Università degli Studi di Padova

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in
Control Systems Engineering

Including Metabolic Load in Bacterial Growth Rate: a Systems Biology Approach

Advisor:
Prof. Luca Schenato

Co-advisor:
Dr. Massimo Bellato

Student:
Elisa Gaetan
2020375

Academic Year 2021/2022
3 October 2022

Abstract

In this thesis, different mathematical models which describe the cell growth in bacterial cells, are studied. In particular the effect of metabolic load is explored.
Beyond the modeling of the cell system, the identification of the most suitable growth rate law represents a key point in the whole analysis.
Despite several growth rate functions are proposed in the literature, most of them are applied to a specific case, namely to a cell in its unperturbed configuration (i.e., no metabolic load is added), which is a particular case.
Since the results are strongly influenced by the growth rate, first the already developed (from the current literature) growth functions were tested using mathematical simulations, in particular highlighting the possible disagreements with the biological knowledge, which allow to reject some of them.
Furthermore, some alternative functions from personal intuitions based on biological evidence were proposed and tested as well.
The final goal is to identify a growth rate function that can be applied to models with different levels of complexity, leading to coherent results.

Contents

Chapter 1

Introduction

1.1 Context and Motivation

In the last years, the study of the biological systems have caught more importance: a deeper understanding of the inner mechanisms of cells, tissues, organs and human body can bring significant contributions in the development of new drugs, therapies, other than biomedical devices in a more general sense.
In this thesis, the system of a single cell of *Escherichia Coli* will be modeled and its growth will be analyzed; in particular the effects of a metabolic load on it will be examined. In other words, the cell will be seen as input-output system, where the load is the input and the growth rate the output.

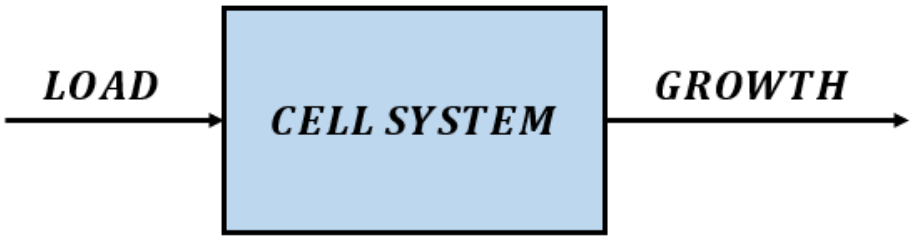

Figure 1.1: Cell as input-output system.

Furthermore, it is also possible to extend the case of study, considering a population of cell and analyzing the relations between them or how they influence each other.
In general, the comprehension of how the cell (or the population) behaves in different configurations and conditions (in other words the design of a realistic model that best approximates the true system) plays a fundamental role in many disciplines. Additionally, it can be used to estimate in advance the results of the laboratory experiments.
Here the discipline of systems biology plays a fundamental role, in particular for the derivation of the mathematical models, which is the preliminary step before to proceed with the analysis of the load effects on the cell growth.

1.2 Systems Biology and its importance

Systems Biology has been defined as an alternative approach in the research field of biology, biochemistry, immunology, biotechnology and biomedical engineering.

Systems biology is based on the understanding that the whole is greater than the sum of the part.

– Dr. Nitin Baliga

Instead of studying the singular components, this discipline is focused on the analysis of the whole system. It can be applied at different levels: organism, tissue or cell. In this particular case, it will be used for the modeling of a single cell system.
It has been introduced for facing the complexity of the biological systems: indeed it was not worth to study each component singularly and then to sum all the contributions, because the relations between the parts can not be sufficiently highlighted. Therefore the understanding of the inner networks has been preferred to the analysis of the single elements.
Additionally, systems biology integrates several scientific fields as engineering; according to that, several concepts and approaches play a fundamental rule for the comprehension and analysis, such as systems theory and mathematical modeling. Furthermore the co-operation between these two disciplines provides new methods for facing the challenges in both areas. Moreover, the possibility to model the biological systems allows to study their dynamics in a simulated environment, that permits to avoid (at least in a first moment) all the problems and/or challenges in the experimental design. At the same time, this type of approach may consent to eliminate some hypotheses or theses, reducing the number of experiments to perform. Thus, this should lead also to the reduction of the research costs.

1.3 Challenges

The challenges handled in this thesis were basically two: to model the cell system and to define a growth rate function.
The first challenge will be faced considering a pool of chemical reactions that describe the processes that occur in the cell; these have been derived from the available literature on the cell biology. Since it is not feasible to examine the totality of the possible reactions that take place in the system, only the most significant for the declared aim will be taken into account. In other words, during the modeling process, some variables will be considered constant or some reactions will be approximated, in order to reduce the number of components that characterize the system.
The second one has been revealed as the most challenging. Most of the articles in the literature consider the growth rate as a constant so just few examples of the growth rate of the bacterial cell can be found and, additionally, no data sets are available, making an eventual process of estimation and/or fitting unfeasible. Therefore, from the available articles, the proposed growth rate functions will be examined and tested with the designed models. Additionally, other functions will be derived based on biological knowledge.

1.4 Personal contribution

Even though it has been proved (by some laboratory experiments) that the integration of a burden in the cell modifies the growth rate, the current literature does not characterize it with a suitable model.
Indeed, as explained previously, the studies on cell modeling do not consider a variable growth rate or, in the best case, they evaluate it in the unperturbed case, namely when there is no metabolic load added. This is a clear limitation and the goal of this thesis is to provide an original model that can be efficiently used to predict the cell behavior (in terms of growth rate) in the loaded case.
Firstly, mathematical models will be derived in order to represent the cell system with a sufficient level of description; afterwards several growth rate functions will be tested and the obtained results will be compared with the biological evidences in order to validate or reject the chosen function. Finally, the most suitable growth rate function for the characterization of the cell growth will be identified.
It is interesting to highlight the future potentialities of this approach. In fact, once a detailed model of the system has been derived, the load in input can be used to control the growth of the cell, namely it can be seen as control input. In this way, a dangerous bacterium may be killed modifying the amount of proteins to be synthesized inside it.
Future studies and developments could be dedicated to implementing and designing a suitable controller. Afterwards, involving further research fields, such as synthetic biology, this controller may be physically assembled.

Chapter 2

Systems Biology of Cell Growth

2.1 Modeling in Systems Biology

In this first section, the bases for the modeling of biological and biochemical systems will be established.
First, it will be specified how to derive the set of differential equations starting from the biochemical reactions, and then two possible analyses of this pool of ODEs will be presented. Afterwards, the biological scenario will be introduced, in order to provide an overview of all the components of the cell, before to proceed with the actual modeling process.

2.1.1 From Kinetics equations to ODE

The biological systems can be modeled in different ways, depending on the level of description and resolution wanted.
However, the shared principle behind all the possible approaches is that the model is used for analysing and making predictions on already existing systems. Starting from the chemical reactions which occur in it, the aim is to derive a set of differential equations.
In this thesis, the chemical reactions have been modelled using the *Reaction Rate Equations* method. According to this method, the system is considered as a set of species S_i that interact with each other and in particular it takes into account the concentration x_i:

$$x_i = [S_i] = \frac{n_{S_i}}{\Omega} \tag{2.1}$$

where n_{S_i} is the number of molecules of the species S_i and Ω is the given volume.
Furthermore, the application of this method is based on a strong assumption: all the reactions occur in a well-stirred volume, which means that the rate of interaction between two species is uniform and it does not have any spatial influence.
The final aim is, as already declared, to describe the system with a set of differential equations

$$\dot{x} = f(x, \theta) \tag{2.2}$$

where $x \in \mathbb{R}^n$ is the vector that contains all the species of the system, $\theta \in \mathbb{R}^p$ is the vector of the parameters and $f : \mathbb{R}^n \times \mathbb{R}^p \longrightarrow \mathbb{R}^n$ rules the change in the concentrations.

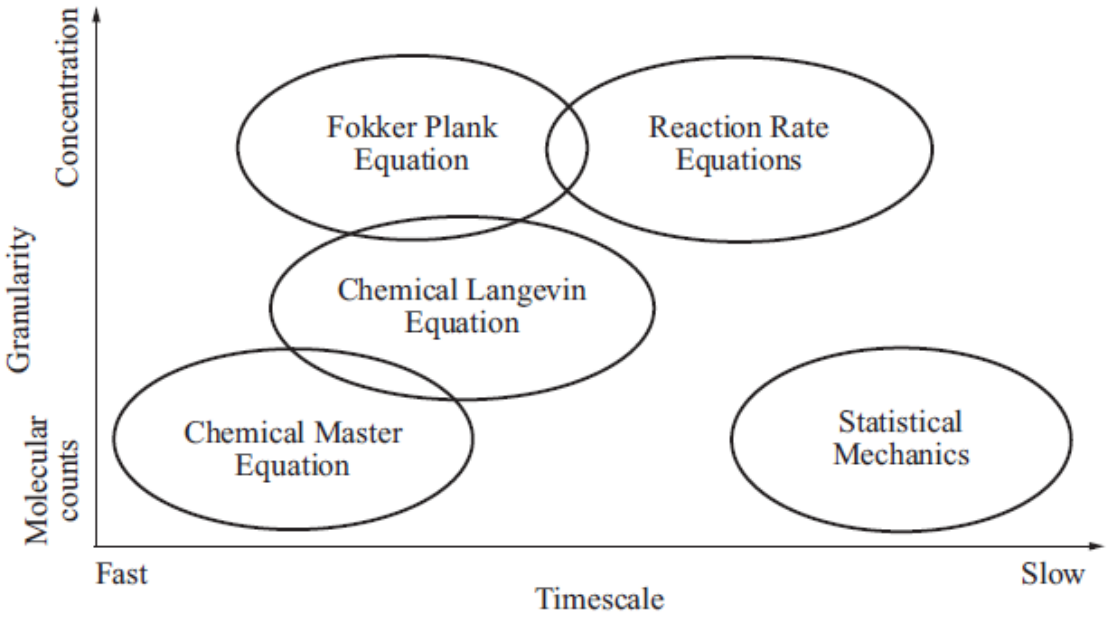

Figure 2.1: Different methods of modeling biomolecular systems [6].

In order to describe the process of derivation of the differential equations, a basic biomolecular reaction is proposed as example, namely:

$$\mathrm{A} + \mathrm{B} \underset{\mathrm{k}_r}{\overset{\mathrm{k}_f}{\rightleftharpoons}} \mathrm{AB} \tag{2.3}$$

The reaction 2.3 can be interpreted as follows: every time the forward reaction occurs, the number of molecules of A (n_A) and B (n_B) must be decreased by one, while the number of molecules of AB (n_{AB}) must be increased by one. The reasoning is the same with the reverse reaction (n_{AB} must be decreased while n_A and n_B must be increased by one). However, just one reaction at a time can occur and it is regulated by the likelihood. In particular, the likelihood of the forward reaction (in a time interval dt) is:

$$a_f(q)dt = (k_f/\Omega)n_A n_B dt \tag{2.4}$$

where q is the configuration of the system and k_f is the parameter that depends on the reaction (rate of association). For the reverse reaction, the likelihood is:

$$a_r(q) = k_r n_{AB} \tag{2.5}$$

where k_r still depends on the reaction (dissociation rate).
Now the equation that expresses the variation of n_{AB} can be written considering also the contribution of the likelihoods:

$$n_{AB}(t + dt) = n_{AB}(t) + a_f(q - \xi_f)dt - a_r(q)dt \tag{2.6}$$

The number of molecules after a time interval dt is the amount at the previous instant n_{AB} plus the contribution of the forward reaction (ξ_f represents the change in the configuration) and minus the contribution of the reverse reaction. Both the contributions of

the likelihood can be roughly seen as the probabilities that the reaction takes place.
To convert 2.6 into an equation that involves the concentrations, it is enough to divide each term for the given volume Ω and to substitute the expression of the likelihoods with 2.4 and 2.5:

$$[AB](t+dt) = [AB](t) + (k_f/\Omega^2)n_A n_B dt - k_r n_{AB}/\Omega dt \tag{2.7}$$

Furthemore, since $n_A/\Omega = [A]$, $n_B/\Omega = [B]$ and $n_{AB}/\Omega = [AB]$:

$$[AB](t+dt) - [AB](t) = (k_f[A][B] - k_r[AB])dt \tag{2.8}$$

If $dt \longrightarrow 0$:

$$\frac{d}{dt}[AB] = k_f[A][B] - k_r[AB] \tag{2.9}$$

which is the differential equation that regulates the changes in concentration of the species AB.
In a quite similar way, the equations of A and B can be derived:

$$\frac{d}{dt}[A] = k_r[AB] - k_f[A][B] \qquad \text{and} \qquad \frac{d}{dt}[B] = k_r[AB] - k_f[A][B] \tag{2.10}$$

For sake of simplicity, in the next chapter the concentrations of the species will not be indicated with the squared brackets (i.e $[AB] \longrightarrow AB$).

2.1.2 Rapid Equilibrium versus Steady State assumption

Once the differential equations have been derived, the dynamics can be investigated.
For a generic system, the study of the equilibrium points is usual performed. The configuration that the system reaches at the equilibrium is called *Steady State* and it can be obtained setting all the differential equations to zero:

$$\frac{df(x,\theta)}{dt} = 0 \tag{2.11}$$

However, a different analysis can be executed with biological system, which takes into account the velocity of the reactions and it is called *Rapid Equilibrium* [12].
Consider a generic enzymatic reaction:

$$\mathrm{E} + \mathrm{S} \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{C} \xrightarrow{\mathrm{k}} \mathrm{E} + \mathrm{P} \tag{2.12}$$

where E is the enzyme, S is the substrate, C the complex and P the final product. Furthermore, a and d are the association and dissociation constant respectively, while k is the catalytic rate constant.
The corresponding set of differential equation (without consider any degradation) is:

$$\dot{S} = -aES + dC \tag{2.13}$$
$$\dot{E} = -aES + dC + kC \tag{2.14}$$
$$\dot{C} = aES - (d+k)C \tag{2.15}$$
$$\dot{P} = kC \tag{2.16}$$

Nevertheless, it can be assumed that the first reaction in 2.12, namely the formation of the complex C is faster than the synthesis of the final product P and consequently it

reaches first the equilibrium.
Setting equation 2.13 equal to zero, one obtains:

$$\frac{ES}{C} = K_d \qquad \text{where } K_d = \frac{d}{a} \text{ is called } \textit{dissociation constant} \tag{2.17}$$

In addition, the total amount[1] of the enzyme $E^{TOT} = E + C$ is usually constant at the equilibrium, thus it can be substituted into the equation at the equilibrium 2.17:

$$\frac{(E^{TOT} - C)S}{C} = K_d \qquad \longrightarrow \qquad C = \frac{S}{K_d + S} E^{TOT} \tag{2.18}$$

Going back to the starting enzymatic reaction, now the amount of product P at the *Rapid Equilibrium* can be computed as follows:

$$P = kC = kE^{TOT}\frac{S}{K_d + S} = P_{max}\frac{S}{K_d + S} \tag{2.19}$$

which is called *Michaelis-Menten kinetics* and it describes the evolution of the final product P as function of the substrate S (Figure 2.2).

