\mathrm{d}X}{\mathrm{d}t}}{\frac{\mathrm{d}Y}{\mathrm{d}t}} = \frac{r_\mathrm{X}}{r_\mathrm{Y}} = \frac{\phi_\mathrm{X}}{\phi_\mathrm{Y}}$$

This means that, in the steady state, the proteome fraction and the production rate differ only by a constant. In other words:

$$\phi_\mathrm{X} = \frac{r_\mathrm{X}}{k_\mathrm{R}}$$

and the constraints (2.14) and (2.16) are different sides of the same coin.

2.3 Cell division

Of course, the model cell cannot grow indefinitely—real cells undergo a cell cycle at the end of which they divide into two daughter cells. We need to include the mechanism that keeps the cell size small, the mechanism that produces new cells: cell division. There are two components to the process of cell division: when it occurs and what the result is. Also, as noted in Section 1.1, the average size of unicellular organisms scales exponentially with the growth rate. This effect is important for our model, because the size of a cell influences the amplitude of noise in the system. A large cell is relatively unaffected by noise, because the fluctuations are small compared to the average. A small cell feels the effects of noise strongly. Because slow-growing cells are smaller, we expect that the effects of noise are more apparent at slow growth. We must therefore include the exponential dependence of average cell size on growth rate.

There are many different types of models that explain the exponential dependence, all using different sensors that organisms may use to determine progress along their way from division to division. For example, cells may have a way to measure their current size, or the time that has passed since the last division. Single-cell data [19, 20] suggest a model[21][2] where cells are regulated to add an amount $\hat{\Delta}$ to the initial size before starting division. This is called an "adder" model. In *E. coli* this model works well, but different unicellular organisms may be wired differently[22] with correspondingly different models[23].

Unfortunately, a simple implementation of the adder model failed to be stable in our simulations. Because the exact way of implementing the exponential dependence of cell size on growth rate does not matter very much for the noise properties of the protein abundances, we chose to implement a so-called "sizer" model, where the cell senses its own size and divides when it has reached a size $\hat{\Delta}$. We let T_D denote the time scale in the exponent and Δ (without the hat) the cell size extrapolated to zero growth rate. Empirically, T_D is about equal to the time DNA takes to replicate, which in *E. coli* is about one hour[3]. We leave noise out of the picture here and simply divide the cell when it reaches a size $\Delta \exp(\mu T_\mathrm{D})$.

Note that the growth rate needs to be known in order to do this—we use the optimal growth rate from section 3.1. How the cell would know its own growth rate so it could regulate this is interesting in its own right, but beyond the scope of this thesis. It is likely linked to the precise choice of the sensor that the cell uses to determine the start of division, and should take the details of the metabolism into account[25, 26].

Now we have described when in our model cell division occurs, it remains to describe what the process actually entails. When the cell size is larger than the threshold, each protein has a probability of $\frac{1}{2}$ to end up in daughter cell 1, and equal probability to end up in daughter cell 2. This amounts to a binomial distribution of proteins between daughter

[2]The authors of this paper state that this model was first put forward by Voorn *et al.*, "Mathematics of cell division in *Escherichia coli*", *Curr Top Mol Genet* 1:187-194, 1993, but I have not been able to confirm this.

[3]Note that cells can double much faster than once per hour, which means that fast-growing cells will contain DNA for multiple generations to come![24]

cells. We now choose to continue our analysis with one of the two daughters. We always choose daughter cell 1. [4]

2.4 Stochasticity of protein production

In addition to the binomial distribution at cell division, stochasticity is introduced in the model at protein production. In this section I will explain how this can be described as a stochastic process, and how we simulate it.

If proteins X are produced at the rate r_{X}, one at a time, we can describe the stochasticity in a chemical master equation (CME) [27]

$$\dot{P}(T,E,R,Q,t) = \frac{a}{1+a}\begin{pmatrix} r_{\mathrm{T}}(\mathbb{E}_{\mathrm{T}}^{-1}-1)+ \\ r_{\mathrm{E}}(\mathbb{E}_{\mathrm{E}}^{-1}-1)+ \\ r_{\mathrm{R}}(\mathbb{E}_{\mathrm{R}}^{-1}-1)+ \\ r_{\mathrm{Q}}(\mathbb{E}_{\mathrm{Q}}^{-1}-1) \end{pmatrix} RP(T,E,R,Q,t) \tag{2.18}$$

where $\mathbb{E}_{\mathrm{X}}$ is the step operator for protein species X, defined as $\mathbb{E}_{\mathrm{X}}^{-1}f(X) = f(X-1)$. Note that we are basically describing four different processes in one equation, namely ribosomes R producing T, E, R, and Q. The ribosomes work only at a fraction $\frac{a}{1+a}$ of their maximal rate and each active ribosome is producing X-proteins at a rate r_{X}.

Most chemical M-equations have not been solved for nontrivial boundary conditions[5], although some progress has been made for single enzymes[28]. Equation (2.18) is no exception and we have been unable to make progress in an analytical sense.

However, it is possible to simulate the CME with a Gillespie algorithm [29]. This works in the following way[30]:

1. Compute the rates, also called propensities, at which the reactions occur. For the production of protein X, this rate is

$$p_{\mathrm{X}} = r_{\mathrm{X}}\frac{a}{1+a}R \tag{2.19}$$

   Denote $p_{\mathrm{tot}} = \sum_{\mathrm{X}} p_{\mathrm{X}}$ for the rate at which any reaction occurs. Index the reactions as $i_{\mathrm{T}} = 1$, $i_{\mathrm{E}} = 2$, $i_{\mathrm{R}} = 3$, and$i_{\mathrm{Q}} = 4$ and denote $p_{i_{\mathrm{X}}} \equiv p_{\mathrm{X}}$.

2. Draw two random numbers y_1, y_2 from the uniform distribution on the unit interval[31].

3. Compute the time δt until the next event occurs. One should draw from an exponential waiting time distribution, so with the first random number given, the waiting time is

$$\delta t = -\frac{\ln(1-y_1)}{p_{\mathrm{tot}}}$$

4. Determine which of the events occurs. The production of one molecule of X occurs when

$$\sum_{i=1}^{i_{\mathrm{X}}-1} p_i < p_{\mathrm{tot}}y_2 < \sum_{i=1}^{i_{\mathrm{X}}} p_i$$

   which is equivalent to saying that the probability for the production of X to occur is $\frac{p_{\mathrm{X}}}{p_{\mathrm{tot}}}$.

[4] This strategy does not guarantee agreement between the statistics of the lineage and the statistics of the population at a particular time point, see chapter 6.

[5] A trivial boundary condition would be $P(T,E,R,t) \propto \delta(R)$ such that $\dot{P}(T,E,R,t) = 0$.

In reality, it is known that the production of proteins does not occur one by one as a Poisson process, but in bursts. This is due to the fact that a gene is first transcribed into a strand of RNA before that strand is translated several times by ribosomes. The RNA transcripts have a short half life, on the order of minutes. It has been experimentally observed that the burst size is geometrically distributed[32].

We do not model transcription explicitly. Instead, we let protein production occur in bursts of varying size.[6] The burst size b is drawn from a geometric distribution. The burst size does not depend on the type of protein—the only variables that differ between proteins are the rates r_{X}.

This means that the simulation procedure outlined above is changed in two ways. The first one is, that step 4 now produces n molecules of type X instead of one, where $n \geq 1$ is distributed geometrically with mean b. The second, and more subtle one, is as follows. Because each reaction produces on average b copies, each reaction has to occur only $\frac{1}{b}$ times as often so as to result in the same average production rate. So the formula (2.19) for computing the propensities is modified to

$$p_{\mathrm{X}} = \frac{r_{\mathrm{X}}}{b} \frac{a}{1+a} R$$

2.5 Gene regulation

Cells do not have a fixed composition. Rather, they adapt to their environment by adjusting their gene expression levels to the conditions that they find themselves in. The regulation of gene expressions typically needs a biochemical network consisting of several additional proteins. One protein will sense some property of the system, typically by binding an intracellular or extracellular molecule, another will bind to that protein, and so on, ending in interactions between a protein and the transcription or translation machinery.

In our model, we use a course-grained description of gene regulation, where the production rates of the proteins are functions of the two metabolic intermediate concentrations. Because the state of the metabolic network is determined by only three proteins, with the additional constraint that their fractional abundances (corresponding to their concentrations) must add up to a constant, two regulatory variables (c and a) should be sufficient to specify the metabolic state.

The mechanism that we use for the regulation of transporter production was described by You *et al.*[7] They also gave evidence for this mechanism of regulation. A small molecule, called cyclic AMP (or cAMP), is responsible for the regulation of catabolic enzymes, by sensing the concentrations of small metabolic precursors. A high precursor concentration inhibits the production of cAMP. The molecule binds to a regulatory protein, called CRP. The presence of CRP-cAMP complexes stimulates the production of catabolic enzymes at the transcript level. The net result of this is, that high concentrations of metabolic precursors inhibit the production of catabolic enzymes, and thereby their own production.

In our model, this scheme is represented by making the production of the transporter dependent on the concentration c of the first metabolic intermediate, which is the one that is produced by the transporter.

The mechanism that we use for the regulation of ribosome production is likewise inspired on a known molecular regulation mechanism. When the pool of amino acids that are available for protein synthesis is small, there will be many uncharged tRNAs present[7]. When these tRNAs bind to the ribosome, protein production stalls. This sends a signal to increase the production of a signalling molecule, ppGpp. This molecule inhibits the transcription of ribosomal RNA and thereby the production of ribosomes. In summary, a small amino

[6] I like to call my bursty model by the names of its metabolites a and c, and its proteins T, E and R, which becomes b-acTER.

[7] tRNA molecules present amino acids to ribosomes. An uncharged tRNA is a tRNA molecule that does not have an amino acid bound to it.

acid pool inhibits the production of ribosomes. The mechanism was studied in detail by Bosdriesz *et al.[33]* in the context of optimal ribosome synthesis.

Our model again glosses over the molecular details and we simply make the production of the ribosomal proteins depend on the concentration a of the amino acids—a large amino acid pool will stimulate the production of ribosomal proteins.

This general scheme of gene regulation was analysed by Scott *et al.*[34] They show that this way of coordinating gene expression with nutrient conditions is able to produce the growth laws from section 1.1. They also show that the occurrence of the growth laws is insensitive to the precise shape of the regulatory functions.

Nonetheless, the regulation must take into account the constraints that we have encountered so far. In particular, we will assume that the maximal production rate r_{X} of protein X is given by some fraction $F_{\mathrm{X}}(c,a)$ of the total turnover rate k_{R} of one ribosome:

$$r_{\mathrm{X}} = k_{\mathrm{R}} F_{\mathrm{X}}(c,a) \tag{2.20}$$

where

$$\sum_{\mathrm{X}} F_{\mathrm{X}}(c,a) = 1 \tag{2.21}$$

As a result, $\phi_{\mathrm{X}} \to F_{\mathrm{X}}(c,a)$ when c and a remain constant for a long time.

The housekeeping proteins are not regulated explicitly. Their production rate is constant, irrespective of the growth conditions:

$$r_{\mathrm{Q}} = k_{\mathrm{R}} \phi_{\mathrm{Q}} \tag{2.22}$$

For the metabolic proteins, we can use the following expressions to ensure that equation (2.21) holds, while using c to regulate T directly and using a to regulate R directly:

$$F_{\mathrm{T}} = (1-\phi_{\mathrm{Q}}) \frac{f_{\mathrm{T}}(c)}{f_{\mathrm{T}}(c) + 1 + f_{\mathrm{R}}(a)} \tag{2.23}$$

$$F_{\mathrm{E}} = (1-\phi_{\mathrm{Q}}) \frac{1}{f_{\mathrm{T}}(c) + 1 + f_{\mathrm{R}}(a)} \tag{2.24}$$

$$F_{\mathrm{R}} = (1-\phi_{\mathrm{Q}}) \frac{f_{\mathrm{R}}(a)}{f_{\mathrm{T}}(c) + 1 + f_{\mathrm{R}}(a)} \tag{2.25}$$

with $f_{\mathrm{T}}(c)$ and $f_{\mathrm{R}}(a)$ yet to be determined regulation functions. Because a large pool of catabolites inhibits the production of transporters, f_{T} must be monotonically decreasing. Likewise, f_{R} must be monotonically increasing, because a large pool of amino acids stimulates the production of ribosomes.

An important feature of this type of regulation, incorporating the constraints, is what we may call *implicit* regulation. If, for example, the production rate of transporters increases due to a small catabolite pool, the production rates of enzymes and ribosomes decreases. This is due to the fact that the machinery that is now producing transporters is not available for the production of enzymes and ribosomes. In the expressions(2.23), (2.24), and (2.25), it shows in an increase in the denominator.

We will choose the following regulation functions:

$$f_{\mathrm{T}}(c) = \left(\frac{\theta_{\mathrm{T}}}{c}\right)^{n_{\mathrm{T}}}$$

$$f_{\mathrm{R}}(a) = \gamma_{\mathrm{R}} \frac{a^{n_{\mathrm{R}}}}{\theta_{\mathrm{R}}^{n_{\mathrm{R}}} + a^{n_{\mathrm{R}}}}$$

Here θ_{T} and θ_{R} are parameters that describe the range of the (dimensionless) concentrations, n_{T} and n_{R} are parameters that describe the steepness of the functions and γ_{R} is a parameter that describes the maximum expression of ribosomes.

This particular choice is made because it appears that these shapes optimise the growth rate. We will describe the process of choosing these functions and parameters in more detail in section 5.1. The shape of $f_{\mathrm{R}}(a)$ is called a Hill function and it is very common as a regulation function, due to its sigmoidal shape. Power laws such as $f_{\mathrm{T}}(c)$ are less common, because they are not bounded, although they are generally used to model regulation functions if the range of possible concentrations is small[18]. In our model, the actual production rate r_{T} of the transporters is bounded, even though f_{T} may not be, because of the presence of f_{T} in both the numerator and the denominator of (2.23).

I have now described every component of our model. For a graphical representation of the general idea, see Figure (2.2). The parameters that we have used in our simulations, which I will discuss in the rest of this thesis, are summarised in Table (2.1).

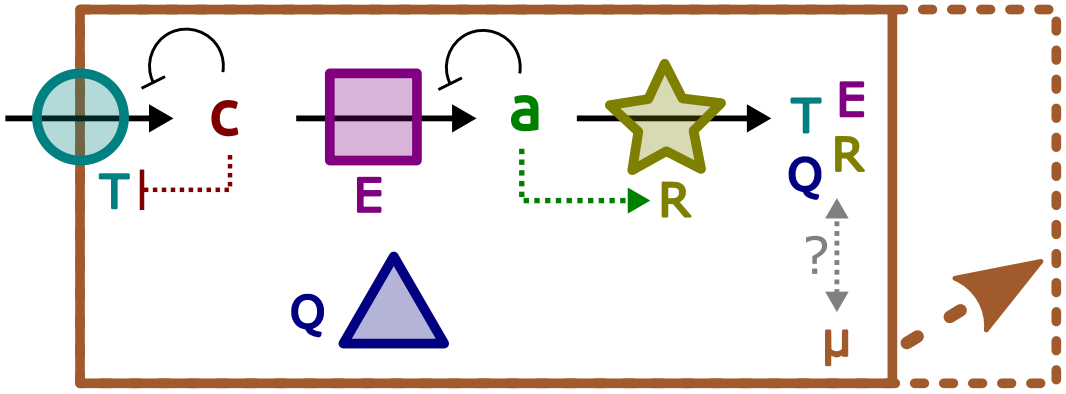


Figure 2.2: Cartoon of the metabolism of Figure 2.1 embedded in a cell with growth rate μ, including the gene regulation via the metabolite concentrations (red and green dashed arrows). A high concentration c of catabolites will inhibit the production r_T of transporter proteins, as denoted by the line segment perpendicular to the arrow. This will have the effect of increasing the productions r_E and r_R of enzymatic and ribosomal proteins. In contrast, a high concentration a of amino acids will stimulate r_R, as denoted by the pointed arrow, effectively decreasing $r_\mathrm{T}$and r_E. The production r_Q of housekeeping proteins is fixed.