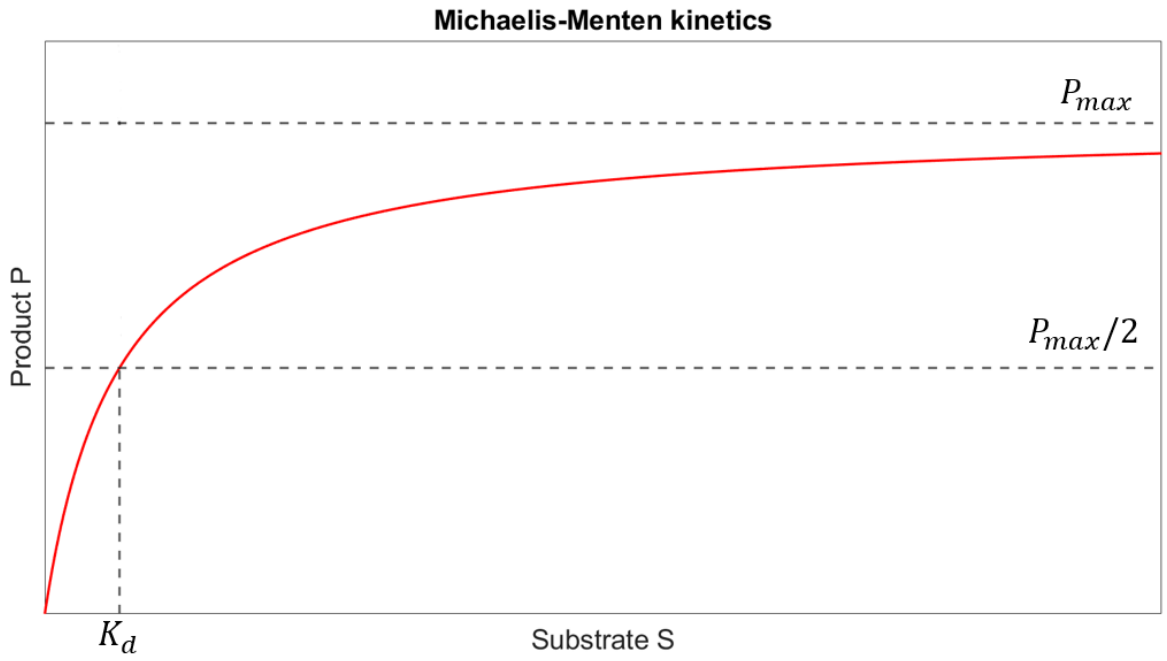

Figure 2.2: Michaelis-Menten kinetics

It may be interesting to compare this result to the *Steady State* one. From equation 2.14:

$$\dot{E} = -aES + dC + kC = -a(E^{TOT} - C)S + dC + kC = 0 \tag{2.20}$$

and then:

$$C = \frac{E^{TOT}}{1 + \frac{d+k}{aS}} = \frac{E^{TOT}}{1 + \frac{K_m}{S}} \qquad \text{where } K_m = \frac{d+k}{a} \text{ is called } \textit{half saturation constant} \tag{2.21}$$

[1] The total amount of a species S_i^{TOT} is defined as summation of the species itself S_i and the complexes that are directly derived from it.

Hence, the product P at the *Steady State* is:

$$P = kC = kE^{TOT}\frac{S}{K_m + S} = P_{max}\frac{S}{K_m + S} \tag{2.22}$$

which is instead called *Briggs-Haldane kinetics.*
In conclusion, the *Rapid Equilibrium* study is based on the assumption $k << d$, which means that the association and dissociation of the substrate are faster than the product formation. Instead, the *Steady State* does not consider any relation between the rates of the reaction. Due to this considerations it is possible to assert that the Michaelis-Menten is a special case of Briggs-Haldane.
It is worth to recall that the set of ODEs (from 2.13 to 2.16) does not take into account any degradation of the components. However, if it would be consider, the equation of P (2.19) at *Rapid Equilibrium* would not change, since it is based on the velocity of the reactions. Instead, the computation of the equilibria at the *Steady State* would lead to a different formulation compared to 2.22.

2.1.3 Biological Scenario

Now the bases for the modeling of the biochemical systems have been established.
Before to proceed, the biological scenario must be introduced, namely the chemical reactions and the components involved.
In this dissertation, the growth of a single cell of *Escherichia Coli* has been analysed.
As already specified in Paragraph 2.1.1, the cell system is modelled starting from a pool of chemical reactions. These can be divided into four main categories: transcription, translation, degradation and synthesis, which are represented in Figure 2.3.

***Transcription*:** Its standard form is:

$$\mathrm{D} + \mathrm{RNAP} \underset{\mathrm{p}^-}{\overset{\mathrm{p}^+}{\rightleftharpoons}} \mathrm{CTR} \xrightarrow{\omega} \mathrm{D} + \mathrm{m} + \mathrm{RNAP} \tag{2.23}$$

During this phase, the *DNA* (D) is converted into *mRNA* (m). More specifically, this happens with the support of the *RNA* polymerase ($RNAP$), which is able to "open" the double helix of the *DNA*. This initial process forms an open complex (CTR). Afterwards, the $RNAP$ can proceed with the formation of the *mRNA* sequence, which contains also a region called *Ribosome Binding Site* (RBS), needed for the next process.
In the following, the reaction of the transcription will be approximated[2] as:

$$\mathrm{D} \xrightarrow{\omega} \mathrm{D} + \mathrm{m} \tag{2.24}$$

***Translation*:** Its standard form is:

$$\mathrm{m} + \mathrm{R} \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{c} \xrightarrow{\beta} \mathrm{m} + \mathrm{R} + \mathrm{P} \tag{2.25}$$

It follows the *Transcription*. Once the *mRNA* is produced, the ribosome R binds it in the RBS region, forming the complex c. Then actual translation process can begin (reaction 2.25). The final product is the protein P.

[2]Under the assumption that $RNAP$ is constant (further details in Appendix B.).

***Degradation*:** All the components are subjected to decay. This is pointed out by the reaction of degradation, which is:

$$\mathrm{A} \xrightarrow{\psi_\mathrm{A}} 0 \tag{2.26}$$

where A is a generic component and ψ_A is the rate of death that depends on the component. In the following, it will be explained that μ_A is the actual summation between the spontaneous decay of A and the growth rate of the cell (its importance will be discussed in the next paragraph).

***Synthesis*:** The last reaction to be considered is the (generic) synthesis. Its general form contains two reactants A and B and a final product C (usually one reactant is a protein, while the other is a RNA):

$$\mathrm{A} + \mathrm{B} \xrightarrow{\epsilon} \mathrm{C} \tag{2.27}$$

where ϵ is the rate of association between the reactants. This reaction will be mainly used for the process of ribosomes formation.

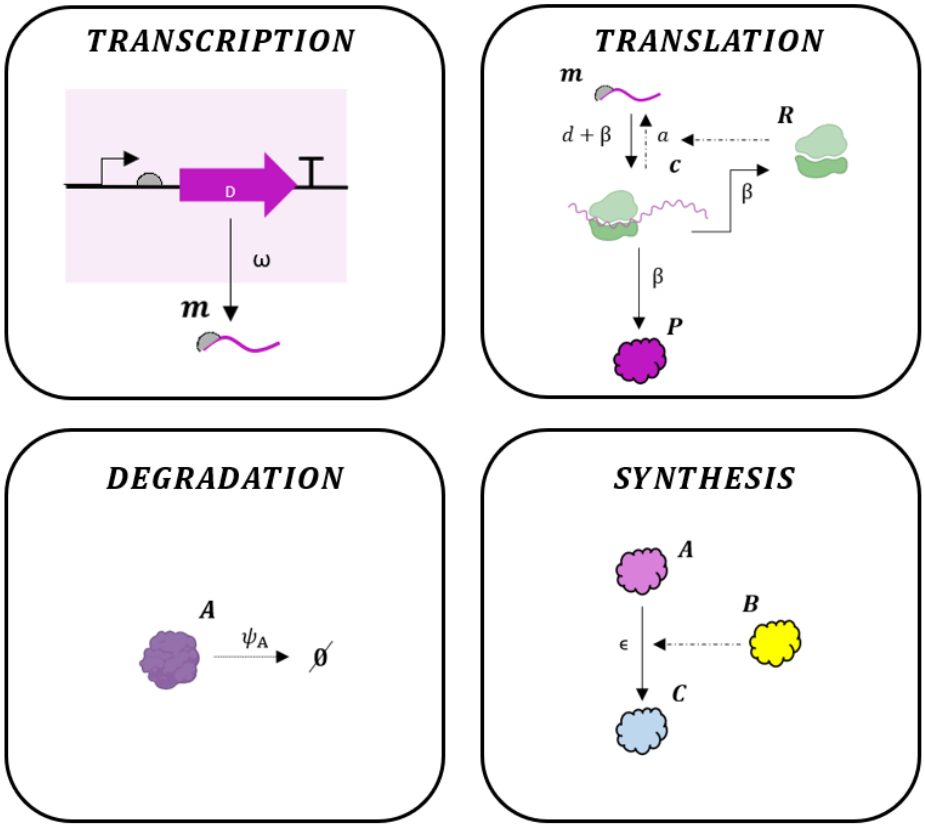

Figure 2.3: Sketches of chemical reactions

In particular, the transcription and translation together lead to the protein synthesis, which is the main biochemical process that takes place in the cell. Indeed proteins are responsible for many aspects of the cellular life, including cell shape and inner organization, product manufacture and waste cleanup, and routine maintenance; furthermore

they receive signals from outside the cell and mobilize intracellular response ([14]). Even though different proteins[3] are assigned to different tasks, all of them share one component: the ribosome R. From this statement, it is straightforward to understand the importance of the ribosomes; they are constituted by some of the 52 ribosomal proteins contained in the cell (*R-Proteins*, P) and three ribosomal *RNAs* (*rRNA* r,[10],[13]), namely:

$$\mathrm{P} + \mathrm{r} \xrightarrow{\sigma_p} \mathrm{R} \tag{2.28}$$

Additionally, since the cell system is very complex and it contains many different species, some have been gathered together based on their components and/or their final products.

***Ribosomal* species :** It is a general compartment (Figure 2.4) that groups together all the species that involve ribosomes R. In particular, in the following there will assume the existence of only three species.

***Basal* species B:** It gathers together all the species that are fundamental for the survival of the cell (i.e., they produce proteins that are needed for the basic cell life).

***R-Protein* species P:** It is known from the literature that the bacterial cell contains 52 types of *R-Proteins* ([13]), which are proteins that participate to the ribosomes synthesis. Since it is not worth to consider them singularly[4], they have been gathered into the more general *R-Protein* species.

***Load* species L:** The load species produces a protein that requires a not negligible amount of energy and consequently influences the system. Usually the green fluorescent protein (*GFP*) is used in the laboratory experiments, since it can be quantified using the spectrophotometer [5].

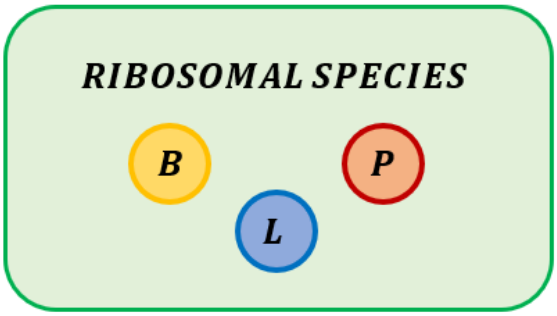

Figure 2.4: Compartment of ribosomal species

[3]Different proteins come from different *mRNAs*.

[4]If every type of *R-Protein* would be considered, there would need 52 equations just for describing the *R-Proteins* dynamics.

[5]Roughly, the sample is crossed by a beam and the device measures the intensity of the exiting light radius as function of the wavelength. In biochemical experiments, the spectrum allows to quantify the level of expression of the proteins.

2.1.4 Biological Feedbacks

The importance of the ribosomes has been already introduced in the previous paragraph. Therefore, the cell tends to control their production in order to guarantee the right amount in order to fulfill the metabolic requirement.
It is known from the literature that there exists an inner feedback that rules the synthesis of the ribosomes and its aim is to prevent to accumulation of them in the cell. It may sound counter intuitive, since the more available ribosomes, the more synthesized proteins, the more fulfilled tasks.
However, the reason why the cell seeks to limit the concentration of ribosomes lies in the needed energy. Indeed, as all the cellular processes, the ribosomes synthesis require energy (i.e. ATP), it is not worth to produce more ribosomes than the needed amount.
Because of the components involved in the synthesis (2.28), the feedback should concern at least one between *R-Protein* and *rRNA*.
One of the ribosome feedback model ([8]) proposed in the literature states that the cell is able to prevent the accumulation by using a feedback regulation on *rRNA*. In particular, this is an auto-regulatory process through a negative feedback loop ([9]). Nevertheless, other authors ([11]) have asserted that it may exist a mechanism that regulates the parallel production of *R-Proteins* and *rRNAs*. This should guarantee to have similar concentrations for both the two reactants, since the lacking amount of one component plays the role of limiting factor in the ribosomes synthesis, as it can be easily observed in 2.28. In other words, it does not make sense to produce a huge amount of $rRNA$ if the *R-Proteins* are missing and vice versa.
Furthermore, in Nomura et al. ([10]), it has been demonstrated that certain *R-Proteins* work as inhibitors of protein synthesis from their own *mRNAs*.

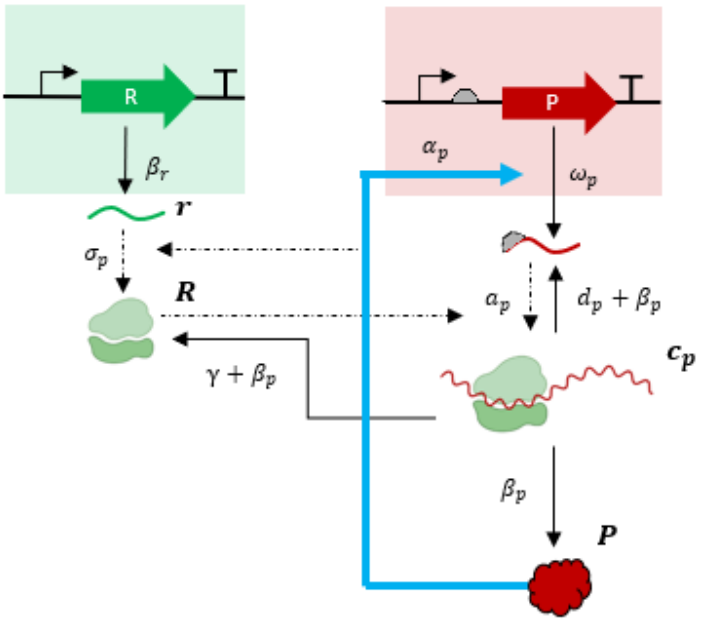

Figure 2.5: Sketch of the inhibitory feedback

A rough scheme of this feedback mechanism is shown in Figure 2.5. As already specified

in the previous paragraph, for sake of simplicity all the different species of *R-Proteins* are gathered together in an unique species P, same for the *rRNAs* which have been grouped together into the generic component r.
Moreover, the regulation of the ribosomes synthesis can be seen as the result of the competition between *rRNA* and *mRNA*. Indeed the set of reactions that regulates the inhibition is the following:

$$\mathrm{P} + \mathrm{m}_p \xrightarrow{\alpha_p} \mathrm{P} : \mathrm{m_p} \tag{2.29}$$

$$\mathrm{P} : \mathrm{m_p} \xrightarrow{\gamma} \mathrm{P} \tag{2.30}$$

$$\mathrm{P} \xrightarrow{\mu} 0 \tag{2.31}$$

In particular, the reaction 2.29 must compete against 2.28.
However it is reasonable to expect that the affinity of *R-Protein* with *rRNA* is higher ([11]) than with *mRNA*, otherwise the inhibition would be stronger than the synthesis, which is a contradiction.

In this first part of the chapter, the bases for the modeling process have been analyzed. Particular emphasis has been attributed to the biological scenario, which must be defined in advance, in order to better understand the following steps.

2.2 Modeling Growth, Load and their relation

In this second part, the problem of defining the cell growth will be introduced.
In particular, the importance of the cell growth will be discussed as first. Afterwards, it will be proved that it is ruled by a parameter, called *Growth Rate*. However, this parameter will not be considered constant. Consequently, the last part will concern the literature related to the growth rate function.

2.2.1 Importance of the Growth

The bacterium *Escherichia Coli*[6] reproduces by the cell division and this process is called *Binary Fission*.
This starts from the replication of the unique circular *DNA*, which is the responsible for the genetic pool of the cell. Because of that, the replication of *DNA* is a fundamental process since it must ensure that the new cell will contain all the proteins needed for the life. Even though the whole process of duplication of the cell will not be debated in this thesis, it is interesting to analyze the needed time and the conditions of the cell growth, other than the effects on the system.
For example, it is known that the binary fission process of cell of *Escherichia Coli*, at $37°C$, takes 40 minutes but it can be reduced to 20 minutes ([1],[7]) in particular conditions of nutrients. The needed time for the duplication of a single cell is called *Doubling Time* τ.
The binary fission has an important influence on the cell system. Indeed, it is worth to recall that this process implies that the genetic material of the cell is divided into two equal parts, one remains at the mother cell, while the other is transferred to the new born. Consequently, the loss of some genetic material due to the duplication must be taken into

[6]The cell of *Escherichia Coli* contains just one circular *DNA*.

account in the modeling. Since it concerns a single cell, and not the entire population, the components can be assumed as simply consumed[7]. Therefore, this can be assumed to be equal to a degradation reaction (Paragraph 2.1.3), namely:

$$\mathrm{A} \xrightarrow{\lambda} 0 \tag{2.32}$$

which indicates that the generic component A of the cell is simply dissipated. The rate λ, which is called *Growth Rate*[8], indicates the velocity of the growth and it is related to time t needed for the growth, and in particular to the parameter τ, as it will show in the following.
The growth phase of a cellular population is ruled by the equation:

$$x(t) = x_0 e^{t\lambda} \tag{2.33}$$

where x is the number of cells in the temporal instant t and x_0 is the initial condition of the population. By the definition of doubling time, equation 2.33 can be rewritten as:

$$2x_0 = x_0 e^{\tau\lambda} \tag{2.34}$$

and consequently:

$$\lambda = \frac{\ln 2}{\tau} \tag{2.35}$$

Even though 2.35 has been derived from an equation that expresses the dynamics of a population of cell, it can be easily adapted to the single cell case. Indeed, all the elements in the cell grow linearly with the cell itself, so it is sufficient to consider x as amount of a certain component instead of as number of cells.
Finally, it is important to not confuse the growth rate λ with the spontaneous decay, even though they are represented by the degradation reaction 2.26, where ψ_A can be either the growth rate or the spontaneous decay rate. Indeed, while the first is related to the growth of the cell, the second descends from the assumption that all the components can not last forever, but they are subject to a natural decline. Consequently, the pool of chemical reactions will contain two different degradation reaction for each component. Thus, they can be summed up in a unique reaction:

$$\mathrm{A} \xrightarrow{\mu_\mathrm{A}} 0 \tag{2.36}$$

where μ_A is called *Degradation Rate* and it is the summation between the growth and the spontaneous decay rate.

In most of the experiments and articles from the literature, the growth rate has been considered constant, namely all the calculations have been made based on the assumption that the cell (or the population) is always in its maximal growth condition.
Instead, the aim of this dissertation is to perturb the system using a metabolic load, in order to appreciate the effect on the growth. Because of that, the growth rate can not be assumed constant anymore, but it will be studied as function that depends on certain components of the cell.

[7] The fact that they will be transferred to a new born cell is not considered, since the system is a single cell.

[8] This parameter is here applied to the single cell model, but it is used also for expressing the velocity of growth of a cellular population.

2.2.2 Literature on Modeling

In the previous paragraph, the importance of the growth rate has been debated. As already stated, this parameter is not constant, so it is fundamental to understand which components (of the cell) the growth rate depends on.
There are several articles and contributions from the literature which assert that the growth function depends on the ribosomal content. However, as discussed in Paragraph 2.1.3, the ribosomal compartment is wide and it includes different species. Nevertheless, some contradictions or inaccuracies compared to the biological evidence can be found, especially because most of the authors have focused their own research on cells without considering the influence of a metabolic burden.

In *Allen G.Marr* [1], the proposed model for the cell growth is a linear function dependent (Figure 2.6) on the total number of ribosomes R^{TOT} [9].

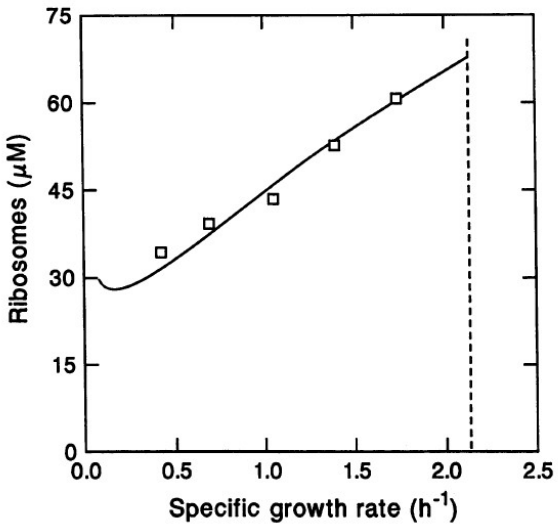

Figure 2.6: The graphs shows the concentration of the ribosomes as function of the specific growth rate. The symbols □ are values computed from a specific data set, while the solid line is the analytical solution of a given equation [1] .

In this work, the metabolic load has not been considered. The critical aspect of this article comes from the assumption that the growth rate depends on R^{TOT}. In fact, this implies that the metabolic burden will not impact the growth rate of the cell. In other words, the addition of the load will simply re-arrange the inner proportions of the compartment R^{TOT}. Hence, it should be possible to observe the same growth in both the two configurations (with and without load).
However, this is not consistent with the biological evidences. Even though some biological experiments show that adding and/or increasing the concentration of the load does not lead to an instant alteration in the growth rate, it is not reasonable to expect that the rate will be maintained invariant compared to the unloaded case.

[9] The total number of ribosomes of the cell is the summation between free ribosomes R and the ribosomal complexes c_i, which are obtained by the bond of $mRNA$ with the ribosome, where i stands for the species.

In *Chenhao Wu et al.* [3], the linearity of the growth rate has been maintained (Figure 2.7), but it now depends on the difference between total active (N_R) and inactive number (N_R^{inact}) of ribosomes and on the total protein mass (M_p):

$$\lambda M_p = \epsilon \cdot (N_R - N_R^{inact}) \tag{2.37}$$

Again, the metabolic load has not been considered. Despite the proposed formulation does not clearly show critical points, it is not possible to draw any conclusion. Indeed, even though the addition of a metabolic load will influence both N_R and N_R^{inact} (preferably increasing N_R and decreasing N_R^{inact}) and consequently modify the growth rate, its new value can not be predicted, since the formulation 2.37 does not consider the eventual addition of the burden.

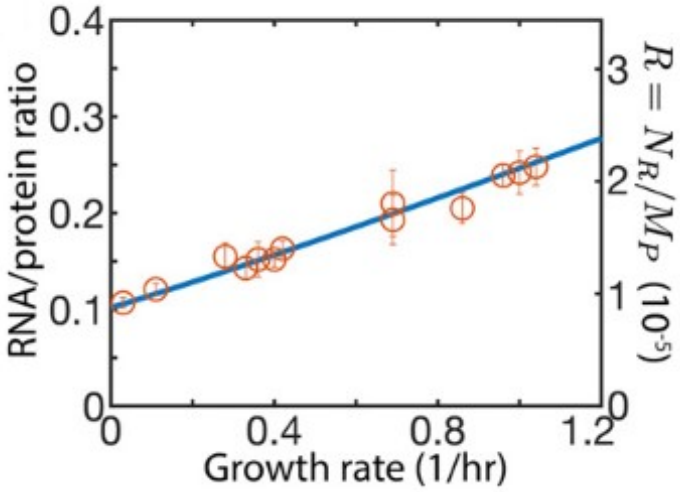

Figure 2.7: Approximate linear relation between the growth rate, the RNA/protein ratio and active ribosomes/protein mass ratio [3]

Differently, in *R.Levin et al.* [4], the proposed formulation of the growth rate is a first order Hill function dependent on R^{TOT}:

$$\lambda = (\lambda_{MAX} - \lambda_{MIN}) \frac{R^{TOT}}{(R^{TOT} + K)} + \lambda_{MIN} \tag{2.38}$$

where λ_{MAX} is the maximum growth rate, λ_{MIN} is the death rate and K is the Hill (shape) parameter. The existence of minimum and maximum values for the growth suggests that it should exist the equivalent for R^{TOT}. In other words, this model states that the cell can not grow if $R^{TOT} < R_{min}^{TOT}$. Furthermore, it has been assumed that there exists also a threshold for the maximum amount of ribosomal content that can be produced by the cell, namely R_{max}^{TOT}. Again, this model seems to have the same issue as the first discussed one ([1]), namely the dependence on R^{TOT} is not suitable.

In *Atkinson et al.* [5], the proposal is a linear growth rate that depends on the ribosome complexes ribosomes bound with $mRNA$):

$$\lambda = \frac{\gamma(a)}{M} \sum_x c_x \tag{2.39}$$

where $\gamma(a)$ is the rate of translation of all the proteins dependent on the the available energy and M is the total proteome content of the cell.

In *Del Vecchio and Murray* [6], the growth rate function is not clearly defined. Instead, the authors have provided the ratios of free ribosomes, which is 30% over the total number of ribosomes $R^{TOT} \approx 34\mu M$ (in the exponential phase). Accordingly to this information, the growth rate has been analytically drawn by fitting a Hill function on the points obtained from the available information.