Table 2.1: Parameters of the system with chosen values, separated into the categories described in chapter 2. The parameters in the regulation are optimised as described in 5.1, with uncertainties in the last digit according to the fits.

Class	Description	Symbol	Value
Metabolism	External substrate concentration	σ	10
	Maximal transporter turnover rate	k_T	variable[a]
	Maximal enzyme turnover rate	k_E	15 h^{-1}
	Maximal ribosome turnover rate	k_R	10 h^{-1} or variable[b]
	ratio of Michaelis constants for catabolite	β_c	0.1
	ratio of Michaelis constants for amino acid	β_a	0.1
Cell division	time scale of cell size/growth rate dependence	T_D	1 h
	cell size at division for zero growth	Δ	1000 molecules
Noise	burst size	b	10 molecules
Regulation	housekeeping proteome fraction	ϕ_Q	0.55
	response size of ribosome regulation	γ_R	1.4920(2)
	threshold for transporter regulation	θ_T	18.1(2)
	threshold for ribosome regulation	θ_R	3.6120(9)
	cooperativity of transporter regulation	n_T	1.509(4)
	cooperativity of ribosome regulation	n_R	1.708(1)

[a] $k_\mathrm{T} \in (0.1, 0.4, 1, 1.5, 2, 2.6, 3.3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 21, 26, 32, 48, 64, 128)\,\mathrm{h}^{-1}$ was used in chapter 3. For the simulations of Sections 4 and 5 the first three of this list were dropped.

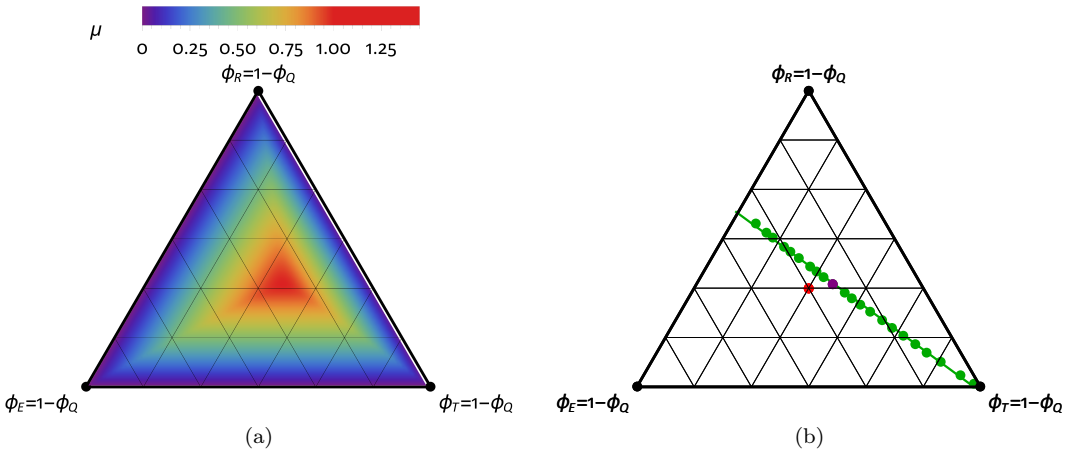
[b] $k_\mathrm{R} \in (1, 2, 3, 4, 5, 6, 7, 8.5, 10, 12)\,\mathrm{h}^{-1}$ was used in section 3.2.

3 Noiseless cells in optimal state

We will build up the model step-by-step. First we will consider a model that only consists of the growth and metabolism described above. Stochasticity and gene regulation will be included at a later stage. Given a set of external and internal parameters, there is a unique optimal state of the system, defined as the state where the growth rate is maximised. We will see that the optimal state reproduces the growth laws and the Monod law. Furthermore, we will consider the effect of perturbations of the cellular composition on growth by introducing an analytic framework built on growth control coefficients, defined in section 3.3.

3.1 Maximising growth rate

As described in sections 2.1 and 2.2, we can compute the growth rate of the system with any given protein composition. One example plot of the growth rate as a function of protein composition is plotted as a so-called simplex plot in Figure 3.1a. Here the operation rate k_T of the transporter is taken to be of the same order of magnitude to the operation rate $k_\mathrm{E}$of the enzyme. Specifically, $k_\mathrm{T} = 10\,\mathrm{h}^{-1}$; recall that $k_\mathrm{E} = 15\,\mathrm{h}^{-1}$. There is a composition that maximises the growth rate, in this case close to the point where all non-housekeeping proteins occur equally frequently with just a little less E. The precise position of the optimum depends on the parameters in Table 2.1. When any of the metabolic proteins do not occur at all, the cell does not grow. There are regions far from the optimum where just one protein is limiting growth. For example, when $\phi_\mathrm{R} = \frac{1}{6}\phi_\mathrm{Q}$ the growth rate is constant anywhere between $\phi_\mathrm{T} = \frac{2}{3}\phi_\mathrm{Q}$ and $\phi_\mathrm{T} = \frac{1}{6}\phi_\mathrm{Q}$. In this area the growth rate is therefore limited by the ribosome abundance.



(a) (b)

Figure 3.1: A point in the triangle corresponds to the configuration of the system as a point $(\phi_\mathrm{T}, \phi_\mathrm{E}, \phi_\mathrm{R}) \in \mathbb{R}^3$ constrained onto the plane $\phi_\mathrm{T} + \phi_\mathrm{E} + \phi_\mathrm{R} = 1 - \phi_\mathrm{Q}$. In the top corner, the only proteins present in the system are R and Q, whereas on the bottom line, no R is present and the system consists of T, E, and Q only. Horizontal lines correspond to constant ϕ_R. The plot works similarly for T (maximal in the bottom right) and E (maximal in the bottom left). (a) Growth rate $\mu\,(\mathrm{h}^{-1})$ as a function of cellular composition, for $k_\mathrm{T} = 10\,\mathrm{h}^{-1}$ and $k_\mathrm{R} = 15\,\mathrm{h}^{-1}$. The growth rate is maximised at $(\phi_\mathrm{T}, \phi_\mathrm{E}, \phi_\mathrm{R}) \approx (0.178, 0.115, 0.156) \approx (0.397, 0.256, 0.347) \times (1 - \phi_\mathrm{Q})$. (b) When varying k_T through the range of values described in Table 2.1, the optimal composition of the cell changes. The green points indicate the optimal composition, with a linear fit through them. For convenience, the optimum from figure (a) is coloured purple and the point where $\phi_\mathrm{T} = \phi_\mathrm{E} = \phi_\mathrm{R} = \frac{1-\phi_\mathrm{Q}}{3} = 0.15$ is shown in red.

With a numerical procedure in hand to compute the growth rate, it is straightforward to use numerical optimisation techniques to find the maximal growth rate, and hence the optimal composition of the cell. I performed this part of the computations in Wolfram Mathematica, using the inbuilt `NMaximize[]` function. I computed the maximal growth rate and the optimal composition for a range of different values of k_{T}. This amounts to different efficiencies of the transporter, corresponding to different qualities of nutrient. The changing optimal composition is plotted in Figure 3.1b.

3.2 Growth laws

The empirical growth laws, discussed in section 1.1, are formulated in terms of composition as a function of growth rate, so we have to invert the data. A plot of the composition as a function of the optimal growth rate is shown in Figure 3.2a. The fractional abundances of proteins indeed follow the linear dependences on growth rate as introduced before. That is, the system reproduces the growth laws in the optimal state.

The controlled parameter in these simulations was the the efficiency k_{T} of the transporter. If k_{T} increases, ϕ_{T} can decrease while keeping the total flux through the transporters constant. Therefore ϕ_{E} and ϕ_{R} can increase, so the total flux through these enzymes can increase[8]. The growth rate thereby increases.

The parameters from Table 2.1 have been chosen such that they resemble the experimental results semi-quantitatively, but we have not fitted the parameters to the data from [6]. It is possible to increase the steepness of e.g. the $\phi_{\mathrm{E}}(\mu)$-line by decreasing the efficiency k_{E} of the enzyme E. Increasing all rate constants k_{X} simultaneously will increase the maximum growth rate, which is given by the interception of $\phi_{\mathrm{T}}(\mu)$ with the horizontal axis. We did change the value of k_{R} independently of the others, as will be described towards the end of this section.

The increasing efficiency in the uptake of nutrients has the same effect as increasing the available amount of nutrient, as quantified by Monod's law. We have plotted the dependence of the growth rate on the transporter efficiency in Figure 3.2b and fitted a function of the form $\mu(k_{\mathrm{T}}) = \mu_{\max}\frac{k_{\mathrm{T}}}{k_{\mathrm{T}}+K}$. The fit is excellent.

We can now analyse the effect of changing nutrient quality k_{T} on the size of the metabolite pools. As shown in Figure 3.2c, the metabolite pools become larger at larger growth rate in a nonlinear fashion. As a metabolite pool increases, the first effect is a further saturation of the enzyme that consumes the metabolite. This is true until they are of order $\beta^{-1} = 10$. For even larger metabolite pools, product inhibition becomes important too. This is illustrated in Figure 3.2d. At low growth rates, the transporter is saturated by the external substrate, whereas with increasing growth rate the enzyme is inhibited more and more by its product, the catabolite. The efficiency of the general enzyme increases; in other words, the saturation by its substrate, the catabolite, is stronger than the inhibition by its product, the amino acids. The efficiency of the ribosomes increases very slowly with growth rate, but is generally high at all growth rates.

Secondly, to investigate the effect on the optimum of antibiotics that target the ribosome, I changed the efficiency k_{R} of the ribosome for several values of k_{T} already indicated in Figure 3.2b. The results on the proteome fraction of the ribosomes and transporters are shown in Figures 3.3a and 3.3b. If k_{R} decreases, the flux through the ribosomes and hence the growth rate decrease. This is counteracted by an increase in ϕ_{R} so ϕ_{T} and ϕ_{E} decrease.

[8] This means that ϕ_{T} decreases less than what would keep the total flux constant, because the total flux through the transporter must increase just like the total flux through the other enzymes.

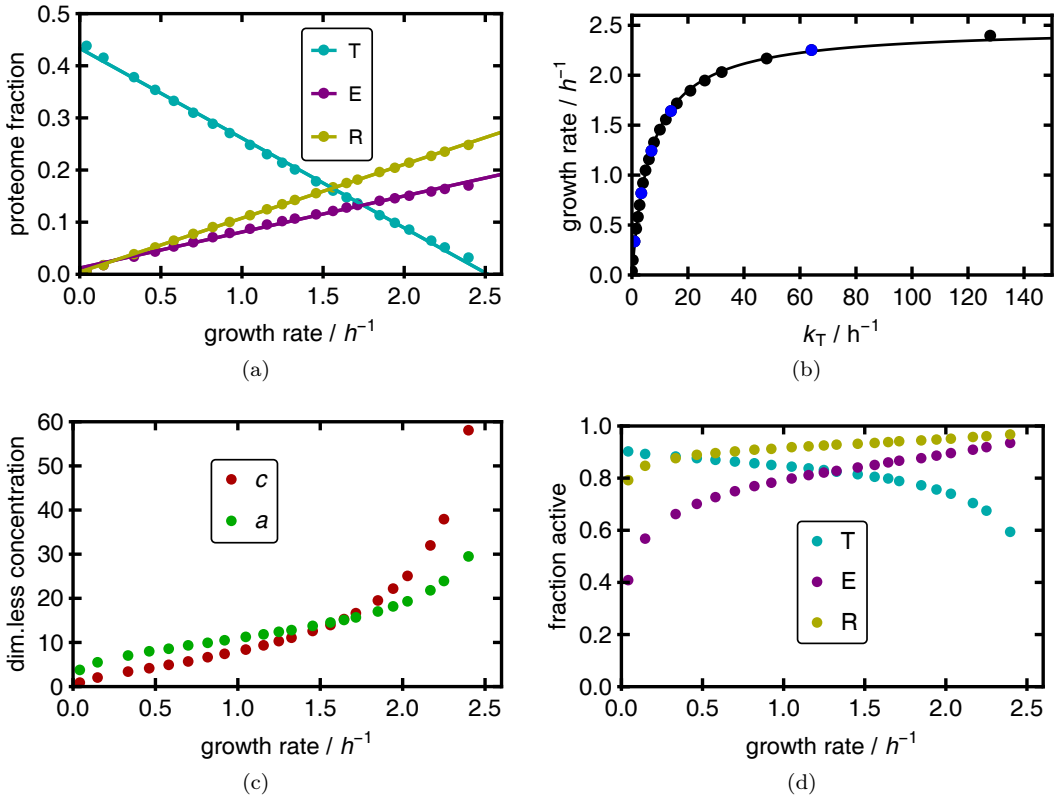
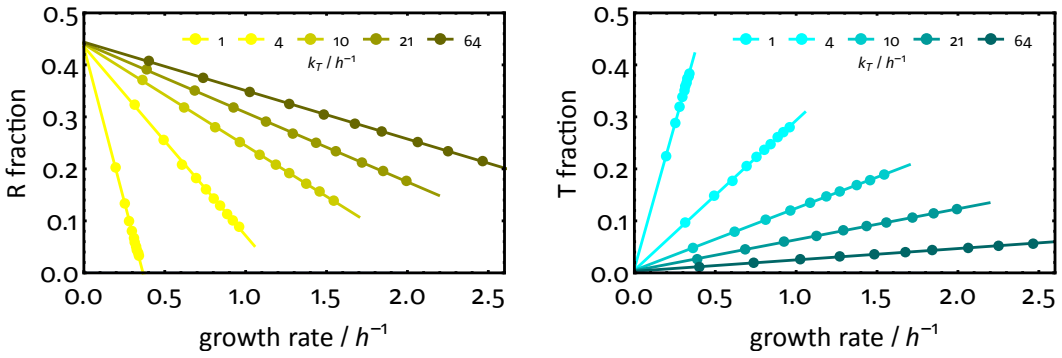


Figure 3.2: Plots describing the optimal state of the system for various nutrient qualities k_{T}, leading to varying growth rate. (a) Amount of transporter (cyan), enzymatic (purple), and ribosomal (yellow) proteins as a fraction of the total—the empiric linear relations are referred to as the growth laws. (b) The growth rate as a function of the transporter efficiency k_{T}, following Monod's law. Shown in blue are the values of k_{T} for which k_{R} was also varied (see Figure 3.3). (c) Concentrations of the metabolic intermediates as a function of the growth rate. The metabolite concentrations are scaled to the Michaelis constants of the reaction that produces them. The red points denote the dimensionless catabolite concentration c, the green points the dimensionless amino acid concentration a. (d) Fraction of the proteins that are in the active state, being the fractions $\frac{\sigma}{1+\sigma+\beta_{\mathrm{c}}c}$ for T, $\frac{c}{1+c+\beta_{\mathrm{a}}a}$ for E, and $\frac{a}{1+a}$ for R, using the now familiar colour coding for variables belonging to the T, E, and R proteins. Fit parameters for (a) and (b) are tabulated in appendix A.



(a) (b)

Figure 3.3: Descriptions of the optimal state for five values of k_{T}, and varying k_{R} with fitted straight lines. Amount of (a) ribosomal and (b) transporter protein. Lowering k_{R} corresponds to the /in vivo/ addition of antibiotics inhibiting the ribosome activity. This leads to an increased R fraction and a decreased T fraction. In the limit where $k_{\mathrm{R}} = 0$ (the ribosomes are fully inhibited), the cell consists of ribosomal protein only.

3.3 Growth control coefficients

We will now quantify the effect that the number of proteins of different types has on the growth rate. We want to know what properties of the metabolic network influence growth, and how strongly. This is especially important when we study the impact that noise has on the growth rate, as we will explain in section 3.4. We will do this by developing the concept of *growth control coefficients*.

We were inspired by metabolic control theory, which introduces the so-called flux control coefficients[35]. They are defined as the logarithmic derivatives of the flux through a metabolic network with respect to the enzyme concentrations, so the flux control coefficient with respect to a protein with concentration $[X_i]$ is given by

$$C_i^J = \frac{[X_i]}{J}\frac{\partial J}{\partial [X_i]}$$

Now we do not use the concentrations of proteins explicitly. Rather, we have developed the theory in terms of protein fractions, because we have taken the size of the cell to be identical to the total number of proteins. This means that an upshift in one concentration will decrease all other concentrations, and taking the partial derivative with respect to each particular concentration would require careful thought about the constraint (2.14). In addition to this, we choose to use the intensive variable μ over the extensive variable J (see section (2.2)).

Because of these issues, we have decided to develop a new framework. In analogy to the flux control coefficents, we define the *growth* control coefficients as the logaritmic derivative of the growth rate with respect to total protein copy number:

$$C_{\mathrm{X}}^{\mu} = \frac{X}{\mu}\frac{\partial \mu}{\partial X} \tag{3.1}$$

is the growth control coefficient of protein X.

We will use the growth control coefficients (sometimes called GCCs for short) to quantify how limiting each enzyme is. We can see that C_{X}^{μ} is positive if and only if $\frac{\partial \mu}{\partial X} > 0$. In other words, when the GCC of protein X is positive, we would increase the growth rate when increasing the copy number of X. Therefore, we say that X is limiting growth in this case.

In the opposite case, when $C^{\mu}_{\mathrm{X}} < 0$, the growth rate would decrease when we would increase X. This means that there is a surplus of X when its growth control coefficient is negative.

We now want to apply this concept to the optimally growing state, and see which enzymes limit the growth rate the most under varying circumstances. We will first develop a summation theorem similar to what has been done for the flux control coefficients C^J. We will then calculate the growth control coefficient C^{μ}_{Q} of the non-metabolic housekeeping protein, and finally the growth control coefficients C^{μ}_{X} for the metabolic enzymes $\mathrm{X} \in \{\mathrm{T}, \mathrm{E}, \mathrm{R}\}$.

Let us first consider a metabolic system with n enzymes $X_1, \ldots, X_n$ and growth rate μ. If we were to increase all copy numbers by the same amount, the growth rate would not change. (Recall that μ is an intensive quantity.) In other words, scaling the amount of enzyme with a scale factor α does not change the growth rate[9]:

$$\mu(\alpha \vec{X}) = \mu(\vec{X}) \tag{3.2}$$

Differentiating both sides with respect to α, using the chain rule, and evaluating the expression in $\alpha = 1$, gives the equality

$$\sum_{i=1}^{n} X_i \frac{\partial \mu}{\partial X_i} = 0$$

Using definition (3.1) of the growth control coefficients gives

$$\sum_{i=1}^{n} C^{\mu}_{\mathrm{X}_i} = 0$$

or, in slightly different notation,

$$\sum_{\mathrm{X}} C^{\mu}_{\mathrm{X}} = 0 \tag{3.3}$$

In words: the growth control coefficients must add up to zero.

In our system, the protein sector Q does not contribute to the metabolism. This means that the expression of Q proteins is a burden to the cell, as its production takes up metabolic resources without any return. However, we assume that cells cannot survive in the long run without these housekeeping proteins. In our model, the fact that Q functions as a load will mean that its growth control coefficient is negative: the cell would grow faster, at least in the short run, if it could express less Q proteins.

Because the housekeeping proteins do not contribute to the metabolism, the amino acid concentration a does not depend on Q, as was already apparent in section (2.1). We can nevertheless compute the growth control coefficient of Q:

$$\begin{aligned}
C^{\mu}_{\mathrm{Q}} &= \frac{Q}{\mu}\frac{\partial \mu}{\partial Q} \\
&= \frac{Q}{\mu}\frac{\partial}{\partial Q}\left(k_{\mathrm{R}}\phi_{\mathrm{R}}\frac{a(T,E,R)}{1+a(T,E,R)}\right) \\
&= \frac{Q}{\mu}\frac{\partial}{\partial Q}\left(k_{\mathrm{R}}\frac{R}{T+E+R+Q}\frac{a(T,E,R)}{1+a(T,E,R)}\right) \\
&= -\frac{Q}{\mu}\frac{1}{(T+E+R+Q)^2}\left(k_{\mathrm{R}}R\frac{a(T,E,R)}{1+a(T,E,R)}\right) \\
&= -\frac{Q}{\mu}\frac{\mu}{T+E+R+Q} \\
C^{\mu}_{\mathrm{Q}} &= -\phi_{\mathrm{Q}}
\end{aligned} \tag{3.4}$$

[9]The total flux will change linearly with α: $J(\alpha\vec{X}) = \alpha J(\vec{X})$. This is used in a similar way as equation (3.2) in the text to derive $\sum_{\mathrm{X}} C^{J}_{\mathrm{X}} = 1$.

So surprisingly, the growth control coefficient of the non-metabolic protein is simply equal to minus the fraction of the proteome allocated to it. Given the summation rule (3.3), this means for the enzymes X that do feature in metabolism,

$$\sum_{\mathrm{X}\neq\mathrm{Q}} C^{\mu}_{\mathrm{X}} = \phi_{\mathrm{Q}} \tag{3.5}$$

This is all we can say in general. We will now consider the growth control coefficients of the metabolic enzymes $\mathrm{X} \in \{\mathrm{T}, \mathrm{E}, \mathrm{R}\}$ in the optimal state for a given cell size. When the cell grows optimally, i.e., $\mu(T, E, R)$ is maximised, it is not possible to increase any copy number without decreasing the growth rate. This is because an increase in the copy number of one enzyme will decrease the concentrations of all other enzymes. It means that

$$\frac{\partial\mu}{\partial T} = \frac{\partial\mu}{\partial E} = \frac{\partial\mu}{\partial R} = K \tag{3.6}$$

for some constant K independent of T, E, and R. To see that this is so, suppose that, for example, $\frac{\partial\mu}{\partial T} > \frac{\partial\mu}{\partial E}$. Then a cell with an infinitesimal amount of extra T and less E would have a higher growth rate than this one, which contradicts the assumption that the growth rate was maximised. It follows that all derivatives must be equal. Now we have

$$C^{\mu}_{\mathrm{X}} = \frac{X}{\mu}\frac{\partial\mu}{\partial X} = \frac{X}{\mu}K \tag{3.7}$$

for all proteins $\mathrm{X} \in \{\mathrm{T}, \mathrm{E}, \mathrm{R}\}$. Using equation (3.5) we obtain

$$\phi_{\mathrm{Q}} = \frac{T + E + R}{\mu}K \tag{3.8}$$

We can substitute equation (3.8) back into equation (3.7) to arrive at

$$\begin{aligned} C^{\mu}_{\mathrm{X}} &= \frac{\phi_{\mathrm{Q}} X}{T + E + R} \\ &= \frac{\phi_{\mathrm{Q}}}{1 - \phi_{\mathrm{Q}}}\phi_{\mathrm{X}} \end{aligned} \tag{3.9}$$

In conclusion, in the situation where protein resources are allocated optimally, the growth control coefficient of a metabolic enzyme is proportional to its concentration.

This is a general result. Relation (3.9) holds independent of the specific model of metabolism. But it is also very specific, because it only holds in the optimum. Any amount of noise will perturb the system such that it no longer holds.

We have already seen that the protein abundances ϕ_{X} are linearly related to the growth rate. Using these growth laws together with equation (3.9), we see that the growth control coefficients will also depend linearly on the growth rate.

3.4 Noise transmission and growth control coefficients

The general intuition concerning biological noise is, that its importance stems from the fact that copy numbers are small. After all, small copy numbers usually result in a large coefficient of variation (the standard deviation divided by the mean). One may be tempted to conclude from this that the proteins that are the least abundant are the most important for the noise properties.

However, the copy number of a certain protein does not necessarily have any implications for the survival of a cell. Properties that do influence the fitness, such as the growth rate, depend on the properties and abundances of many proteins in often large networks. The noise in these properties depends not only on the noise in the copy numbers, but also on the transmission coefficients. For the case of growth rate, these transmission coefficients are

the growth control coefficients C^{μ}_{X}. As we have seen in the previous section, they depend on the copy number of the proteins. This means that the noise in the growth rate will depend nonlinearly on the copy numbers of the network proteins.

As a simplified model, consider the deviation $\Delta\mu = \mu - \bar{\mu}$ of the growth rate μ from its optimum $\bar{\mu}$. Suppose that the growth rate depends linearly on the amounts $X_1, \ldots, X_n$ of some proteins $\mathrm{X}_1, \ldots, \mathrm{X}_n$ that are considered as independent random variables:

$$\mu = \sum_i \frac{\partial \mu}{\partial X_i} X_i$$

We write σ_i^2 for the variance of the distribution that describes the copy number X_i. If the protein abundances are independent, the variance σ_μ^2 of the distribution of $\Delta\mu$ is given by

$$\sigma_\mu^2 = \sum_i \left(\frac{\partial \mu}{\partial X_i}\right)^2 \sigma_i^2 = \sum_i \left(\frac{\mu}{X_i} C_i^{\mu}\right)^2 \sigma_i^2$$

We can rewrite this in terms of the coefficients of variation (the standard deviations divided by the means). These are quantities that describe the amount of noise relative to the quantity that is noisy:

$$\left(\frac{\sigma_\mu}{\mu}\right)^2 = \sum_i (C_i^{\mu})^2 \left(\frac{\sigma_i}{X_i}\right)^2 \equiv \sum_i (V_{i\to\mu})^2$$

In the optimum and at low concentrations, $C_i^{\mu} \sim X_i$. Furthermore, the coefficient of variation of a quantity that is produced in a Poisson-like process decreases with its size as the square root[10]: $\frac{\sigma_i}{X_i} \sim X_i^{-1/2}$. Therefore, $V_{i\to\mu} \sim X_i$.

It follows that the most abundant proteins have the largest effect on the noise in the growth rate, even though the noise in their concentrations may be relatively small.

[10]When proteins are produced in exponentially distributed bursts, their abundance is Gamma distributed[36] and has standard deviation $\sigma_i = \sqrt{bX_i}$ with b the burst size.

4 Stochasticity in optimally fixed system

In the previous section, we have discussed the optimised system. The first extension to this is including noise. I therefore performed the simulations described in chapter 2 with one exception: I left gene regulation (section 2.5) out for the moment. I kept the production rates of the enzymes fixed at the optimal rate, computed in section 3.1. This approach enables us to begin analysing the effects of growth and noise on each other.

Recall that, at low growth rates, the cells can be very small. This means that cell division (see section 2.3) is a significant source of cell-to-cell variability due to the binomial partition. In fact, when the number of proteins of a single type in the mother cell is very small, it is possible that one daughter cell inherits no copies at all. This means the cell cannot grow, so that the simulation aborts. Because of this, it was not possible to simulate at very low growth rates, which is at very low k_T. I only simulated for $k_\mathrm{T} > 1.5$.

4.1 Growth laws

The introduction of noise keeps intact the linear dependence of the proteome fractions on the growth rate (see Figure 4.1a). Moreover, the dependence of growth rate on transporter efficiency k_T still follows a Monod law (see Figure 4.1b).

Although the general trends seen in the optimum are still valid, there are some slight differences between the growth and Monod laws in the optimal system and those in the simulated system with fixed optimal rates. Firstly, the linear fits of the ribosome and enzyme fraction show a slight offset: $\phi_\mathrm{E}(0) > 0 \,\mathrm{and}\, \phi_\mathrm{R}(0) > 0$. Secondly, the maximum growth rate, as seen both in the intersection of $\phi_\mathrm{T}(\mu)$ with the horizontal axis in Figure 4.1a and in the asymptote of the Monod law in Figure 4.1b, is somewhat smaller in the simulated system as compared with the optimal system. Thirdly, the Monod constant K is smaller.

The fact that the growth rate is smaller in the simulation than in the optimised state is straightforwardly explained. Because the simulation includes noise in the production of proteins, the fractional abundances will fluctuate over time (even though the *rates* in the stochastic process of protein production are constant). This means that the system will spend most of its time some distance away from the optimum, which necessarily means that the growth rate is lower. This effect is stronger at smaller growth rate, because then the cell is smaller. A smaller cell means that the relative size of the fluctuations is larger, so that the effect of decreasing the growth rate is larger. This explains the larger Monod constant.

We can see the effect of fluctuations on the metabolism clearly in Figure 4.1c. Recall that, in the optimised system, the metabolite concentrations increased monotonically with the growth rate. In contrast, in the fixed-rate system we see that the average metabolite concentrations become very large at low growth rates in the fixed-rate simulations. Recall that the ratio of concentrations where the metabolic intermediates function as substrates for the next reaction and where they function as inhibitors for the previous reaction, β_c and β_a, were taken to be 0.1. Because the (dimensionless) metabolite concentrations are larger than $\beta_\mathrm{c,a}^{-1} = 10$ over the whole range of simulations, product inhibition is important in all cases. This is seen in Figure 4.1d, showing that the enzyme activities are clearly repressed in the noisy system as compared to Figure 3.2d.

The departure from the optimum is clearly seen in the average growth control coefficients, plotted in Figure 4.2a. As we demonstrated in section 3.3, the GCCs should follow the growth laws if the system is always at the state that optimises growth. This is clearly not the case in the noisy simulations. The most notable difference is, that the average of C^μ_T is much smaller than the optimum of $C^\mu_\mathrm{T} = \frac{\phi_\mathrm{Q}}{1-\phi_\mathrm{Q}}\phi_\mathrm{T}$ for all but the very highest growth rates, and even becomes negative at very slow growth. Because $\sum_\mathrm{X} C^\mu_\mathrm{X} = 0$ (Equation 3.3) and $C^\mu_\mathrm{Q} = -\phi_\mathrm{Q}$ (Equation 3.4) must still hold, the average C^μ_E and C^μ_R are much larger than their optimal values.