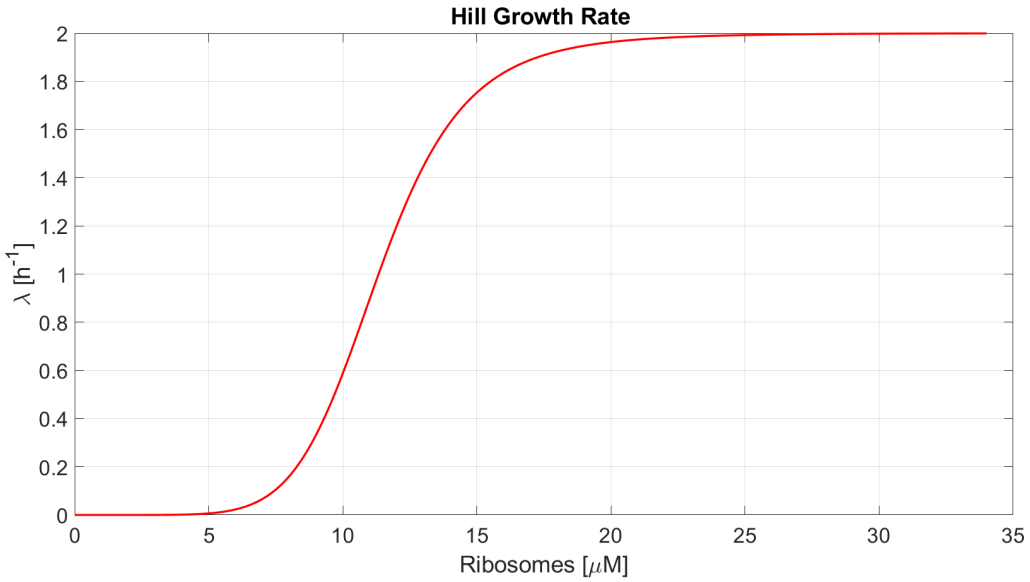

Figure 2.8: Derived growth rate function [6]

In this second part of the chapter, the importance of modeling the cell growth has been highlighted. In particular, several articles from the literature have been roughly analyzed, in order to understand which growth rate functions have been already proposed.
In the following, different models of the cell system will be designed. Their inner dynamics, and in particular their growth, will be analyzed, by using different growth rate functions; some of them will be derived from the exposed articles, others will be inspired by the biological evidences and the articles.

Chapter 3

Mathematical Modeling and Analysis

In this chapter, the analysis of the cell growth will be discussed, considering also the load effect on the system. In particular, various growth rate functions will be tested with three models (which describe the cell system), with different levels of complexity.
First, an overview of all the models and the growth functions will be presented; then every model will be analyzed individually.

3.1 General Hypotheses and Assumptions

In the following, three models with increasing level of complexity will be presented.
Starting from the structure of simplest model, some chemical reactions have been added, in order to obtain a model the more comparable as possible to the real biological system.

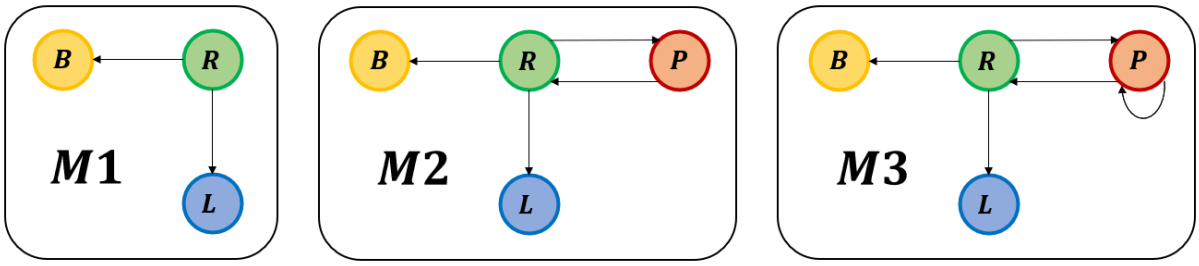

Figure 3.1: Comparison between the rough structures of the three models

The first model *M1* considers just three species, namely the basal B, the ribosomal R and the load L species.
The basal species is referred to the components and the products (*mRNAs*, complexes and proteins) that come from the biochemical reactions that are fundamental for the survival of the bacterial cell. The R compartment simply consists on the free ribosomes. Finally, the load species groups together the *mRNA*, the complex and the protein that are involved in the chemical reaction of the metabolic burden.
In model *M2*, the complexity is increased by adding the *R-Protein* P species. This compound gathers together the components (*mRNAs* and the complexes), other than the final products (proteins) that contribute to the synthesis of *R-Proteins*, which are needed for the synthesis of the ribosomes.

Finally, the third model *M3* presents the same species of *M2*, but it is assumed that there exists a loop that involves the compound P. In the following, this will be identify as a negative feedback mechanism.

The three models here presented share some general hypotheses and assumptions.
As it has been already discussed in Paragraph 2.1.1, the modeling process starts from the biochemical reactions that occur in the bacterial cell system. However, these reactions involve many reactants, products and parameters. Thus, the following hypothesis and assumptions aim to simplify the models.

Assumption A: It is worth to remind that the transcription process in the protein synthesis involves also RNA polymerase (see Paragraph 2.1.3). For every model, the hypothesis of $RNAP$ constant has been assumed (see Appendix B for more details) in order to simplify the transcription reactions (and consequently reduce the number of differential equations).

Hypothesis B: As already discussed in Paragraph 2.2.1, all the components in the cell are subject to spontaneous decay, which is not the same for everyone. However, since it has the same order of magnitude, the spontaneous decay will be considered equal for proteins, ribosomes and ribosomal complexes (regardless of the species), in order to reduce the number of parameters in the models. It will be indicated with δ.
The same reasoning can be applied on $mRNA$ and $rRNA$ (again independently of the species): their spontaneous decay will be considered equals and indicated with γ.

Assumption C: The last assumption regards the degradation rate of the $mRNA$ and $rRNA$. As declared in Paragraph 2.2.1, the degradation rate is defined as summation between the spontaneous decay and the growth rate. From the literature, it is known that the spontaneous decay γ is faster than the growth rate λ of the cell. This leads to the following approximation:

$$\mu_{mRNA,rRNA} = \gamma + \lambda \approx \gamma \tag{3.1}$$

which proves that the degradation of a $mRNA$ and $rRNA$ molecule just depends on its decay and not on the growth rate of the cell and consequently it can be assumed constant over time.

For each model, different growth rate functions have been tested.

Linear Growth Rate: At first, a linear growth rate function has been proposed. It has been obtained from the fitting of the data collected in [1] (these data have been collected in a single cell without the addition of a metabolic load to be synthesized). In the cited paper, the hypothesized independent variable is the total amount of ribosomes in the cell R^{TOT}. Even though this choice does not show issues in the unperturbed case, it would lead to some incoherent results if the metabolic load would be added.
Indeed, accordingly to Nikolados et al. [2], the burden impacts the system and in particular it must cause a decrease in the growth rate. Nevertheless, the addition

of metabolic load would not influence the total number of ribosomes at all if R^{TOT} is set as independent variable.[1].
Despite the discussed issue, the data have been used anyway due to unavailability of other sets of data but the independent variable has been changed.
Two different formulations of growth rate function have been proposed.

Linear c_b: The first formulation considers c_b, namely the complex of the basal species, as independent variable. This choice has been made considering the meaning of the term *basal*. Indeed it is referred to all the biochemical reactions that are fundamental for the life of the cell, so it seems reasonable to get the growth rate depend on it.

Linear R^{active}: The second alternative sets R^{active}, namely the summation of all the complexes [2], as independent variable. This has been formulated starting from the evidences showed in Atkinson et al. [5].

In both cases, the new independent variable is assumed to replace R^{TOT}.

Hill Growth Rate: Alternatively, a Hill growth rate function has been considered. As in the linear case, two formulations have been examined. Both depend on c_b.

Hill$_{DV}$ c_b: This growth rate functions has been formulated starting from the assumptions presented in Del Vecchio and Murray [6] (as in most of the articles, they did not consider an eventual addition of load). The authors asserted that, in the maximal growth conditions, the total amount of ribosomes is fixed and it is shared between the free ribosomes and basal species. Since a clear formulation of the function has not been reported, an approximated version has been derived. It has been found out that only the Hill function respects the given indications.

Hill c_b: For sake of completeness, a further Hill function has been proposed. It has been obtained from the fitting of the data in [1].

The comparison between all the discussed growth rate functions is reported in Figure 3.2. In the panel A the fitted linear growth rate function is reported, while the panel B shows the differences between the two Hill formulations. In particular, it has been highlighted the value 34 μM in the x-axis and it represents the maximal amount of ribosomes in the cell hypothesized by Del Vecchio and Murray [6].

It is not worth to analyze every growth rate function with all the models.
Indeed, the final aim is to identify the best growth rate function that produces coherent results, independently of the complexity of the model. According to that, it does not

[1] The total number of the ribosomes is the summation between the free ribosomes R and the other complexes c_i (i.e., in *M1*: $R^{TOT} = R + c_b + c_\ell$). Remind that the complex c_i is obtained by the bound of *mRNA* and R. Before adding the load, $c_\ell = 0$. Afterwards, in order to be synthesized, the burden requires a certain amount of free ribosomes. This can be obtained from the compound of the free ribosomes (if it is not empty) or it can be stolen from the other complexes. This means that, while c_ℓ is increasing, the other components are decreasing. In conclusion, this proves that the addition of the load recombines the amount of the individual components of R^{TOT}, but not R^{TOT} itself.

[2] In *M1*: $R^{active} = c_b + c_\ell$; in *M2* and *M3*: $R^{active} = c_b + c_p + c_\ell$.

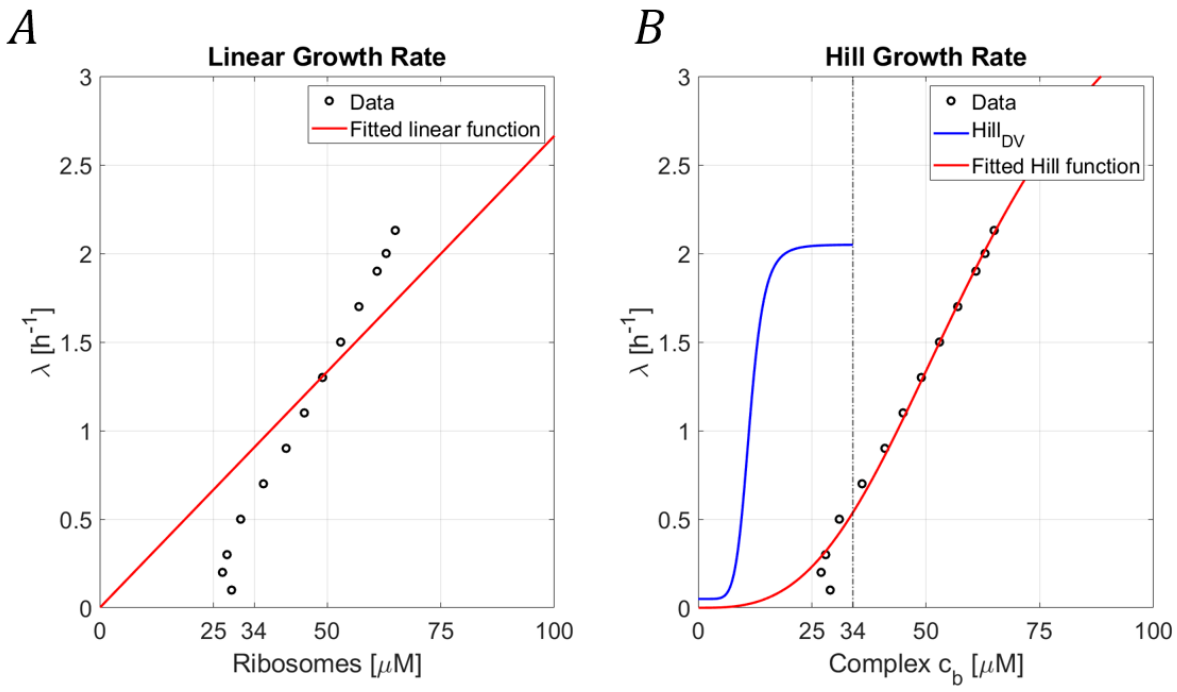

Figure 3.2: Comparison between the growth rate functions

make sense to test a function with more complex structures if it has already led to some issues.
First, every growth rate functions will be tested with the first model *M1*. Successively, the functions that have produced suitable results will be analyzed with the second model *M2* (for sake of completeness, also one of the functions that has presented some issues will be tested as well, in order to confirm its inadequacy). Lastly, the most complex model *M3* will be examined considering just one growth rate function, which is resulted to be the best one so far.
The performed tests are resumed in the Table 3.1.

Growth Rate Function λ	**Variable**	**Model 1**	**Model 2**	**Model 3**
Linear	c_b	**X**		
Linear	R_{active}	**X**	**X**	**X**
Hill_{DV}	c_b	**X**	**X**	
Hill	c_b	**X**		

Table 3.1: Performed tests

3.2 Model $M1$

The first proposed model is shown in Figure 3.3.
The structure involves three different types of ribosomal species: the free ribosomes R, the basal c_b and c_ℓ complexes. Together, they form the total amount of ribosomes of the cell R^{TOT}.

Starting from the *DNA*, the transcription produces the *mRNA* with *RBS* (*Ribosome Binding Site*). This can bind the ribosome with rate a_i (where i indicates the species, namely $i = \{b, \ell\}$). However, this link can be broken with dissociation rate d_i.
In the building phase of the model, an hypothesis has been made: c_i can release the free ribosome with rate γ.
Furthermore, the translation process creates the protein with rate β_i from the ribosomal complex c_i, but also returns the *mRNA* and the free ribosome used for the bound (this justifies the arrow with rate $d_i + \beta_i$ from the complex to the *mRNA* and the one with rate $\gamma + \beta_i$ from the complex to the ribosome).
In Figure 3.3, also the degradations (γ for *mRNA* and μ for ribosomes, ribosomal complexes and protein) are considered.
It is worth to specify that an approximation has been made. The process of formation of the ribosomes R should involve both *rRNA* and *R-Protein*, while in this case it is similar to a transcriptional process, namely R has been derived directly from the *DNA* with rate β_r.