However, the growth control coefficients are very sensitive measures of optimality. As shown in Figures 4.2b, 4.2c, and 4.2d, they are either very large or quite small (negative)

when the system is not close to the optimum. In other words, the system is limited by one component if it is far from the optimal growth rate. This has the effect of making the distributions of the growth control coefficients in the noisy simulation highly skewed and possibly bimodal, as plotted in Figures 4.2e and 4.2f.

From these considerations it is clear that the system where the proteins are produced stochastically, but with the production rates fixed at the optimum, finds itself far from the optimum most of the time.

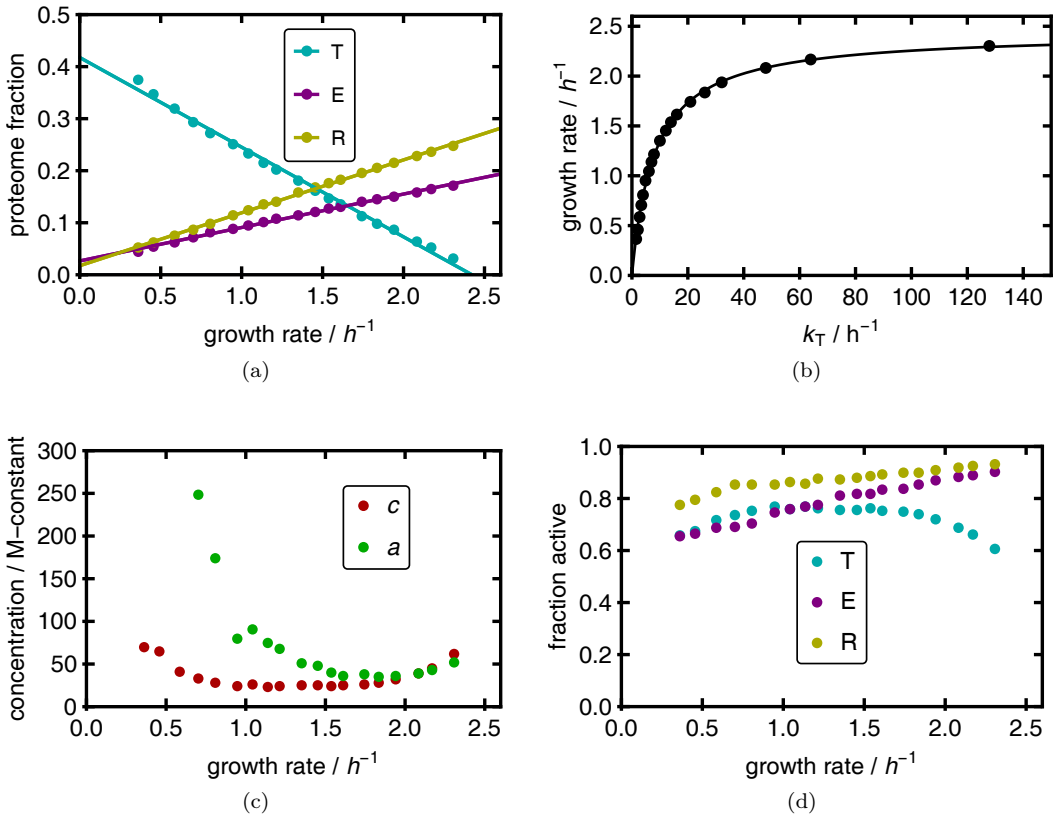


Figure 4.1: Plots describing the state of the simulated system at varying nutrient quality k_T, where for each value of k_T all protein production rates were held fixed at the optimum. (a)–(d) as in Figure 3.2. (a) Comparing the growth laws in the simulated system with fixed production rates with the optimised system, we observe two differences. Firstly, the simulated transporter fraction departs somewhat from the fit at low and high growth rate, although the points at intermediate growth rates still follow a straight line. Secondly, the fits to ϕ_R and ϕ_E show a small offset in the limit of zero growth. (b) At low to intermediate values of k_T, the growth rate is suppressed compared to the optimised system. (c) The average metabolite concentrations at low growth rates are very high, the values for a at $k_T \leq 2.6$ were even omitted from the figure. (d) Keeping the production rates results in suppressed activities of transporters at low growth rates. Fit parameters for (a) and (b) are tabulated in appendix A.

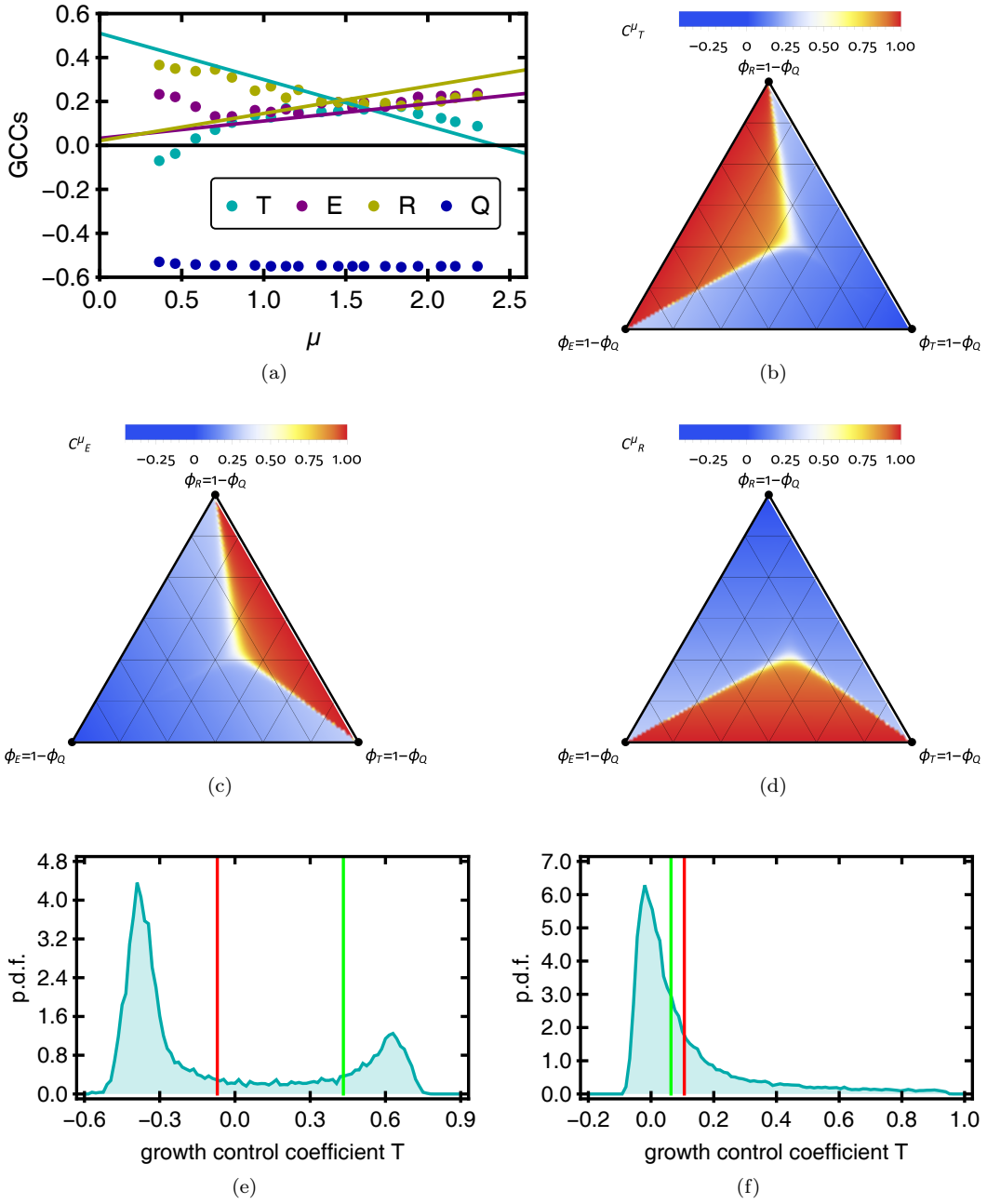


Figure 4.2: (a) Growth control coefficients C^{μ} of the four protein types as a function of growth rate μ, upon variation of k_{T}. The straight lines are the optimal values of C^{μ}, given the growth laws of Figure 3.2a and equation 3.9. (b)-(d) Plots of the growth control coefficients (b) C^{μ}_{T}, (c) C^{μ}_{E}, (d) C^{μ}_{R}, as a function of cell composition, at $k_{\mathrm{T}} = 10$. In the red areas, the respective control coefficient is large, so the corresponding enzyme is highly limiting, while in the blue areas, the control coefficient is small, so the corresponding enzyme is overabundant. The transition between these two conditions is sharp. (e)-(f) Histograms of C^{μ}_{T} for (e) $k_{\mathrm{T}} = 1.5$, (f) $k_{\mathrm{T}} = 64$, with the optimal value denoted as a green line and the average value as a red line. The system spends most of its time away from the optimum, with C^{μ}_{T} either below or above the optimal value. The bimodality disappears for increasing growth rate, when the optimal T-fraction approaches zero.

4.2 Distributions of variables

We can now investigate the distribution of variables in the same manner as we did for the growth control coefficients. We will focus on the growth rate and the protein fractions. We will discuss how the distributions compare at low, medium and high growth rate.

We will start with the general features of the distributions. For this we focus on medium growth rate, which results from transporter efficiency $k_T = 10$. The histogram of the growth rate probability density function is plotted in Figure 4.3a At first sight, it looks reasonably like a Gaussian, but remarkably, it is possible to grow faster than the "optimal" growth rate. This is odd, because it would appear that the cell has no way of growing faster than the optimum by definition. However, this reasoning assumes that all other parameters remain constant. In particular, it assumes that the proteome fraction of the housekeeping proteins ϕ_Q remains fixed. This is not the case: for the Q proteins as well as the metabolic ones, the *rate* of production is fixed, but their abundance fluctuates due to the stochastic production. A downward fluctuation of ϕ_Q causes a growth rate greater than "optimal".

Of course, ϕ_Q is not the only quantity that fluctuates; the other protein fractions do so too. Their distributions are plotted in Figures 4.3b, 4.3c, and 4.3d. Any fluctuation in these protein fractions (while ϕ_Q remains constant) has the effect of decreasing the growth rate from the optimum. As explained further in the next section (4.3), this means that the mode of the growth rate distribution lies below the optimal growth rate. At this intermediate transporter efficiency, the average protein abundances agree very well with their optimal values.

We will now discuss the effects on the distributions when the growth rate varies (because of varying transporter efficiency), starting with the case of low average growth rate (when $k_T = 1.5$). The distributions of growth rate and fractional protein abundances are plotted in Figure 4.4. We see here that the growth rate distribution has a large tail to the left, meaning that growth is especially hindered by fluctuations when the growth rate is already small. The following observation explains this: a small growth rate means that proteins are produced at a smaller rate. But the only way of "escaping" from a state of slow growth, is producing (the right kind of) proteins. Because all production rates are small, the cell gets trapped in a state of slow growth. This is more apparent when the average growth rate is small, because then the relative fluctuations in the growth rate are larger.

It appears also that the average ϕ_T is somewhat larger than the optimum, while the average ϕ_E and ϕ_R are equal to their respective optimal values. The key here is again in the asymmetry: the distributions of ϕ_E and ϕ_R are skewed to the right, which means that most of the time, there are fewer enzymes and ribosomes present than what would be optimal. That means that there should be more transporters. The skewness arises because the average copy numbers are as small as a few times the burst size.

For completeness, the distributions at high growth rate ($k_T = 64$) are plotted in Figure 4.5. The general trends already seen at intermediate growth rate are confirmed. If anything, the average abundances are even closer to their optimal values. This is because the relative size of the fluctuations is smaller, as the copy numbers are much bigger. Even though the fractional abundance of transporters is about equal at $k_T = 64$ to the fractional abundance of enzymes and ribosomes at $k_T = 1.5$, the former's distribution is not noticeably skewed, unlike the latter two. This is also because the total copy number is still quite large, due to the size of the cell being much greater at high growth rates.

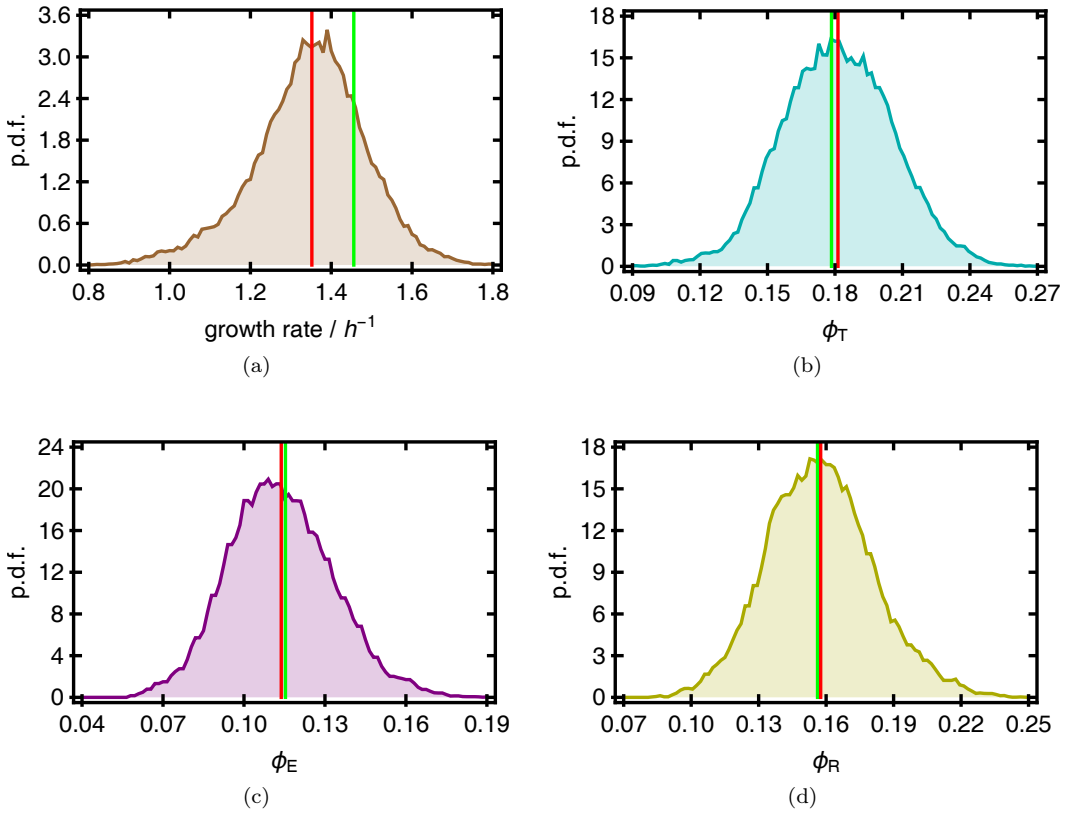


Figure 4.3: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 10$ with the optimal value indicated by a green line and the average value by a red line. Note that the histogram of the growth rate in (a) is slightly skewed to the left, that the most frequent value is about equal to the average but smaller than the optimum, and that faster than "optimal" growth is possible due to fluctuations in ϕ_Q.

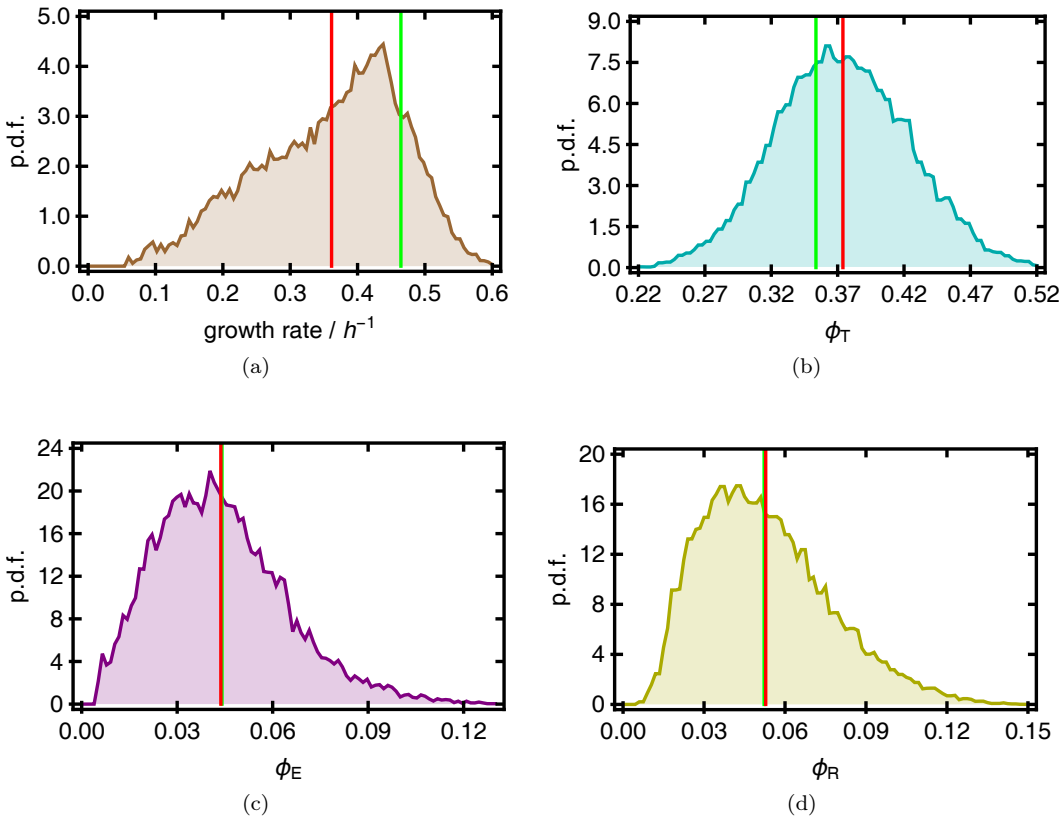


Figure 4.4: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 1.5$ with the optimal value indicated by a green line and the average value by a red line. (a) The growth rate histogram is very much skewed. When the growth rate is small, the system runs at a slower pace. It therefore takes a long time to exit a state of low growth rate, which increases the frequency of such states. (b) The growth control coefficient of transporter proteins, C_T^μ, should be relatively large at low nutrient quality and hence slow down growth. However, we have seen in Figure 4.2e that C_T^μ is usually negative for $k_T = 1.5$. This means that adding more T molecules usually slows down growth. When slow-growing states are more abundant, the result is that states with higher ϕ_T occur more frequently.

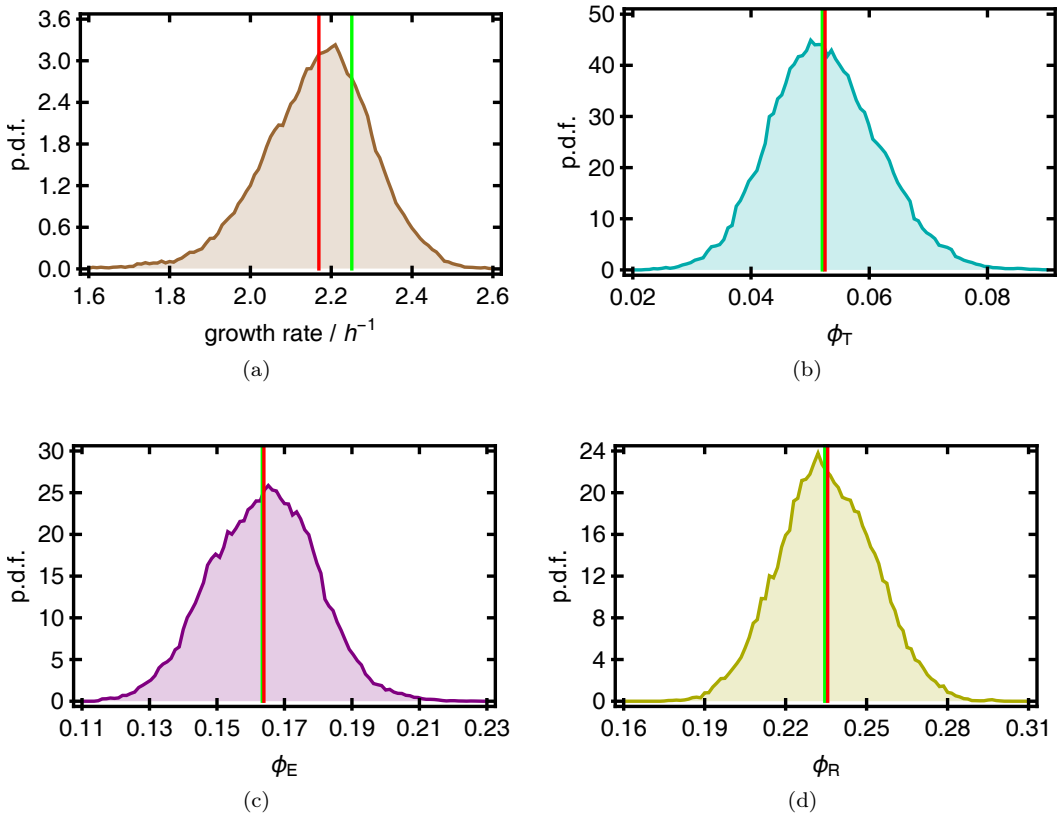


Figure 4.5: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 64$ with the optimal value indicated by a green line and the average value by a red line.

4.3 Dimensionality of noise

As we have seen, the mode of the growth rate distribution is smaller than the optimal value. We will now present a simplified model of noisy protein expression that explains this, even without a housekeeping sector. It will appear that the number of different sources of fluctuations (which we call dimensionality) is important.

Consider the growth rate μ as a function of n independent random variables $X_1, \ldots, X_n$ (interpreted as the copy numbers of n proteins that constititute the metabolic network responsible for growth). Suppose that the X_i's are normally distributed. In the vicinity of the optimum, we can approximate μ by a second-order Taylor expansion. In this approximation, the growth rate depends quadratically on the distance to the optimum in each variable X_i.

Now μ is a random variable too. In fact, it is equal to the optimal growth rate $\tilde{\mu}$ minus a sum of independent squares of n normal variables. This means that, by definition, $\tilde{\mu} - \mu$ is distributed according to a chi-squared distribution with n degrees of freedom. This is equivalent to saying that μ is distributed according to a Gamma distribution with shape $k = \frac{n}{2}$. The distribution of μ therefore diverges at $\mu = \tilde{\mu}$ for $n = 1$, it has mode equal to the optimum for $n = 2$ when it is simply exponentially distributed, and the mode of μ is smaller than $\tilde{\mu}$ for $n > 2$. In the large-n limit, the distribution approaches a normal distribution according to the central limit theorem.

4.4 Crosscorrelations

We can now study the effects that noise in the protein abundances and noise in the growth rate have on eachother. We will compute the crosscorrelations $X_{\phi-\mu}$ of these quantities, defined as

$$X_{\phi-\mu}(\Delta t) = \frac{\int \mathrm{d}t \Big(\phi(t) - \tilde{\phi}\Big)\Big(\mu(t+\Delta t) - \tilde{\mu}\Big)}{\sigma_\phi \sigma_\mu}$$

where $\tilde{\phi}$, $\tilde{\mu}$ denote the averages and σ_ϕ, σ_μ the standard deviations of ϕ, μ respectively. They are plotted in Figure 4.6.

The only process that can negate fluctuations is dilution, as a result of the growth in cell size. This means that the typical time scale of the crosscorrelations in the system without regulation is the time associated with growth, i.e. the division time. The crosscorrelations indeed decay on this time scale.

For small growth rates, the crosscorrelation of the transporter fraction and growth rate is negative. In other words, expression of transporters above the mean inhibits growth. This agrees with the negative average values for the growth control coefficients found in the low-growth case (Figure 4.2a).

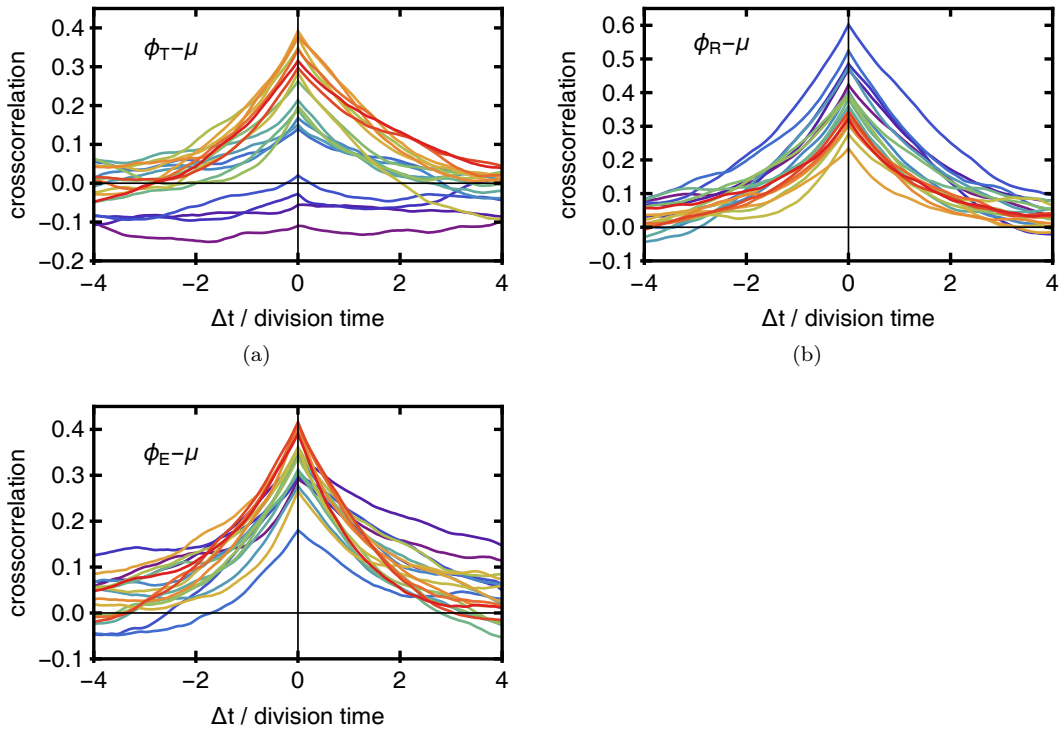


Figure 4.6: Crosscorrelations of the proteome fractions of (a) transporters ϕ_T, (b) ribosomes ϕ_R, and (c) enzymes ϕ_E, with the growth rate μ. The crosscorrelation at time Δt is the correlation of $\phi_X(t)$ and $\mu(t+\Delta t)$ for all times t. The plots show the crosscorrelations at all values of k_T that were simulated, ranging from $k_T = 1.5$ in purple, through blue, $k_T = 10$ in green, and yellow, to $k_T = 128$ in red. The time, on the horizontal axis, was scaled for each condition to the average division time (the total time taken for the entire simulation, which ran for 1000 cell cycles, divided by 1000).

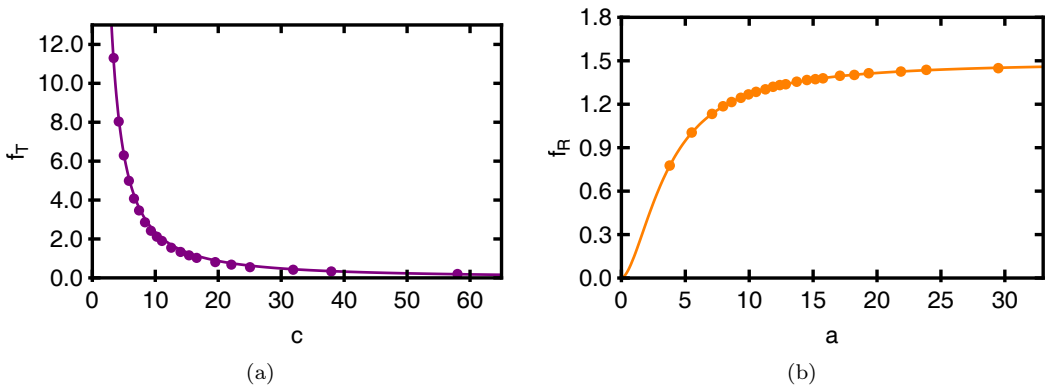
5 Regulated system

In this section we will analyse the system in full simulations. The production rates of the proteins will be determined by the regulation as described in section 2.5. In these simulations, the protein composition again determines the concentrations of metabolic intermediates. These concentrations in turn determine the rate of protein production. We will first describe how we choose the regulation functions so as to regulate the system close to the optimal state. We will see that this regulation is able to reproduce the growth laws. The regulation is sensitive to small variations in the regulation functions and unable to optimise the system exactly. We will also see that the regulation reduces the effects of noise. The trade-off between a non-optimal average state and less time spent very far away from the optimum is decided in favour of the latter—the regulated system grows faster than the non-regulated, fixed-rate system. We will end this section with discussing the noise properties of the system, by way of the crosscorrelations.

5.1 Choosing regulation functions optimally

As we have seen in section 2.1, all variables, namely T, E, R, c, and a, vary monotonically with the growth rate μ. This means they also vary monotonically among themselves. As we are interested in determining $f_{\mathrm{T}}(c)$ and $f_{\mathrm{R}}(a)$, we can use the fact that $a(\mu)$ and $c(\mu)$ are monotonic to write the regulation functions in terms of μ, instead of in terms of a and c. Likewise, the growth laws allow us to write $\phi_{\mathrm{T}}(\mu)$, $\phi_{\mathrm{E}}(\mu)$, and $\phi_{\mathrm{R}}(\mu)$. In the regulatory steady state, a and c do not change and $F_{\mathrm{X}} = \phi_{\mathrm{X}}$ (see section 2.5). This allows us to write equations (2.23), (2.24), and (2.25) as

$$\begin{aligned}
\phi_{\mathrm{T}}(\mu) &= (1-\phi_{\mathrm{Q}})\frac{f_{\mathrm{T}}(\mu)}{f_{\mathrm{T}}(\mu)+1+f_{\mathrm{R}}(\mu)} \\
\phi_{\mathrm{E}}(\mu) &= (1-\phi_{\mathrm{Q}})\frac{1}{f_{\mathrm{T}}(\mu)+1+f_{\mathrm{R}}(\mu)} \\
\phi_{\mathrm{R}}(\mu) &= (1-\phi_{\mathrm{Q}})\frac{f_{\mathrm{R}}(\mu)}{f_{\mathrm{T}}(\mu)+1+f_{\mathrm{R}}(\mu)}
\end{aligned}$$



(a) (b)

Figure 5.1: Fits of the regulation functions used in determining the productions of (a) transporters given catabolite concentration, (b) ribosomes given amino acid concentration.