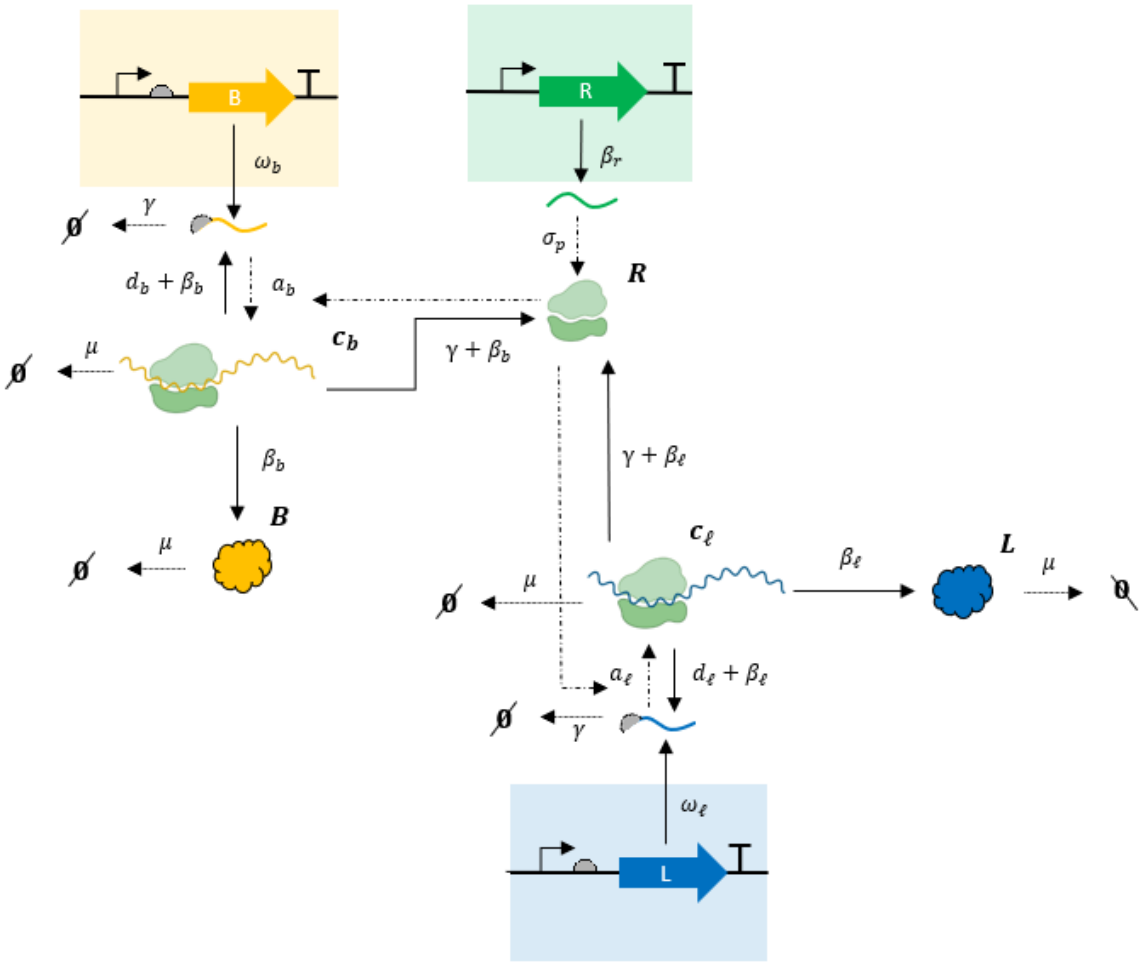

Figure 3.3: First Model. The yellow and blue expression cassettes represent the basal and load species respectively. The green cassette stands for the free ribosomes.

Thus, this model is characterized by the following set of chemical reactions:

$$
\begin{aligned}
&\mathrm{D}_b \xrightarrow{\omega_b} \mathrm{D}_b + \mathrm{m}_b \\
&\mathrm{D}_\ell \xrightarrow{\omega_\ell} \mathrm{D}_\ell + \mathrm{m}_\ell \\
&\mathrm{D}_r \xrightarrow{\beta_r} \mathrm{D}_r + \mathrm{R} \\
&\mathrm{m}_b + \mathrm{R} \underset{\mathrm{d}_b}{\overset{\mathrm{a}_b}{\rightleftharpoons}} \mathrm{c}_b \xrightarrow{\beta_b} \mathrm{m}_b + \mathrm{R} + \mathrm{B} \\
&\mathrm{m}_\ell + \mathrm{R} \underset{\mathrm{d}_\ell}{\overset{\mathrm{a}_\ell}{\rightleftharpoons}} \mathrm{c}_\ell \xrightarrow{\beta_\ell} \mathrm{m}_\ell + \mathrm{R} + \mathrm{L} \\
&\mathrm{m}_b, \mathrm{m}_\ell \xrightarrow{\gamma} 0 \\
&\mathrm{c}_b, \mathrm{c}_l \xrightarrow{\gamma} \mathrm{R} \\
&\mathrm{R}, \mathrm{c}_b, \mathrm{c}_\ell, \mathrm{B}, \mathrm{L} \xrightarrow{\mu} 0
\end{aligned}
$$

and mass conservation laws:

$$
\begin{aligned}
R^{TOT} &= R + c_b + c_\ell && (3.2)\\
m_b^{TOT} &= m_b + c_b && (3.3)\\
m_\ell^{TOT} &= m_\ell + c_\ell && (3.4)\\
& && (3.5)
\end{aligned}
$$

As already explained, the reactions can be translated into a system of differential equations, as it follows:

$$
\begin{aligned}
\dot{m}_b &= \omega_b D_b - a_b m_b R + d_b c_b + \beta_b c_b - \gamma m_b && (3.6)\\
\dot{m}_\ell &= \omega_\ell D_\ell - a_\ell m_\ell R + d_\ell c_\ell + \beta_\ell c_\ell - \gamma m_\ell && (3.7)\\
\dot{c}_b &= a_b m_b R - d_b c_b - \beta_b c_b - \gamma c_b - \mu c_b && (3.8)\\
\dot{c}_\ell &= a_\ell m_\ell R - d_\ell c_\ell - \beta_\ell c_\ell - \gamma c_\ell - \mu c_\ell && (3.9)\\
\dot{R} &= \beta_r D_r - \sum_j a_j m_j R + \sum_j d_j c_j + \sum_j \beta_j c_j + \gamma \sum_j c_j - \mu R \\
&\quad \text{with } j = b, \ell && (3.10)\\
\dot{B} &= \beta_b c_b - \mu B && (3.11)\\
\dot{L} &= \beta_\ell c_\ell - \mu L && (3.12)\\
\dot{R}^{TOT} &= \beta_r D_r - \mu R^{TOT} && (3.13)\\
\dot{m}_b^{TOT} &= \omega_b D_b - \gamma m_b^{TOT} - \mu c_b && (3.14)\\
\dot{m}_\ell^{TOT} &= \omega_\ell D_\ell - \gamma m_\ell^{TOT} - \mu c_\ell && (3.15)
\end{aligned}
$$

A further approximation can be made. Considering the relation between γ and μ ($\gamma >> \mu$) and knowing that $m_i^{TOT} \geq c_i$ (easily derived from equations 3.3 and 3.4), then equations 3.14 and 3.15 become:

$$
\begin{aligned}
\dot{m}_b^{TOT} &\approx \omega_b D_b - \gamma m_b^{TOT} && (3.16)\\
\dot{m}_\ell^{TOT} &\approx \omega_\ell D_\ell - \gamma m_\ell^{TOT} && (3.17)
\end{aligned}
$$

In the following analysis, the approximated version will be considered.
The set of differential equation can be studied in *Steady State* condition. In particular,

while equations 3.16 and 3.17 have a closed form solution, namely:

$$m_b^{TOT} = \frac{\omega_b D_b}{\gamma} \tag{3.18}$$

$$m_\ell^{TOT} = \frac{\omega_\ell D_\ell}{\gamma} \tag{3.19}$$

the others must be solved numerically. The numerical solutions have been computed using the method *fsolve* from *Matlab*.

3.2.1 Parameters' Value based on Maximal Growth Conditions

Before to proceed with the analysis of the system, some parameters must be estimated, since their value is not available in the literature.
As an interesting starting point one could examine the conditions that lead to the maximal growth of the cell.
In order to establish those conditions, the model without the metabolic load must be considered.
However, since the literature does not provide an unique formulation for the maximal growth, the estimations depend on the chosen assumptions. In this case, the point of view of Del Vecchio and Murray [6] has been pursued.
As claimed by them, the partition of the ribosomes is the following:

$$c_b = \frac{2}{3} R_{max}^{TOT} \tag{3.20}$$

$$R = \frac{1}{3} R_{max}^{TOT} \tag{3.21}$$

where $R_{max}^{TOT} \approx 34\mu M$ is the total number of ribosomes in the maximal growth condition and $\mu_{max} = \lambda_{max} + \delta \approx 2.05 h^{-1}$.
Consider now the differential equations of R^{TOT} (3.13) and m_b^{TOT} (3.14) at the steady state.
From 3.13, the value of $\beta_r D_r$ can be derived simply by the substitution of R^{TOT} and μ with R_{max}^{TOT} and μ_{max} respectively, namely their value at the maximal growth:

$$\beta_r D_r = \mu_{max} R_{max}^{TOT} \tag{3.22}$$

To obtain $\omega_b D_b$, the equation of the manifold related to the basal species (see Appendix C for the derivation of the manifold) must be combined with 3.14:

$$\begin{cases} c_b = \frac{R}{R+K_b} m_b^{TOT} \\ 0 = \omega_b D_b - \gamma m_b^{TOT} - \mu c_b \end{cases} \tag{3.23}$$

Replacing c_b and R with 3.20 and 3.21 respectively and substituting m_b^{TOT} in the second equation, the value of $\omega_b D_b$ in the maximal growth condition has been obtained:

$$\omega_b D_b = \frac{2}{3} R_{max}^{TOT} (\gamma - \mu_{max}) + \frac{2}{3} \gamma K_b \approx \frac{2}{3} R_{max}^{TOT} (\gamma + K_b) \tag{3.24}$$

where the approximation is possible since $\gamma >> \mu_{max}$.
Under the assumption that K_b is very small [3] a further approximation is possible, namely:

$$\omega_b D_b \approx \frac{2}{3} R_{max}^{TOT}(\gamma + K_b) \approx \frac{2}{3} R_{max}^{TOT}\gamma \tag{3.25}$$

3.2.2 Analysis of R^{TOT} variation

Ahead of studying the load effect, a prior analysis on the distribution of the total amount of ribosomes R^{TOT} among the different ribosomal species (R, c_b and c_ℓ) has been performed.
The concentration of R^{TOT} is limited in the cell. Because of that, the species must compete for obtaining the needed ribosomes for their reactions. Thus, the aim of this study is to understand which parameters regulate the competition for the ribosomes between different species.
The dynamics of the components is described by equations 3.8, 3.9 and 3.10 and their value at the *Steady State* can be computed simply setting the differential equations equal to zero. These do not have a closed form solution and then they must be solve numerically. However, under the assumption that the reactions of association and dissociation of complexes are fast processes (compared for example to the final synthesis of the proteins), a further analysis has been proposed, namely the study of the *Rapid Equilibrium* (more details in Paragraph 2.1.2).
According to that, it is possible to derive the manifolds of c_b and c_ℓ (as explained in Appendix C):

$$c_b = \frac{R}{R + K_b} m_b^{TOT} \qquad c_\ell = \frac{R}{R + K_\ell} m_\ell^{TOT} \tag{3.26}$$

while the value of R can be computed by using the mass conservation laws of R^{TOT} and the manifolds:

$$R^{TOT} = R + c_b + c_\ell \qquad \longrightarrow \qquad R^{TOT} = R + \frac{R}{R + K_b} m_b^{TOT} + \frac{R}{R + K_\ell} m_\ell^{TOT} \tag{3.27}$$

As it can be easily noticed, this formulation leads to results which are not affected by the chosen growth rate function. Instead, it is just subject to the total amount of *mRNA* content (m_i^{TOT}) and the strength of the ribosome-*mRNA* bound, which corresponds to the dissociation constant $K_i = d_i/a_i$.
Equations 3.26 and 3.27 must be solved numerically [4]. A closed form solution of these equations (which are Hill functions) is possible only if the manifolds are approximated with a piecewise linear function (further details in Appendix D):

$$c_i = \frac{R}{R + K_i} m_i^{TOT} \qquad \longrightarrow \qquad c_i \approx \begin{cases} \frac{m_i^{TOT}}{K_i} R & , R \leq K_i \\ m_i^{TOT} & , R > K_i \end{cases} \tag{3.28}$$

[3] Remind that K_b is defined as the ratio between a_b and d_b. This assumption means that the association rate between the ribosome and the *mRNA* is stronger than the dissociation.
This is reasonable because the basal species is committed to the fundamental processes of the cell, namely the proteins produce from the basal complex are necessary for the cell's life. In this term, assuming a strong bound between the ribosome and the basal *mRNA* sounds plausible.
[4] Remember that R^{TOT} is the independent variable in this analysis.

In Figure 3.4, the evolution of different components is plotted as function of R^{TOT}. The continuous lines represent the real evolution, obtained from the numerical solutions of the equations 3.26 and 3.27. The dashed lines indicate their approximation, based on the general competition analysis (Appendix D).