We can now write f_T and f_R in terms of the proteome fractions ϕ_T, ϕ_E, and ϕ_R for each value of μ:

$$f_T = \phi_T \frac{1 + \frac{\phi_R}{1-\phi_Q-\phi_R}}{1 - \phi_Q - \phi_T - \frac{\phi_T\phi_R}{1-\phi_Q-\phi_R}}$$

$$f_R = \phi_R \frac{1 + \frac{\phi_T}{1-\phi_Q-\phi_T}}{1 - \phi_Q - \phi_R - \frac{\phi_R\phi_T}{1-\phi_Q-\phi_T}}$$

In these expressions, we can use the values of ϕ_T, ϕ_E, and ϕ_R that we found in the optimal state for a range of different conditions k_T. This gives us the values of f_T and f_R in these states too. We can then plot the pairs (c, f_T) and (a, f_R), as in Figure 5.1. This allows us to fit regulation functions $f_T(c)$ and $f_R(a)$ as we described in section 2.5.

5.2 Growth laws

Straight lines are still good fits for the proteome fractions as functions of the growth rate, as seen in Figure 5.2a. We can see that the offsets, $\phi_E(0)$ and $\phi_R(0)$, are a bit more pronounced than they already were in the fixed-rate system. The maximal growth rate is higher than in the noisy system without regulation, as evidenced by both the intercept of $\phi_T = 0$ and the asymptote of the Monod law in Figure 5.2b.

As plotted in Figure 5.2c, the metabolic intermediates are present in higher concentrations than in the optimised system, but at small growth rate they are far lower than in the unregulated system. The amino acid concentrations are high enough to inhibit the enzyme that produces them at all growth rates. Because a is high, the ribosome is more saturated as well. Both effects are apparent in Figure 5.2d.

This also translates to the GCCs (see Figure 5.2e): C_E^μ is smaller than its optimal value, while C_R^μ is larger (so the ribosomes are more and the enzymes are less limiting than they would be in the optimal situation). It is clear from this picture, that the regulated system grows more robustly than the non-regulated system, because the growth control coefficients are all positive and show the same trends for changing growth rate as in the optimal state.

5.3 Simplex plots

We will use the triangular plots to analyse the differences between the proteome constitutions in the optimal, fixed-rate and regulated systems. The plots are in Figure 5.3.

We can see that the average composition of the fixed-rate and the optimal system are very much alike, and only show an asymmetry at high ϕ_T. This happens at low k_T, so at low growth rate, when the cell is at its smallest and stochasticity is at its most important. The effect of noise is asymmetric in such a way that the system find itself, on average, with more transporters and less enzymes than it would have optimally.

The same effect is seen at low k_T for the regulated system. However, the plot in Figure 5.3b also shows that the composition of the regulated system is always slightly different from the optimal system, even though we tried to regulate the system such that the optimal state would arise in steady state. This suggests that the system is sensitive to the precise details of regulation.

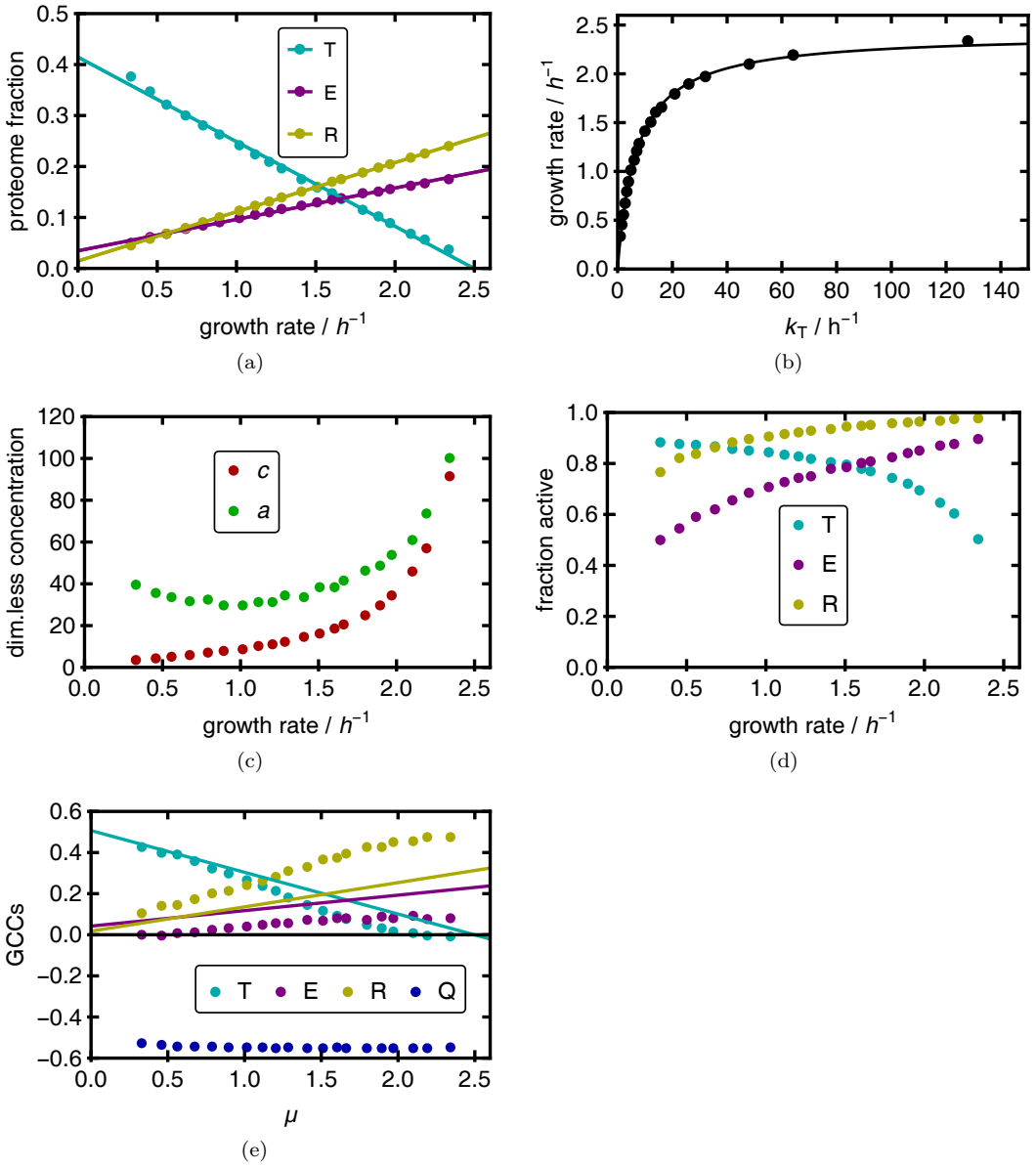
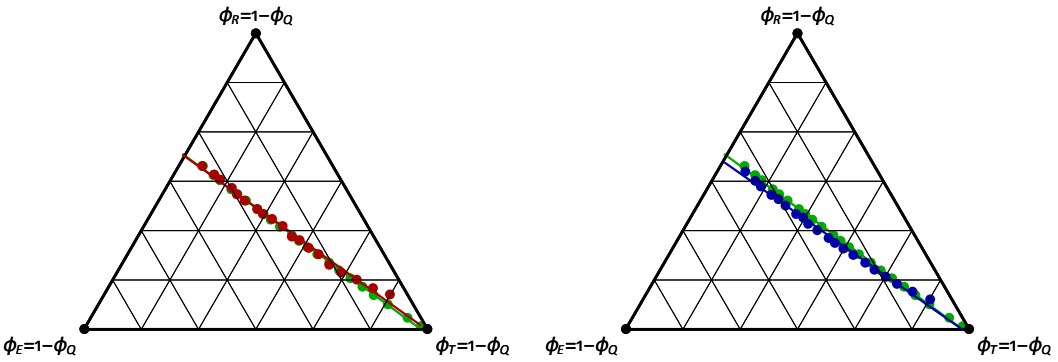


Figure 5.2: Plots describing the state of the simulated system at varying nutrient quality k_T. The production rates were regulated following the regulation functions fitted in section 5.1. (a)–(d) as in Figures 3.2 and 4.1. (e) growth control coefficients as in Figure 4.2a. (a) The fits to ϕ_R and ϕ_E show offsets in the limit of zero growth that are larger than in the non-regulated system. (b) The growth rate as a function of transporter efficiency follows the same Monod law in the regulated simulations as in the optimised state. (c) The average metabolite concentrations in the regulated simulation are only slightly larger than in the optimum. (d) The fractions of active transporters enzymes are lower in the regulated system compared to the optimum, while the ribosomes are more saturated. All of the fractions show a smooth, monotonous dependence on growth rate. (e) The GCCs of the metabolic proteins almost all positive throughout the range of conditions. The average GCCs do not follow the optimal values well, but the trend is the same in both the regulated and the optimised system: transporters become less limiting while the enzymes and especially ribosomes become more limiting with increasing growth rate. Fit parameters for (a) and (b) are tabulated in appendix A.



(a) (b)

Figure 5.3: Plots showing the composition of the cell with varying k_{T}, as in Figure 3.1b. Both plots show the optimal state in green, compared with (a) the fixed-rate system in red, (b) the regulated system in blue.

5.4 Distributions of variables

The fact that the regulation places the system not quite at the optimal state is also reflected in the distributions of growth rate and cellular composition, plotted for medium, small, and large growth rate in Figures 5.5, 5.4, and 5.6.

In general, the regulation narrows the distributions. This means that the system spends more time in fast-growing states near the optimum. However, because the regulation is not perfect, the modes of the distributions of the protein abundances are somewhat different from the optimum. The net effect is still an increase in the average growth rate compared to the fixed-rate system. We observe that there is a trade-off between regulation and noise: tight regulation means less time spent in states that are very far from the optimum, but if the regulation is itself too far from the optimum, the average growth rate may decrease. What this analysis does not take into account, is the cost of maintaining the regulation by expressing more regulatory proteins. This was analysed in more detail in[37].

There are three more observations to be made from the distributions of abundances. Firstly, although all distributions of proteome fractions are more narrow than in the fixed-rate system, this is lesss so for the distribution of ϕ_{E}. This can be easily explained: the production of the enzymes, unlike that of the other two metabolic proteins, is not directly regulated.

Secondly, the production of enzymes at $k_{\mathrm{T}} = 1.5$ (slow growth) is about as skewed in the regulated system as in the fixed-rate system. However, in the regulated system very low expression of enzymes never occurs. This is due to the shape of the function that determines the production rate of enzymes: $r_{\mathrm{E}} \sim \frac{1}{1+f_{\mathrm{T}}+f_{\mathrm{R}}}$. The constant numerator means that it is hard to regulate the enzyme expression to become very low.

Finally, it appears that the distribution of ϕ_{T} is narrowed more by regulation in conditions of fast growth, and the distribution of ϕ_{R} is narrowed more in conditions of slow growth. We can explain this by the fact that the main cause of fluctuations is the stochasticity in the expression of the housekeeping sector, because F_{Q} is always taken to be constant. In other words, the Q protein production is unregulated. In the case of, for example, an upward fluctuation in ϕ_{Q}, all other proteins will be expressed at a lower fraction. However, when the proteome fraction of a metabolic protein is smaller, its fractional abundance will be decreased less by this effect. In conditions of fast growth, ϕ_{T} is small and its stochasticity will be mainly determined by fluctuations in its own production. These fluctuations are suppressed by regulation. Likewise, in conditions of slow growth, ϕ_{R} is small so its fluctuations are suppressed more by regulation in this case.

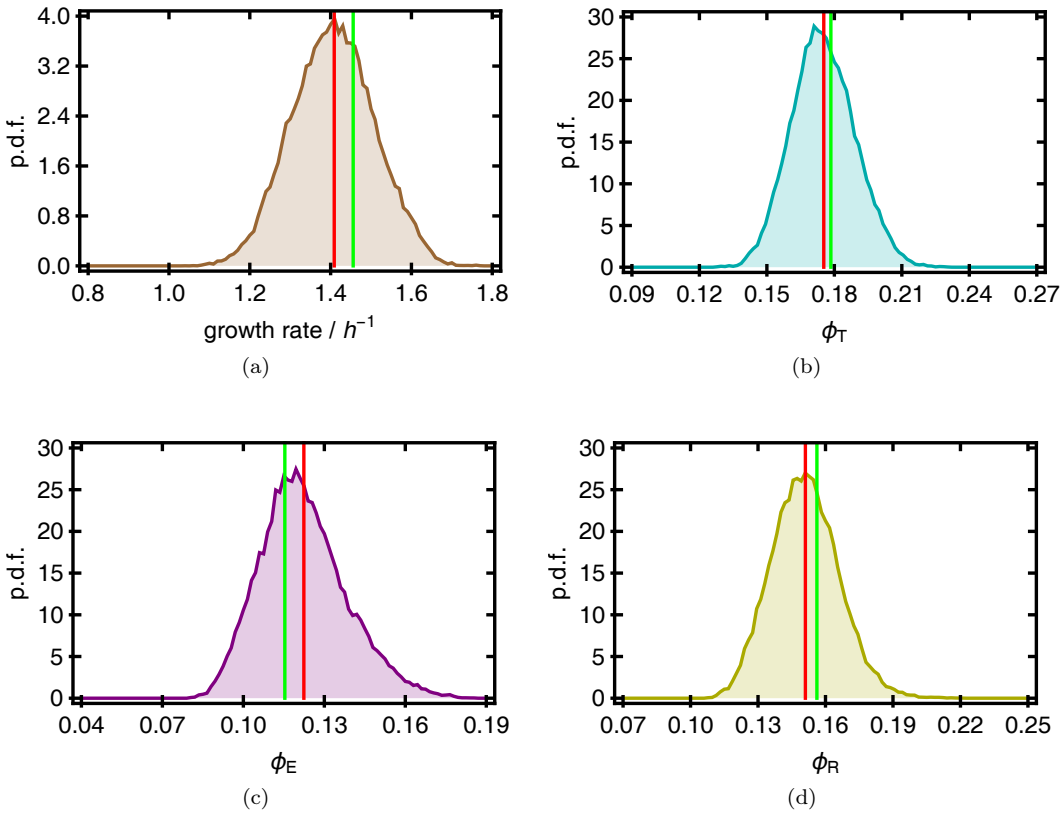


Figure 5.4: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 10$ with the optimal value denoted as a green line and the average value as a red line.