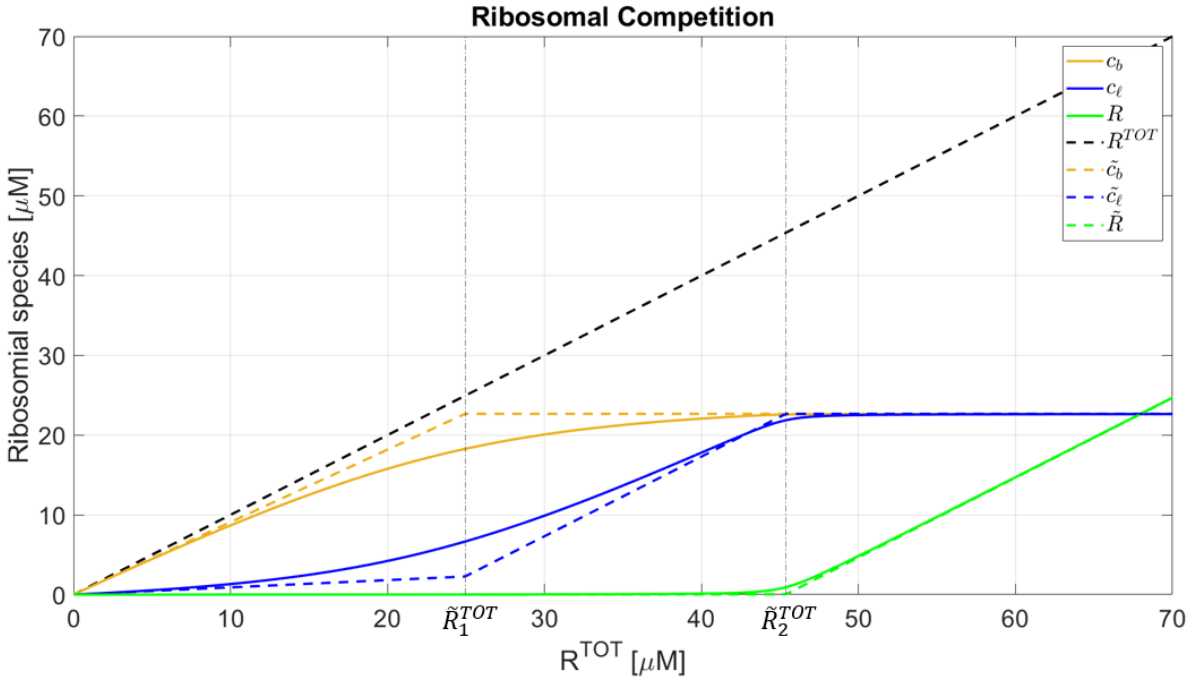

Figure 3.4: Competition between ribosomal species with $K_b << K_\ell$ and $m_b^{TOT} = m_\ell^{TOT}$. The yellow, blue and green dashed lines represent the approximate evolution of c_b, c_ℓ and R respectively. $\tilde{R}_1^{TOT}$ and $\tilde{R}_2^{TOT}$ are the breaking points.

The approximate functions is used to predict the behavior of the ribosomal components. Indeed, the study of the breaking points gives information regarding the rise of the complexes and, in particular, when they start to approach their maximal value. From the approximate system in 3.28, it is straightforward that the first complex to reach its maximum has the smallest K. Thus the order of rising is just regulated by the dissociation constants. This is coherent also from a biological point of view. At the beginning, the amount of available ribosomes is limited. When the species start to compete, only the one with the strongest association rate (which implies the smallest K) can rise, while the others do not have enough strength to bind the needed ribosomes. Furthermore, they are not even able to steal the ribosomes from the already synthesized complex. Thus, when the complex has reached its maximal value, which means that it does not need other ribosomes, the species with the second smallest dissociation constant can start to grow. Moreover, the computations in Appendix D shows that the breaking points $\tilde{R}_1^{TOT}$ and $\tilde{R}_2^{TOT}$ depend from both the dissociation constants and the total amount of $mRNA$. This is reasonable because, as already explained, one species can start to grow when the previous (i.e., species with a lower dissociation constant) has reached its maximal value [5], which is m^{TOT}.

[5] This statement can be easily proved computing the value of 3.28 for $R \longrightarrow \infty$.

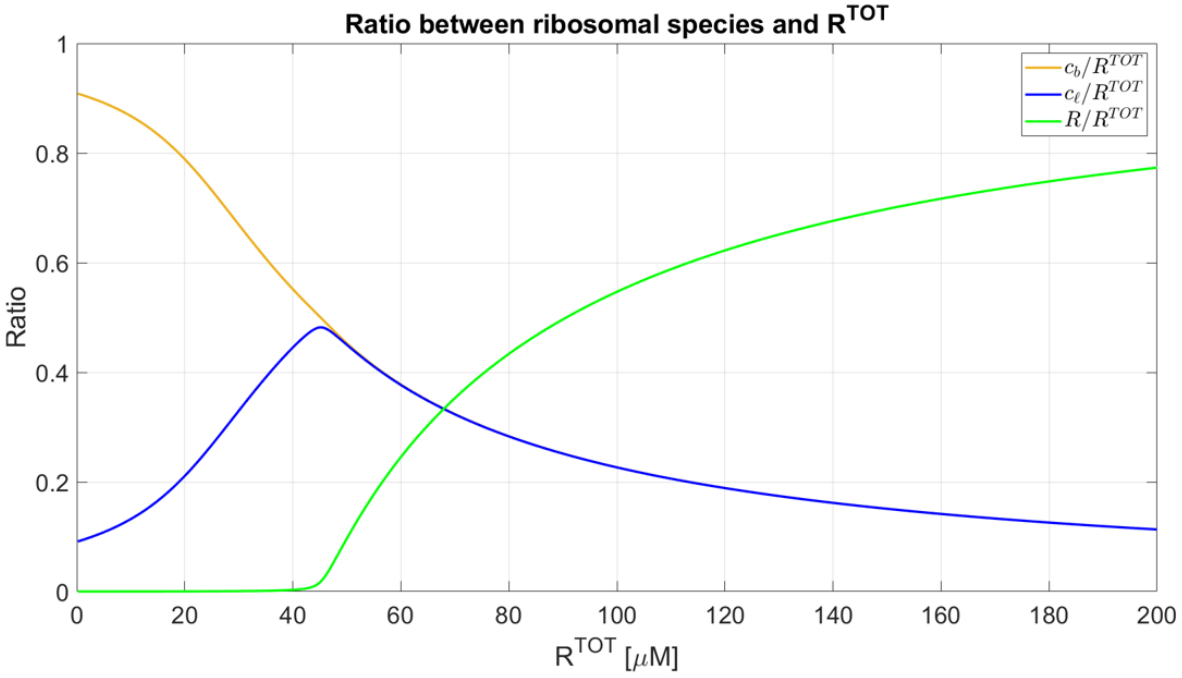

Figure 3.5: Ratio between ribosomal species and R^{TOT}

Furthermore, the ratios between the ribosomal complexes and the total amount of ribosomes are shown in Figure 3.5. This clearly shows that only one species can grow at a time.
To sum up, it has been proved that the competition between the ribosomal species is regulated by the dissociation constants and the total amount of *mRNAs*.

3.2.3 Analysis of Load effect

Afterwards, it is interesting to extend the analysis in order to better understand the load effects on the system [6], in particular on the cell growth, starting from the dynamics described by equations from 3.6 to 3.15.

In this study, the equilibria have been studied as function of the burden. This is characterized by three parameters, which are: the dissociation constant K_ℓ, the transcription rate ω_ℓ and the translation rate β_ℓ. From a biochemical point of view, the easiest to adjust is ω_ℓ [7]. Thus, the assumed choice is to vary the transcriptional rate, while the other parameters are kept invariant.
In particular, the evolution of c_b, c_ℓ, R and R^{TOT} have been considered. Despite the *mRNAs* and the proteins are changing as well, they do not influence the growth of the cell directly (indeed, considering all the cited growth rate functions, they depend uniquely on the ribosomal components). Due to that, the study of the evolution of these elements is not fundamental for the aim of the analysis.
Moreover, differently from the previous case, now the degradation μ is involved in the system. In fact, even though the evolution of the complexes can be still derived from the

[6] Further details regarding the *Matlab* implementation are available in Appendix G
[7] The transcription rate can be regulated by modifying the promoter of the load.

manifolds equations 3.26, the differential equation of R^{TOT} 3.2 at the equilibrium is:

$$R^{TOT} = \frac{\beta_r D_r}{\mu} \tag{3.29}$$

This means that the results are strongly influenced by the chosen type of growth rate.

Following the order in Table 3.1, all the growth rate functions have been tested and the simulations are collected together in Figure 3.6.
The first growth rate that has been analyzed is the *Linear* function dependent on c_b, shown in Panel *A*. Even though the obtained growth rate seems to respect the biological evidences (i.e., the increase of the load should not immediately lead to a reduction of the growth rate), some issues can be found in the graph of the evolution of the equilibrium points. Indeed, the augmentation of the transcription rate of the burden can not cause an increment in the total amount of the ribosomes R^{TOT}. This fact is also inconsistent with the plotted growth rate: the drop of the growth can not justify the rise of the ribosomes.
Afterwards, the *Linear* growth function dependent on R^{active} has been considered. The result is shown in Panel *B*. Differently from the previous case, now R^{TOT} is initially decreasing while μ is rising. After a starting transitory, the growth rate remains constant. This is coherent with the behavior of the ribosomal complexes, since $R^{active} = c_b + c_\ell$.
Later, some other tests have been executed. Instead of a linear, a Hill growth rate function has been examined.
First, the Hill formulation from Del Vecchio [6] has been considered ($Hill_{DV}$) and shown in Panel *C*. Here, the same issue as with linear function dependent on c_b can be observed. Again, while μ is decreasing, R^{TOT} is rising.
Additionally, it is worth to notice that the growth rate has a starting value of $\approx 2h^{-1}$, while in all the other cases it starts from lower values. The explanation comes from the computation of the maximal growth conditions. Indeed, as specified in Paragraph 3.2.1, the values of $c_{b,max}$, R_{max} and R^{TOT}_{max} have been derived from [6] in absence of the load ($\omega_\ell = 0$), and consequently they ensure the maximal growth rate only with this type of growth rate and only at the beginning of the analysis, namely when the burden has not been applied yet.
Lastly, the *Hill* growth rate obtained from the fitting of data (from [1]) has been examined (Panel *D*). Since the only difference with the previous proposal is the parameters that characterize the Hill, the results are similar, so that the observed issues.

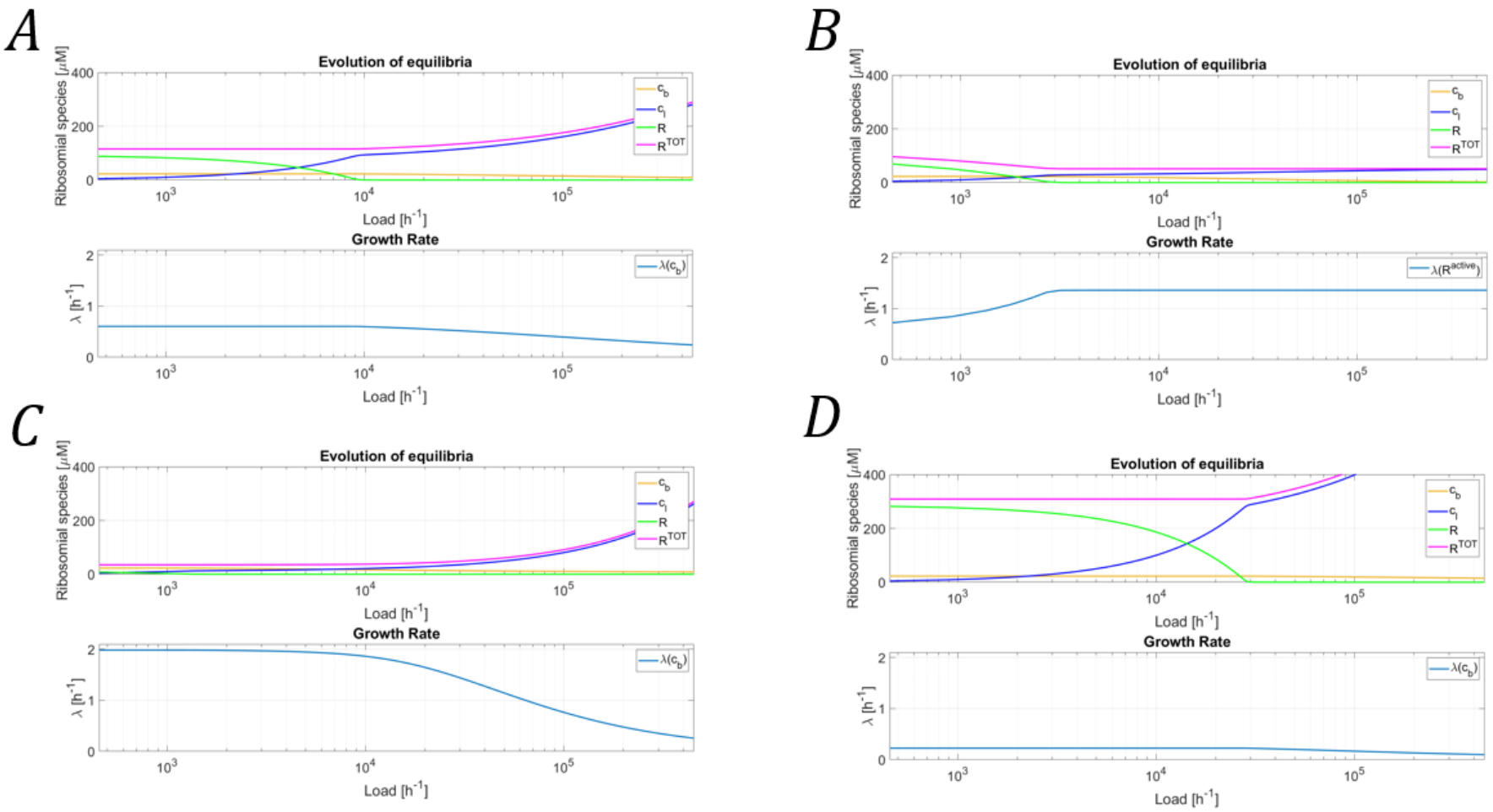

Figure 3.6: Comparison between simulations with different growth rate functions. *A* used the *Linear* function dependent on c_b. *B* used the *Linear* function dependent on R^{active}. *C* used the $Hill_{DV}$ function dependent on c_b. *D* used the *Hill* function dependent on c_b.

3.3 Model *M2*

An extension of the first model has been formulated and it is shown in Figure 3.7.
The main difference with *M1* regards the process of formation of the free ribosomes R. While in the previous formulation it was approximated by the reaction $\mathrm{D}_r \xrightarrow{\beta_r} \mathrm{D}_r + \mathrm{R}$, now a more realistic reaction have been examined, in order to improve the conformity between the model and the actual system, and it involves *rRNA* and *R-Protein*. Consequently, this modification necessarily leads to the addition of a new ribosomal component, namely the *R-Protein* species.
All the other assumptions regarding the transcription and translation phases, the association and dissociation processes and the degradation are invariant with respect to the first proposed model.