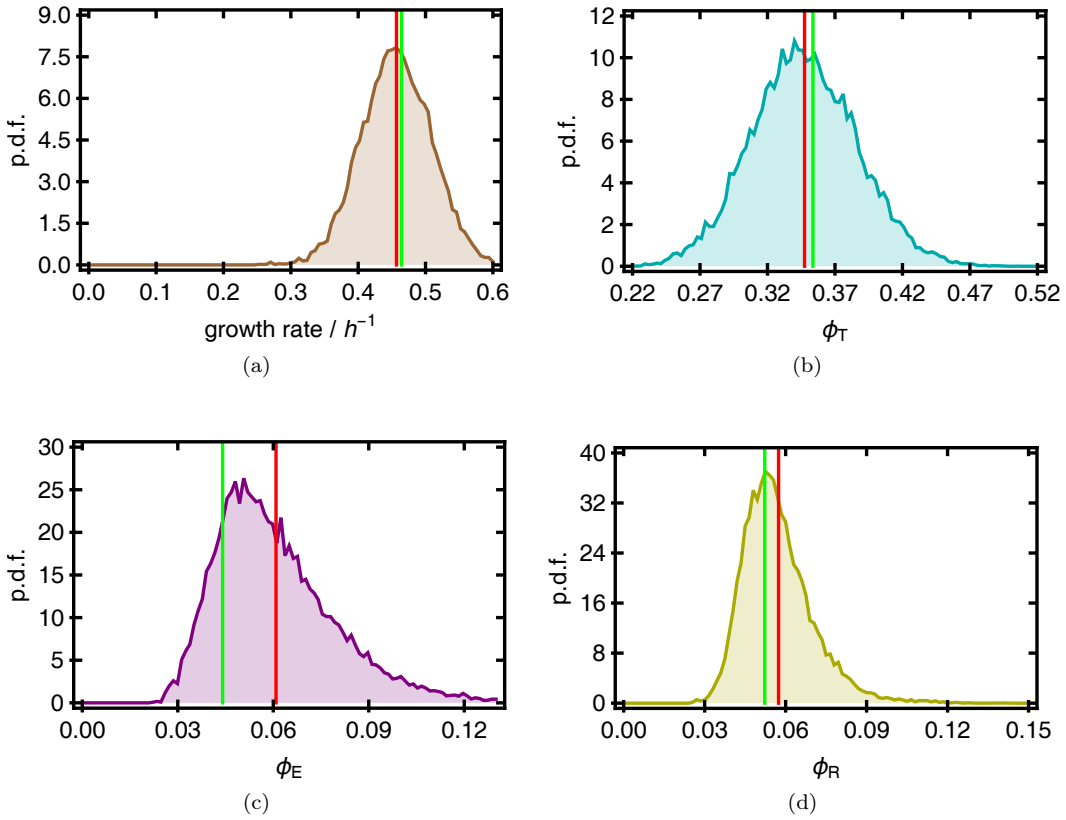


Figure 5.5: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 1.5$ with the optimal value denoted as a green line and the average value as a red line.

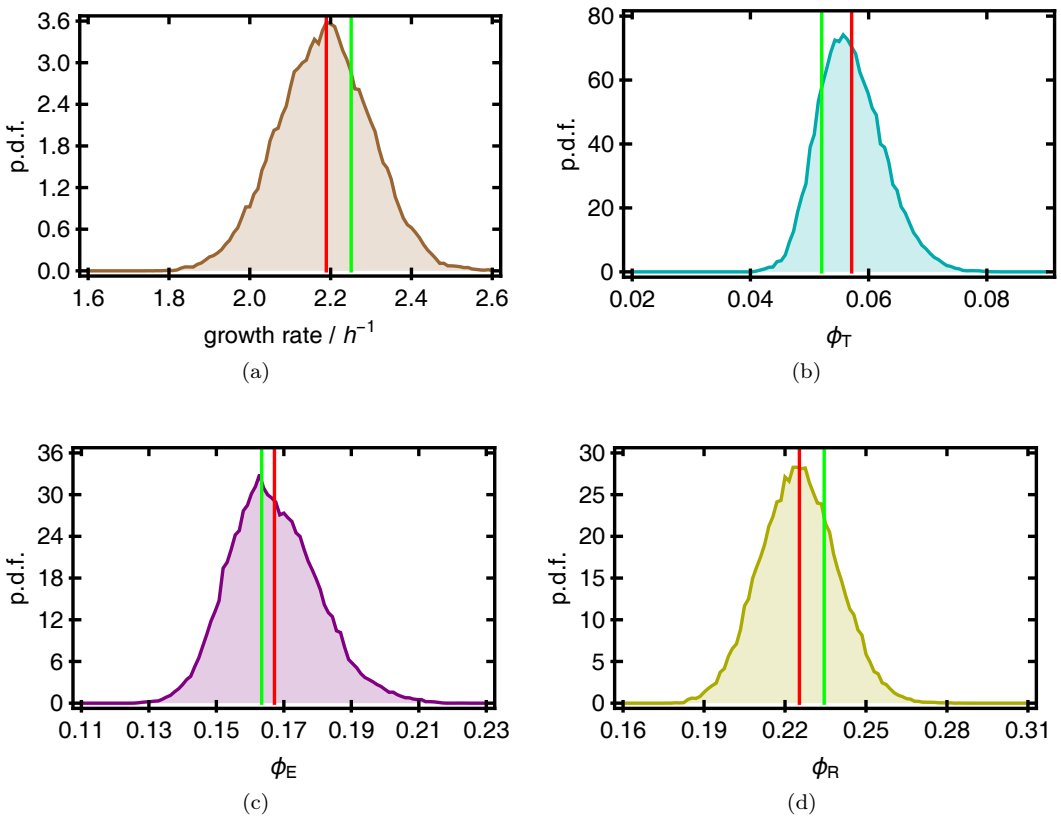


Figure 5.6: Histograms of (a) growth rate, (b) T fraction, (c) E fraction, (d) R fraction, for $k_T = 64$ with the optimal value denoted as a green line and the average value as a red line.

5.5 Crosscorrelations

We will now finally turn our attention to the effects that stochasticity and gene regulation have on each other. We will first compare the timescales of fluctuations in the fixed-rate and in the regulated system. We will then discuss the production rate versus growth rate, and proteome fraction versus growth rate crosscorrelations.

The timescale of fluctuations can be quantified by the autocorrelation function, which is essentially the crosscorrelation function of a quantity with itself. For example,

$$A_\mu(\Delta t) = X_{\mu-\mu}(\Delta t) = \frac{\int \mathrm{d}t\Big(\mu(t) - \tilde{\mu}\Big)\Big(\mu(t+\Delta t) - \tilde{\mu}\Big)}{\sigma_\mu^2}$$

is the autocorrelation function of μ. The autocorrelation function typically decays exponentially with time. The time scale over which the decay takes place is called the autocorrelation time τ. We have computed A_μ and A_ϕ for the different simulations that we have run and fitted exponential functions to them to obtain the autocorrelation times in Figure 5.7.

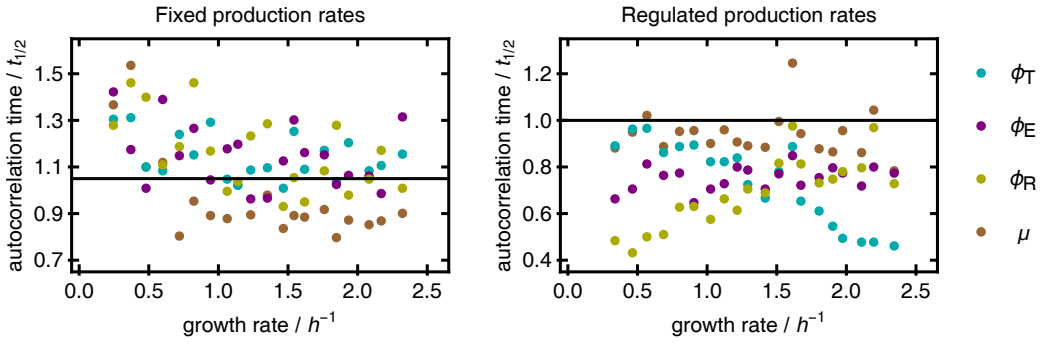
We can use this analysis to compare the system with fixed production rates to the system with regulated production rates. Firstly, the typical time scale in which fluctuations are diluted away is precisely that: the time scale of dilution, that is the doubling time. In the unregulated case, the autocorrelation time of the growth rate is smaller than that of the protein abundances. This occurs because the growth rate is a function of the three abundances. The growth rate will decorrelate whenever the first of the three abundances decorrelates. This will happen earlier than whenever the first of only one abundance decorrelates, which corresponds to the autocorrelation time of the abundances.

Furthermore, it is clear that the autocorrelation times of the three proteome fractions are smaller in the regulated system than in the fixed-rate system. This is the effect of regulation: if a perturbation in one proteome fraction occurs, the system will perceive this disturbance and change its production rates accordingly. In the case of, for example, an upfluctuation in the abundance of transporters, the rate of transporter production will decrease and the other rates will increase. This restores the balance more quickly than when the production rates would have been kept constant. Curiously, the autocorrelation time of the growth rate appears to be relatively constant throughout growth conditions and even between the two differently coordinated systems. We may speculate that this means that the fluctuations in the growth rate are highly influenced by the fluctuations in the abundance of the housekeeping sector. This makes sense, because the housekeeping proteins make up about half of the entire cell.

Additionally, we see in the regulated system that the autocorrelation times of $\phi_\mathrm{T}$and ϕ_R show clear, and opposite, trends with varying growth rate. The autocorrelation time of ϕ_T decreases with growth rate while the autocorrelation time of ϕ_R increases with growth rate.

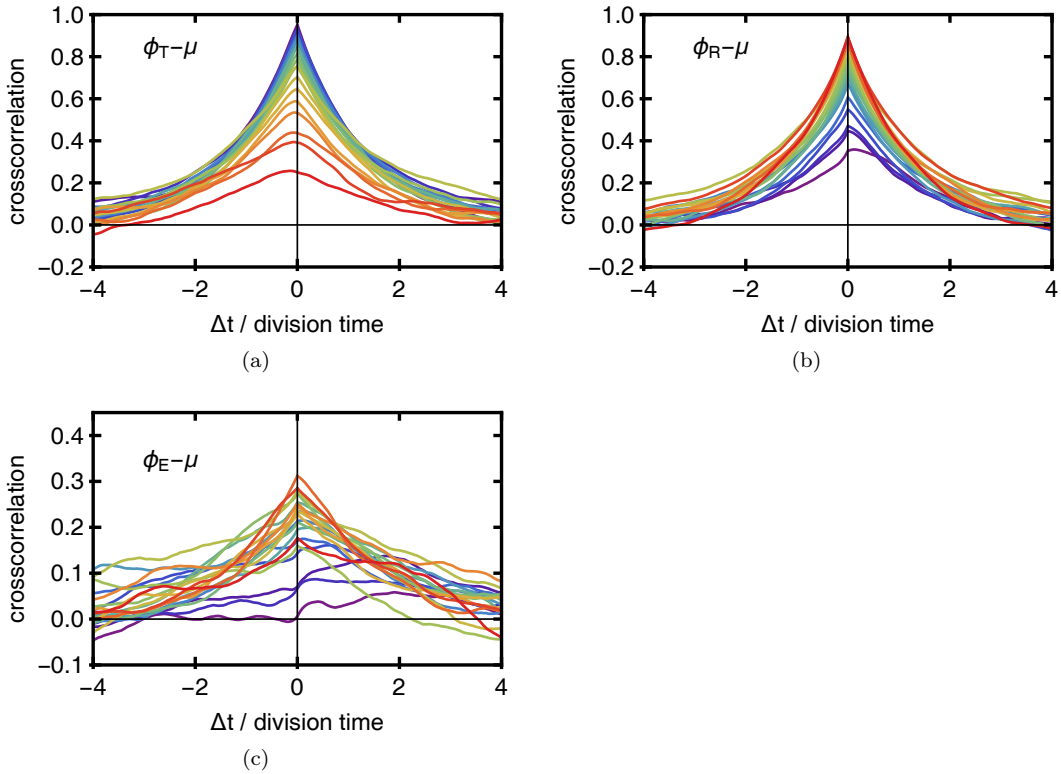
In addition to the autocorrelation times, we have also calculated the crosscorrelation between the proteome fractions and the growth rate for the regulated system, just as we did in Section 4.4 for the fixed-rate system. We show the results in Figure 5.8. The corrrelations (the crosscorrelations at zero lag Δt) paint the same picture as the analysis on the growth control coefficients. At small growth rate (blue), caused by low transporter efficiency k_T, the concentration (fractional abundance) of transporters is at its most limiting. This means that the $\phi_\mathrm{T} - \mu$ correlation is large. With increasing growth rate, the $\phi_\mathrm{T} - \mu$ correlation decreases, indicating a shift in limitation away from the transporters. In fact, the limitation shifts to the ribosomes, because these show the opposite trend.

The $\phi_\mathrm{T} - \mu$ crosscorrelation is especially interesting, because it was measured in [13]. Our transporter protein corresponds to the catabolic protein fraction, which is represented by the *lac* enzymes in the experiments. The experiments also see a decrease in the $\phi_\mathrm{T} - \mu$ crosscorrelation at zero lag upon increasing growth. In addition, they find a peak at negative lag Δt for high growth rates, when the $\phi_\mathrm{T} - \mu$ crosscorrelation at zero lag is almost zero. Our model is not able to reproduce this.



(a) (b)

Figure 5.7: Autocorrelation times of growth rate μ and proteome fractions ϕ_T, ϕ_E, and ϕ_R for varying k_T in the case of (a) fixed production rates (see chapter 4), and (b) regulated production rates. The autocorrelation times τ are plotted in units of the average cell cycle time, in the same way as the crosscorrelation functions from Figure 4.6. Values for τ were computed by fitting exponentials $\exp(-\frac{t}{\tau})$ to the autocorrelation functions for $0 < t < 5t_{\frac{1}{2}}$.



(a) (b) (c)

Figure 5.8: Crosscorrelations of the proteome fractions of (a) transporters ϕ_T, (b) ribosomes ϕ_R, and (c) enzymes ϕ_E, with the growth rate μ, as in Figure 4.6.

In addition to the crosscorrelation of proteome fractions with growth rate, we are able to compute the crosscorrelation of production rates with growth rate as well, because the production rates are no longer fixed. The results of this are plotted in Figure 5.9. The production rates are here

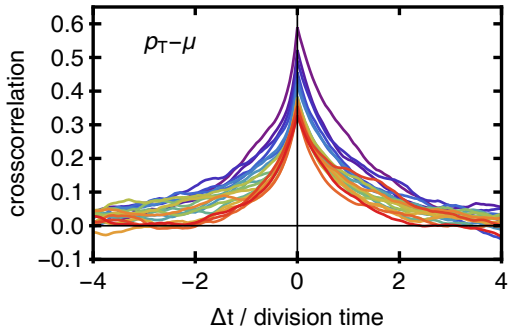
$$p_{\mathrm{X}} = r_{\mathrm{X}}\phi_{\mathrm{R}}\frac{a}{1+a}$$

This is the total production rate of protein X per unit of volume, including the regulated rate r_{X}, the concentration of ribosomes ϕ_{R} and the degree of saturation of the ribosomes by the amino acids $\frac{a}{1+a}$. We want to analyse this combined quantity, because p_{T} was measured in [13].