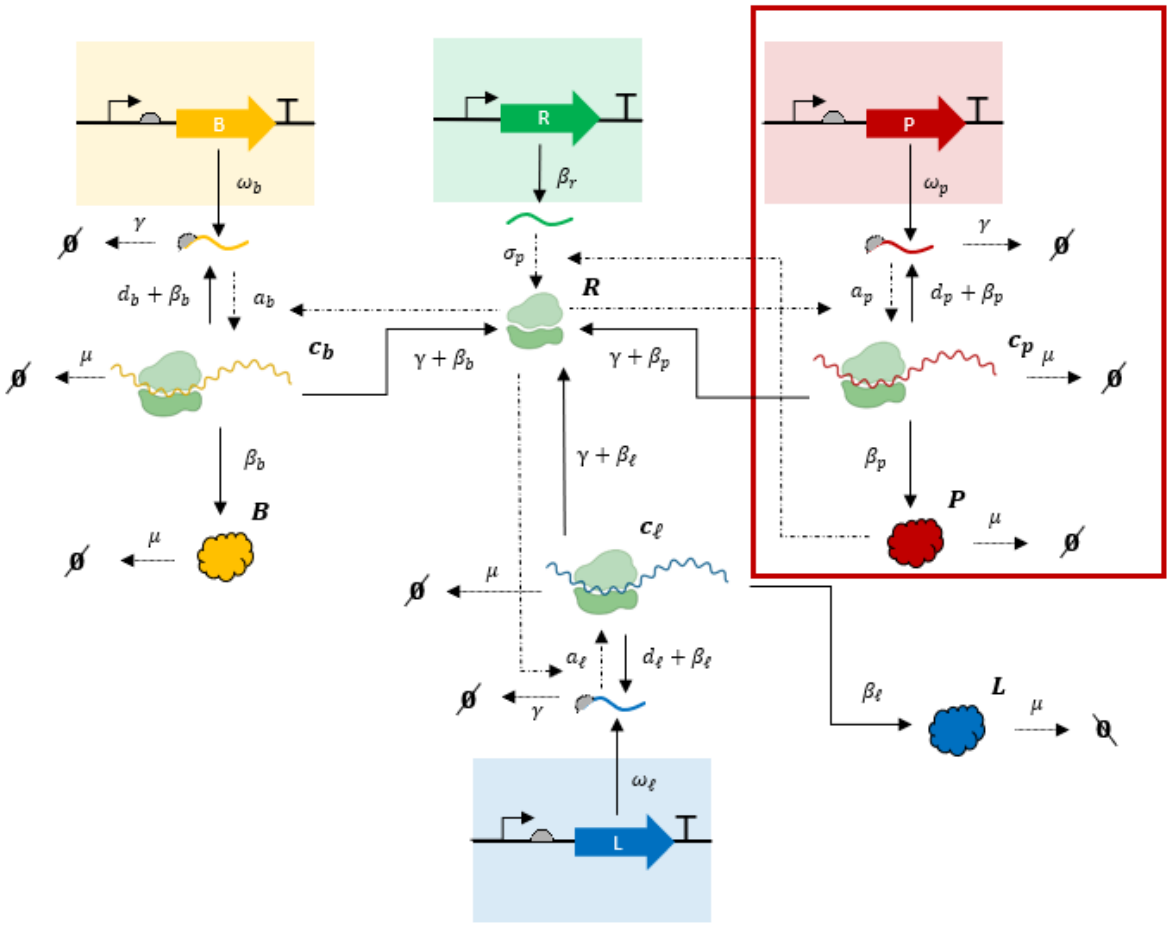

Figure 3.7: Second Model. The yellow, red and blue expression cassettes represent the basal, the *R-Protein* and load species respectively. The green cassette stands for the free ribosomes. The red square highlights the added components, in respect of the first model.

Therefore, the set of chemical reaction has been modified. New reactions have been added

in order to replace $\mathrm{D}_r \xrightarrow{\beta_r} \mathrm{D}_r + \mathrm{R}$:

$$\begin{aligned}
&\mathrm{D}_r \xrightarrow{\beta_r} \mathrm{D}_r + \mathrm{r}\\
&\mathrm{D}_p \xrightarrow{\omega_p} \mathrm{D}_p + \mathrm{m}_p\\
&\mathrm{m}_p + \mathrm{R} \underset{\mathrm{d}_p}{\overset{\mathrm{a}_p}{\rightleftharpoons}} \mathrm{c}_p \xrightarrow{\beta_p} \mathrm{m}_p + \mathrm{R} + \mathrm{P}\\
&\mathrm{r} + \mathrm{P} \xrightarrow{\sigma_p} \mathrm{R}\\
&\mathrm{m}_p, \mathrm{r} \xrightarrow{\gamma} 0\\
&\mathrm{c}_p \xrightarrow{\gamma} \mathrm{R}\\
&\mathrm{P} \xrightarrow{\mu} 0
\end{aligned}$$

Now β_r is the rate of formation of *rRNA* and σ_p is the rate of synthesis of ribosomes. While the mass conservation law 3.2 must be augmented, others must be added to the previous set of laws:

$$\begin{aligned}
R^{TOT} &= R + c_b + c_p + c_\ell && (3.30)\\
m_p^{TOT} &= m_p + c_p && (3.31)\\
r^{TOT} &= r + R^{TOT} && (3.32)\\
P^{TOT} &= P + R^{TOT} && (3.33)
\end{aligned}$$

Starting from the system of ODEs already formulated for the first model, new differential equations must be added and the equations 3.10 and 3.2 must be modified:

$$\begin{aligned}
\dot{m_p} &= \omega_p D_p - a_p m_p R + d_p c_p + \beta_p c_p - \gamma m_p && (3.34)\\
\dot{c_p} &= a_p m_p R - d_p c_p - \beta_p c_p - \gamma c_p - \mu c_p && (3.35)\\
\dot{R} &= \beta_r D_r + \sigma_p r P - \sum_j a_j m_j R + \sum_j d_j c_j + \sum_j \beta_j c_j\\
&\quad + \sum_j \gamma c_j + \mu R \qquad \text{with } j = b, p, \ell && (3.36)\\
\dot{r} &= \beta_r D_r - \sigma_p r P - \gamma r && (3.37)\\
\dot{P} &= \beta_p c_p - \sigma_p r P - \mu P && (3.38)\\
\dot{R}^{TOT} &= \sigma_p r P - \mu R^{TOT} && (3.39)\\
\dot{m}_p^{TOT} &\approx \omega_p D_p - \gamma m_p^{TOT} && (3.40)\\
\dot{r^{TOT}} &= \beta_r D_r - \gamma r - \mu R^{TOT} && (3.41)\\
\dot{P^{TOT}} &= \beta_p c_p - \mu P^{TOT} && (3.42)
\end{aligned}$$

where the equation 3.40 has been derived considering the relation between γ and μ discussed previously.
As in the first case, it is possible to derive the closed solution of equation 3.40 (the closed form solutions of m_b^{TOT} and m_ℓ^{TOT} are invariant):

$$m_p^{TOT} = \frac{\omega_p D_p}{\gamma} \qquad (3.43)$$

The other differential equations should be solved using *fsolve* numerical method from *Matlab*. However, thanks to a further approximation, it is possible to derive the closed

form solution of r starting from equation 3.41. The detailed process is described in Appendix F and it leads to:

$$r = \frac{\beta_r D_r}{\gamma} \tag{3.44}$$

3.3.1 Choice of Parameters' Value

Differently from the first case, the definition of the parameters in condition of maximal growth presents some issues.
The main difficulty comes from the unavailability of data from experiments which have discriminated between basal and *R-Protein* species.
This can be explained due to the complexity of the experiments. Indeed, measuring the total amount of ribosomes inside a cell is not straightforward itself; being able to distinguish between different types of species is even harder.
Additionally, information regarding the amount of *rRNA* and *R-Protein* are reported as relative quantities ([7]).
These evidences make the model not *a priori* identifiable and consequently the definition of the parameters in the maximal growth conditions is not possible.
Consequently, the assumed choice is to use the already estimated parameters, namely the transcription rate ω_b for the basal species and the rate β_r of the process of ribosomes formation, also in this model.
However, the values of ω_p, σ_p and β_p are still missing and needed for the analysis of the load effect.
Regarding ω_p, it seems reasonable that the transcription rates of basal and *R-Protein* species have the same order of magnitude. Indeed both their products are fundamental: while the basal proteins are needed for the survival of the cell, the *R-Proteins* are essential in the process of formation of ribosomes, which are required for every synthesis, even the basal one. This may justify the made hypothesis on the order of magnitude of the transcription rates of basal and *R-Protein* species. According to that and for sake of simplicity, they are assumed to be equivalent.
Concerning the synthesis rate σ_p between the *rRNA* and *R-Protein*, it has ideally the same meaning as β_r in the first model. Hence, it is wise to set σ_p equal to β_r.
Lastly, the translation rate β_p has been assumed with the same order of magnitude of the generation rate β_r of the free ribosomes in the first model.
In congruence with all the stated assumptions, it is trivial to expect that these parameters do not lead to an effective condition of maximal growth, independently on the growth rate function.

3.3.2 Analysis of R^{TOT} variation

As previously, the first analysis that has been made regards the ribosomal competition, namely the distribution of R^{TOT} among the four ribosomal species c_b, c_p, c_ℓ and R.
Again, it does not depend on the chosen growth rate function but uniquely on the manifolds:

$$c_b = \frac{R}{R + K_b} m_b^{TOT} \qquad c_p = \frac{R}{R + K_p} m_p^{TOT} \qquad c_\ell = \frac{R}{R + K_\ell} m_\ell^{TOT} \tag{3.45}$$

and the mass conservation law of R^{TOT}:

$$R^{TOT} = R + c_b + c_p + c_\ell \tag{3.46}$$

where c_b, c_p and c_ℓ must be substituted with their manifolds. Again, it is possible to approximate the equations 3.45 with piecewise linear functions with the same reasoning performed in *M1*.
The Figure 3.8 shows the evolution of the competition as function of R^{TOT}. The continuous lines represent the real evolution, obtained from the numerical solutions of equations 3.45 and 3.46. The dashed lines indicate their approximation, based on the general competition analysis explained in Appendix D.

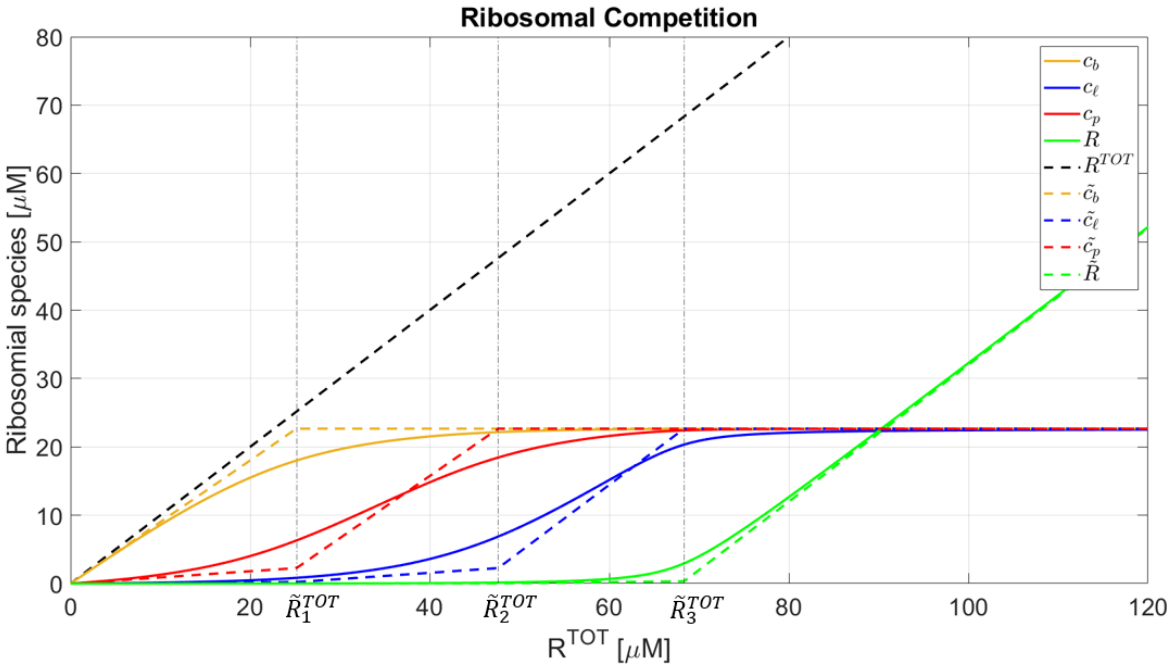

Figure 3.8: Competition between ribosomal species with $K_b << K_p << K_\ell$ and $m_b^{TOT} = m_p^{TOT} = m_\ell^{TOT}$. The yellow, blue, red and green dashed lines represent the approximate evolution of c_b, c_p c_ℓ and R respectively. $\tilde{R}_1^{TOT}$, $\tilde{R}_2^{TOT}$ and $\tilde{R}_3^{TOT}$ are the breaking points

The considerations made for model *M1* persist.
The competition is strongly regulated by the dissociation constants and the total amount of *mRNA*. In particular, the first rule the order of growth of the complexes, while both manage the values of the breaking points. These are fundamental for understanding when a complex reaches its maximal value and consequently another species can start to rise. Since this model considers one ribosomal species more than *M1*, the analysis in Appendix D referred to the general competition must be augmented to the three species case. This

leads to three breaking points (instead of two), which are:

$$\tilde{R}_1^{TOT} = \left(1 + \frac{m_b^{TOT}}{K_b} + \frac{m_p^{TOT}}{K_p} + \frac{m_\ell^{TOT}}{K_\ell}\right) K_b \tag{3.47}$$

$$\tilde{R}_2^{TOT} = \left(1 + \frac{m_p^{TOT}}{K_p} + \frac{m_\ell^{TOT}}{K_\ell}\right) K_p + m_b^{TOT} \tag{3.48}$$

$$\tilde{R}_3^{TOT} = \left(1 + \frac{m_\ell^{TOT}}{K_\ell}\right) K_\ell + m_b^{TOT} + m_p^{TOT} \tag{3.49}$$

Once more, a graph that displays the ratios between the species and the total amount of ribosomes is proposed (Figure 3.9) and it leads to the same observation made for *M1*, namely only one species at time can grow.