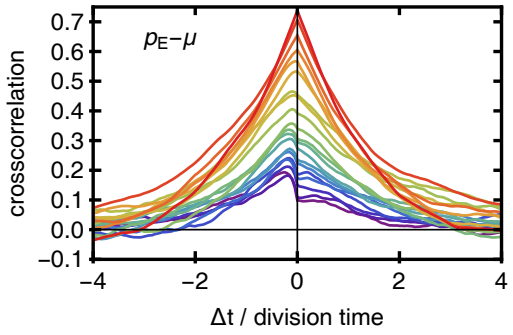
The crosscorrelation of p_{T} and μ in our model shows two features. Firstly, it decreases with increasing growth rate, just like the crosscorrelation of ϕ_{T} and μ. It also appears that the time scale (rescaled to the division time) at which the crosscorrelation decays becomes smaller when the growth rate increases. This is exciting, because the same phenomenon is visible in the experimental data. The data also show that the $p_{\mathrm{T}} - \mu$ crosscorrelation is skewed to positive lag Δt for small growth rate, and that the skewness decreases with increasing growth rate. The presence or absence of this effect in our model cannot be clearly determined from our simulation data, because the signal is too weak.

The crosscorrelations of p_{E} and p_{R} with μ have not been measured, but we can of course compute them anyway. They both look very similar and their magnitudes strongly increase with growth rate. It appears that they are dominated by the ϕ_{R}-contribution to the productions p_{E} and p_{R}. However, in conditions causing slow growth, there is an additional time scale visible in the crosscorrelations. They show a clear peak at small negative Δt. This means that, on average, an upward fluctuation in production of enzymes and ribosomes *follows* an upward fluctuation in the growth rate.

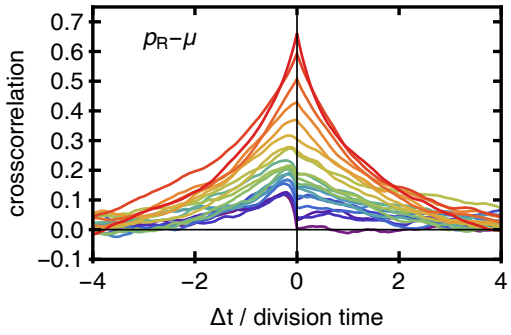
We cannot explain this effect, but we can speculate on the existence of an additional timescale. The effect occurs in cells that grow slowly, and are therefore quite small. In addition, enzymes and ribosomes occur in small copy numbers. These two considerations mean that a cell will produce only a handful of bursts of E (and R) proteins per cell cycle. One burst of E (or R) production will therefore have a large effect on the fractional abundance of E (or R), whereas one burst of T production will have only a small effect—it takes many bursts of T to even out the fluctuation caused by the one burst of E (or R). The time scale of the negation of one burst is therefore large at slow growth, and invisible at fast growth when compared to the cell cycle time.



(a)



(b)



(c)

Figure 5.9: Crosscorrelations of the production rates $p_X = r_X \phi_X \frac{a}{1+a}$ of (a) transporters ϕ_T, (b) ribosomes ϕ_R, and (c) enzymes ϕ_E, with the growth rate μ, plotted in a similar way as Figures 4.6 and 5.8.

6 Discussion

In this final section of my thesis, I will first describe the main results of our research. I will finish by mentioning several ways in which I think my work could be continued.

As we have seen in chapter 3, our minimal metabolic model reproduces the growth laws when its constitution maximises the growth rate for a given parameter set, even when stochasticity is introduced in the system. Moreover, we have implemented a simple model for regulation that still keeps the linear relationships between the average expression and growth rate intact. Interestingly, introducing stochasticity and regulation in the system creates small offsets in the growth laws.

We have introduced the growth control coefficients as a measure to quantify the coupling from protein abundance to growth rate. They describe by how much proteins are limiting growth. They also function as transmission coefficients for the propagation of fluctuations in protein copy numbers to fluctuations in growth rate. We have seen that the growth control coefficients are proportional to the fractional abundance of their respective proteins in the growth-maximised state. This means that stochasticity in the growth rate is mainly caused by stochasticity in the copy numbers of frequently occurring proteins.

The growth control coefficients are sensitive to perturbations away from the optimal state. In the unregulated, fixed-rate experiment we have seen that the GCCs were far from proportional to the fractional abundances, while in the regulated experiment they were closer to their optimal values. This indicates that regulation is able to keep the system from going far away from its optimum. This was verified by the distributions of protein copy numbers: these were considerably narrower in the regulated simulations than in the fixed-rate simulations.

In the regulated system, the average protein composition was suboptimal. However, the average growth rate was larger than in the unregulated system. Therefore the narrowing of the copy number distributions was more important than the suboptimal regulation.

We have also seen that the timescale of fluctuations is of the order of the cell cycle time. The timescale of fluctuations in gene copy numbers dropped when regulation was introduced to the system. Additionally, in the regulated system, the autocorrelation time of the transporter protein concentration decreases with increasing growth rate, while that of the ribosomal protein concentration increases.

The crosscorrelation functions of protein abundances versus growth rate show a gradual shift in limitation from transporters to ribosomes when the growth rate increases. Also, the time scale of the transporter production versus growth rate crosscorrelation decreases. These two results from our simulations agree with experiments. In addition to this, at slow growth we observe an additional peak in the crosscorrelations of the enzyme and ribosome production versus growth rate.

I will now discuss some limitations of our model, and suggest improvements and topics for further research. Firstly, we have used only four proteins in total, to describe four sectors of proteins. An obvious way to extend the model would be to increase the number of proteins in order to include additional protein sectors, and investigate the effect of additional ways of limitation. For instance, it may be worthwile to separate the catabolic sector into multiple parts, such that one of them may be controlled in the same way as IPTG concentrations control the *lac* protein expression in the experiments.

Another way of extending the proteome would be to include more details on the regulation of ribosomal proteins. This would likely prompt us to distinguish transcription and translation, so as to include the effects of rRNA regulation.

Furthermore, fluctuations in the housekeeping fraction remain large, because they are unregulated. It should be interesting to investigate possible mechanisms of Q-sector regulation.

A problem of our model is the following: because we use so few enzymes, the cell size is artificially small. This means that fluctuations in the total protein number, due to direct stochasticity in the protein copy numbers, are relatively large compared to actual cells. We

would expect that stochasticity in the size of real cells is dominated by long-time stochasticity in the regulation of the production rates rather than stochasticity in the copy numbers. This may be solved by relaxing the requirement that the cell size is equal to the total protein number. However, additional regulation is required to ensure that fluctuations in the average density remain small.

On a related note, in our model fluctuations in one enzyme have a strong effect on the concentrations (proteome fractions) of the other proteins, due to the effect of dilution. This important feature led us to develop the growth control coefficients in terms of the total abundances. When the cell contains many different proteins that all occur in low copy numbers, the concentrations are a good measure of the amounts of proteins present. In that case, the flux control coefficients may be sufficient.

A different limitation of our analysis is, that our simulations gathered data on lineages of bacteria. Our analysis is not directly applicable to experimental data on populations, where faster growing cells are selected for and therefore occur more frequently. In other words, the time average over a single lineage differs from the ensemble average over a population of cells. It is possible to correct for this by selecting daughter cells carefully at the time of cell division.

In summary, the modelling approach that I presented in this thesis is able to describe both intracellular stochasticity and culture-level regularities reasonably accurately, even though it glosses over many specific details. Coarse-grained modelling in the spirit of this thesis is a valuable tool to furthen our understanding of complex networks, and I believe that it can be used to study stochasticity in biological cells in more detail.

Acknowledgements

I am thankful for all the help I got while conducting this research. My sincere thanks go to my supervisor, Dr Rutger Hermsen, for focusing my attention in many a long meeting, planned or spontaneous, for letting me experience what it means to be a scientist, and for convincing me that an education in physics can be well spent in biology.

I am grateful to the ITF, in particular Prof René van Roij, for giving me the opportunity to carry out my Master's research externally, and to the TBB group for welcoming me with open arms.

I further thank Prof Sander Tans and his group, particularly Martijn Wehrens and Noreen Walker, for several interesting discussions about the stochasticity experiments, Prof Frank Bruggeman for stressing the difference between lineage and population dynamics, and fellow group members Hilje Doekes, Misha Sheinman, and Laurens Krah for their helpful suggestions and comments.

References

[1] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[2] Lev S. Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, 2014.

[3] Vahid Shahrezaei and Samuel Marguerat. Connecting growth with gene expression: Of noise and numbers. *Current Opinion in Microbiology*, 25:127–135, 2015.

[4] Jacques Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.

[5] Moselio Schaechter, Ole Maaløe, and Niels O. Kjeldgaard. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Journal of General Microbiology*, 19(3):592–606, 1958.

[6] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330:1099–1102, 2010.

[7] Conghui You, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim, Carl W. Gunderson, Yi-Ping Wang, Peter Lenz, Dalai Yan, and Terence Hwa. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500:301–306, 2013.

[8] Sheng Hui, Josh M. Silverman, Stephen S. Chen, David W. Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R. Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11:784, 2015.

[9] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[10] Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz. Gene regulation at the single-cell level. *Science*, 307:1962–1965, 2005.

[11] Mary J. Dunlop, Robert Sidney Cox, III, Joseph H. Levine, Richard M. Murray, and Michael B. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, 40(12):1493–1498, 2008.

[12] Istvan T. Kleijn. *Single-cell variability in a bacterial signalling network investigated by FRET*. Bachelor's thesis, Utrecht University, 2013.

[13] Daniel J. Kiviet, Philippe Nghe, Noreen Walker, Sarah Boulineau, Vanda Sunderlikova, and Sander J. Tans. Stochasticity of metabolism and growth at the single-cell level. *Nature*, 514:376–379, 2014.

[14] Herbert E. Kubitschek, William W. Baldwin, Sally J. Schroeter, and Rheinhard Graetzer. Independence of buoyant cell density and growth rate in *Escherichia coli*. *Journal of Bacteriology*, 158(1):296–299, 1984.

[15] Herbert M. Sauro. Appendix C: Enzyme kinetics in a nutshell. In *Control Theory for Bioengineers*, pages 375–382. Ambrosius Publishing, version 0.57, first edition, 2015.

[16] Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.

[17] Ingrid M. Keseler, Amanda Mackie, Martin Peralta-Gil, Alberto Santos-Zavaleta, Socorro Gama-Castro, César Bonavides-Martínez, Carol Fulcher, Araceli M. Huerta, Anamika Kothari, Markus Krummenacker, Mario Latendresse, Luis Muñiz Rascado, Quang Ong, Suzanne Paley, Imke Schröder, Alexander G. Shearer, Pallavi Subhraveti, Mike Travers, Deepika Weerasinghe, Verena Weiss, Julio Collado-Vides, Robert P. Gunsalus, Ian Paulsen, and Peter D. Karp. EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Research*, 41(D1):D605–D612, 2013.

[18] Michael A. Savageau. *Biochemical Systems Analysis.* Addison-Wesley, Reading, 1976.

[19] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of *Escherichia coli*. *Current Biology*, 20(12):1099–1103, 2010.

[20] Srividya Iyer-Biswas, Charles S. Wright, Jonathan T. Henry, Klevin Lo, Stanislav Burov, Yihan Lin, Gavin E. Crooks, Sean Crosson, Aaron R. Dinner, and Norbert F. Scherer. Scaling laws governing stochastic growth and division of single bacterial cells. *Proceedings of the National Academy of Sciences*, 111(45):15912–15917, 2014.

[21] Sattar Taheri-Araghi, Serena Bradde, John T. Sauls, Norbert S. Hill, Petra Anne Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun. Cell-size control and homeostasis in bacteria. *Current Biology*, 25(3):385–391, 2015.

[22] Suckjoon Jun and Sattar Taheri-Araghi. Cell-size maintenance: Universal strategy revealed. *Trends in Microbiology*, 23(1):4–6, 2015.

[23] Ariel Amir. Cell size regulation in bacteria. *Physical Review Letters*, 112(20):208102, 2014.

[24] William D. Donachie. Relationship between cell size and time of initiation of DNA replication. *Nature*, 219:1077–1079, 1968.

[25] Stephen Vadia and Petra Anne Levin. Growth rate and cell size: A re-examination of the growth law. *Current Opinion in Microbiology*, 24:96–103, 2015.

[26] Lydia Robert. Size sensors in bacteria, cell cycle control, and size control. *Frontiers in Microbiology*, 6:515, 2015.

[27] Nicolaas G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier, Amsterdam, 2nd edition, 1992.

[28] Ramon Grima, Nils G. Walter, and Santiago Schnell. Single-molecule enzymology à la Michaelis–Menten. *FEBS Journal*, 281:518–530, 2014.

[29] Daniel T. Gillespie, Andreas Hellander, and Linda R. Petzold. Perspective: Stochastic algorithms for chemical kinetics. *The Journal of Chemical Physics*, 138:170901, 2013.

[30] Daniel T. Gillespie. Simulation methods in systems biology. In Marco Bernardo, Pierpaolo Degano, and Gianluigi Zavattaro, editors, *Formal Methods for Computational Systems Biology*, pages 125–167. Springer Berlin Heidelberg, 2008.

[31] Walter E. Brown. Random number generation in C++11. Technical report, 2013.

[32] Long Cai, Nir Friedman, and X. Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440:358–362, 2006.

[33] Evert Bosdriesz, Douwe Molenaar, Bas Teusink, and Frank J. Bruggeman. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS Journal*, 282(10):2029–2044, 2015.

[34] Matthew Scott, Stefan Klumpp, Eduard M. Mateescu, and Terence Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10(8):747, 2014.

[35] Henrik Kacser, James A. Burns, and David A. Fell. The control of flux. *Biochemical Society Transactions*, 23(2):341–366, 1995.

[36] Nir Friedman, Long Cai, and X. Sunney Xie. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Physical Review Letters*, 97:168302, 2006.

[37] Sorin Tănase-Nicola and Pieter Rein ten Wolde. Regulatory control and the costs and benefits of biochemical noise. *PLoS Computational Biology*, 4(8):e1000125, 2008.

A Fit parameters

Table A.1: Fit parameters with uncertainties for the linear fits that determine the growth laws, in the form $\phi_{\mathrm{X}}(\mu) = \phi_{\mathrm{X},0} + \alpha_{\mathrm{X}}\mu$.

system	protein X	$\phi_{\mathrm{X},0}$	α_{X}
optimal (Figure 3.2a)	T	0.4334 ± 0.0024	-0.1722 ± 0.0016
	E	0.0120 ± 0.0016	0.0691 ± 0.0011
	R	0.0046 ± 0.0008	0.1031 ± 0.0005
fixed rates (Figure 4.1a)	T	0.418 ± 0.004	-0.172 ± 0.003
	E	0.0266 ± 0.0015	0.0643 ± 0.0010
	R	0.0174 ± 0.0011	0.1017 ± 0.0008
regulated (Figure 5.2a)	T	0.415 ± 0.004	-0.1660 ± 0.0024
	E	0.0345 ± 0.0012	0.0616 ± 0.0008
	R	0.0145 ± 0.0003	0.09654 ± 0.00019

Table A.2: Fit parameters with uncertainties for the fits to the Monod curves, in the form $\mu = \mu_{\max}\frac{k_{\mathrm{T}}}{K+k_{\mathrm{T}}}$.

system	$\mu_{\max}\,(\mathrm{h}^{-1})$	$K\,(\mathrm{h}^{-1})$
optimal (Figure 3.2b)	2.480 ± 0.011	6.90 ± 0.10
fixed rates (Figure 4.1b)	2.439 ± 0.008	8.15 ± 0.08
regulated (Figure 5.2b)	2.414 ± 0.013	6.96 ± 0.12