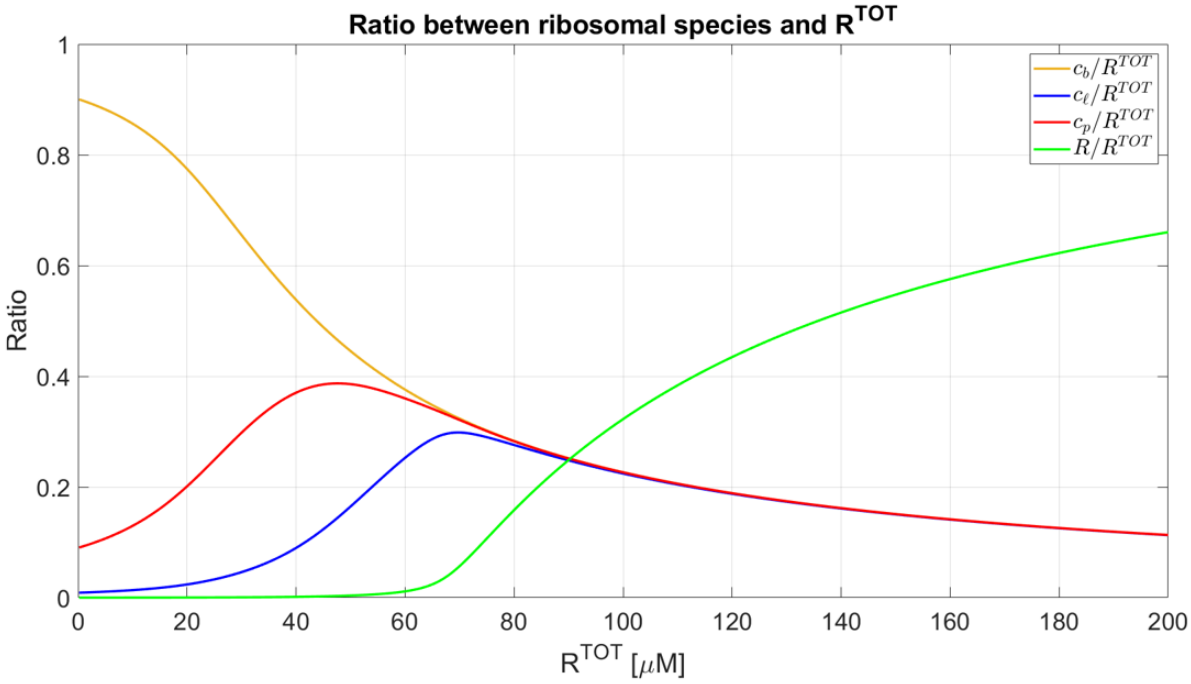

Figure 3.9: Ratio between ribosomal species and R^{TOT}

3.3.3 Analysis of Load effect

In the way that it was studied for the first model, now the analysis of the load effect is proposed.
Again, it is strongly influenced by the growth rate function that has been chosen. Since the previous study have revealed that the *Hill* and the *Linear* functions dependent on c_b are not suitable for the description of the cell growth, this model has not been tested considering all the proposed functions of growth rate. Indeed, two functions have been analyzed: the *Linear* function dependent on R^{active} has been investigated as best option so far, while the *Hill* function dependent on c_b has been analyzed in order to prove its final inaccuracy.
Once more, the parameter that has been varied during the simulation is the transcription

rate of the load ω_ℓ, for the same reasons as before (i.e., the feasibility of laboratory experiments).
Both the simulations are collected in Figure 3.10.
First, the Hill function based on Del Vecchio and Murray ([6]) assumptions ($Hill_{DV}$) has been tested and the results are shown in Panel *A*. As it has been verified with *M1*, increasing the metabolic load makes R^{TOT} grow. This is not consistent with the biological evidences and limitations [8] since it would imply that the total amount of ribosomes can grow endless, just by the rise of the burden.
Thus, since this growth rate function has been already demonstrated to produce inconsistent results with *M1* and *M2*, another further test with a more complex model (*M3*) is not necessary. Hence, it can be already asserted that the *Hill* function, independently of its parameters, is not suitable for the description of the cell growth.

Later, the analysis considering a *Linear* growth rate depending on R^{active} has been performed (Panel *B*). Differently from the previous test, now R^{TOT} does not increase with the growth of the load. At first glance, a possible issue could be noticed: the growth rate is slightly rising while R^{TOT} is slowly decreasing and this might be counter intuitive. However, this is simply a transitory of the system for certain values of the load and it can be explained as follows.
The complex c_ℓ needs ribosomes to be assembled. This amount can come directly from the set of free ribosomes in the cell or it can be stolen from the other complexes. For weak metabolic loads, the requirements of ribosomes is not substantial and then it can be almost entirely provided by the compartment R; at the same time, the complex c_p is used to synthesize new ribosomes. Thus, in this starting configuration of the system: R and c_p are decreasing (c_p can only partially replace the used R, since in turn it requires ribosomes), c_b can be considered unchanged and c_ℓ is growing. Consequently, R^{TOT} (which is the summation of all the ribosomal components) must slightly decrease, since the growth of c_ℓ can not exceed the drop of c_p and R[9]. However, at the same time R^{active} (which is the summation of c_b, c_p and c_ℓ) must increase because the growth of c_ℓ is faster than the drop of c_p and c_b is considered unvaried.
Afterwards, suppose to strengthen the metabolic burden: a decrease of the growth rate is now expected.
As shown in Panel B, the growth rate can still rise until the load (namely the transcription rate ω_ℓ) reaches a threshold value ω_ℓ^*. This value leads the system to a particular configuration of the equilibrium points, where $c_p = 0$ and thus no more ribosomes can be synthesized. Therefore the cell can not form any other complexes and this fact necessarily leads to the death of the cell itself.

[8] A bacterial cell can not contain an unlimited number of ribosomes.

[9] R is the "fundamental unit" in the system since it is needed for the formation of every complex. It is not reasonable to suppose that the system could grow while the ribosomes are decreasing.

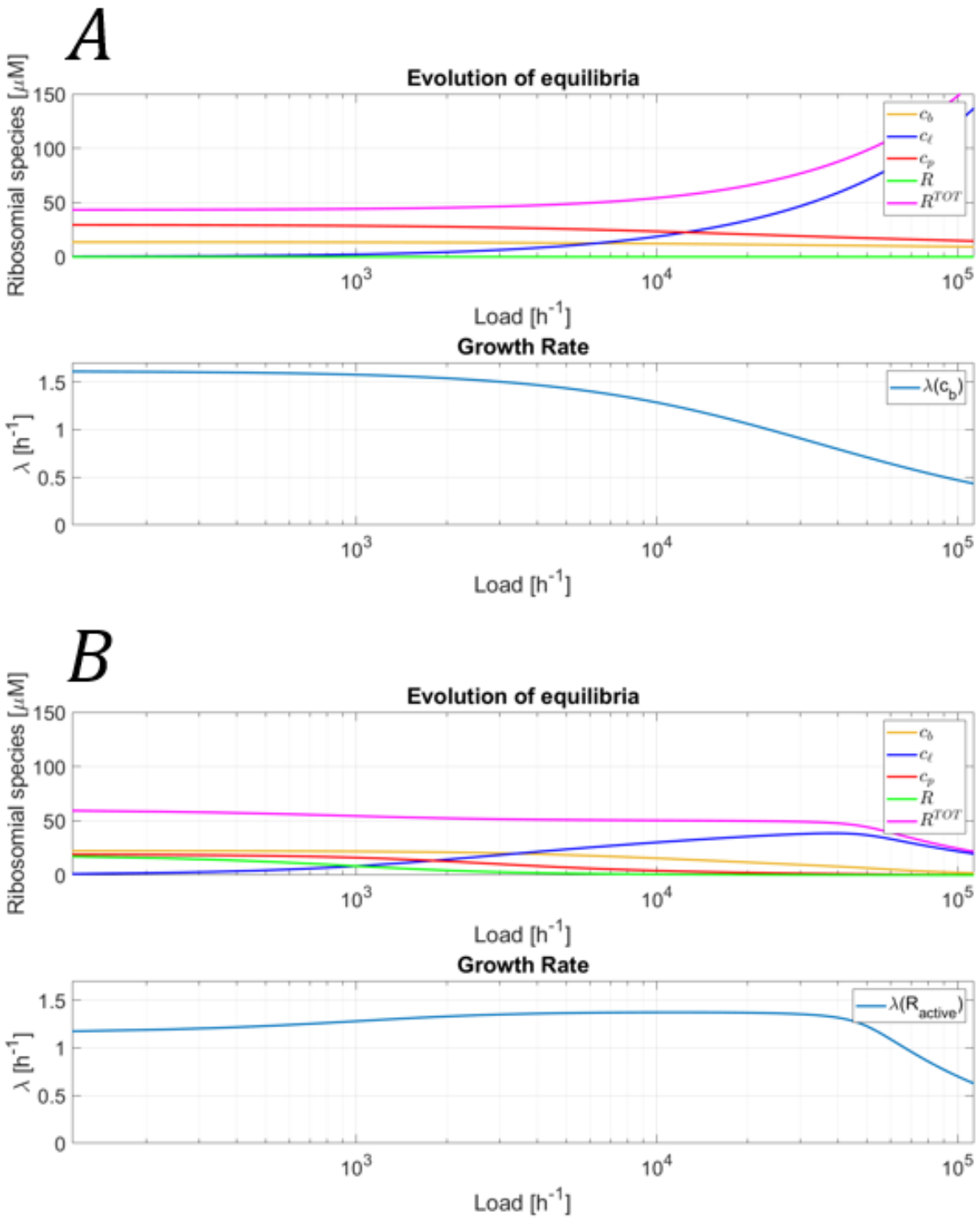

Figure 3.10: Comparison between simulations with different growth rate functions.*A* used the *Linear* function dependent on R^{active}. *B* used the $Hill_{DV}$ function dependent on c_b.

3.4 Model $M3$

The last developed model is the most complex and it still involves the process of formation of the free ribosomes but it also considers one negative feedback (known from literature and better described in the Paragraph related to the biological feedbacks 2.1.4). Roughly, the aim of this feedback is to avoid the endless growth of the free ribosomes, limiting the formation of the *R-Protein mRNA*. According to that, it is reasonable to expect that the feedback assumes a remarkable importance when the metabolic load is null or weak, while it becomes irrelevant when the load is strengthened (because the request of ribosomes is increased and then they tend to bind the *mRNA* instead of be accumulated).

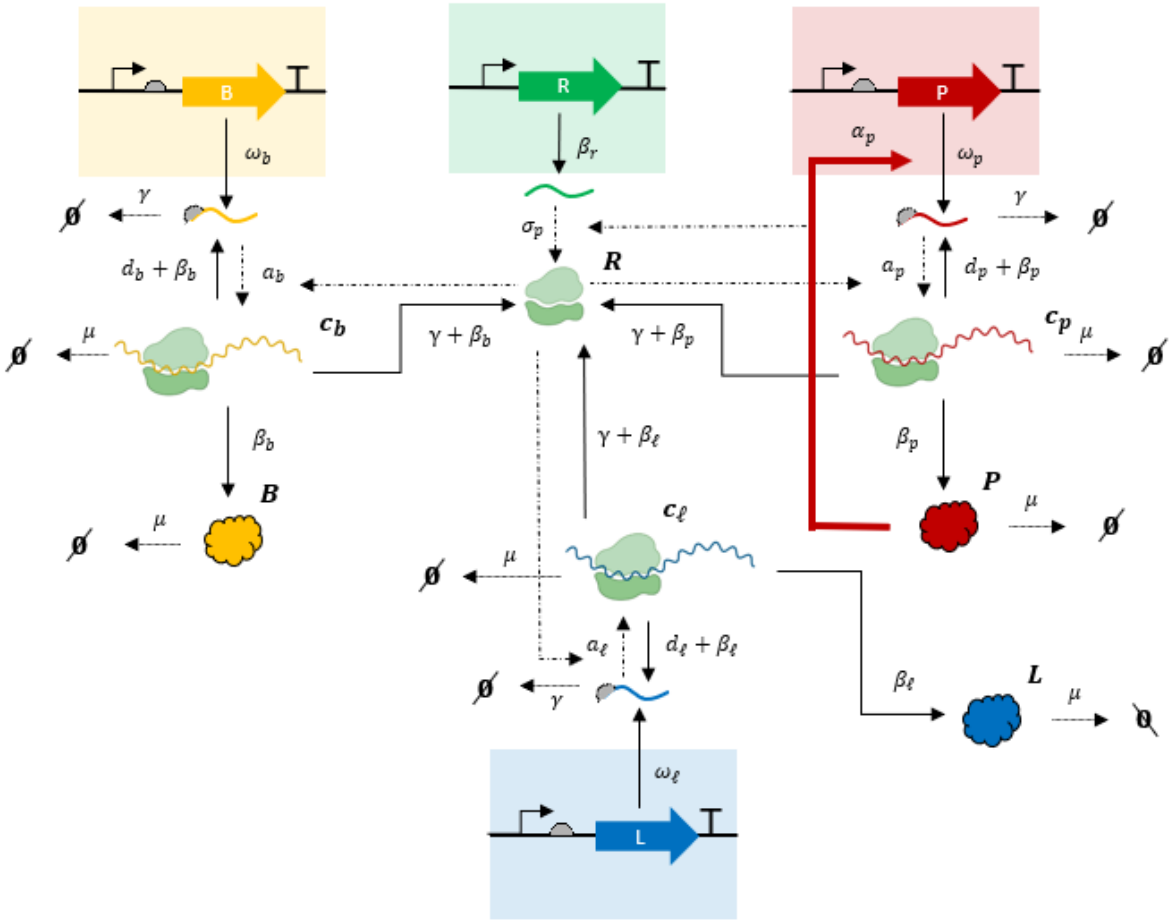

Figure 3.11: Third Model. The yellow, red and blue expression cassettes represent the basal, the *R-Protein* and load species respectively. The green cassette stands for the free ribosomes. The red arrow highlights the added feedback.

According to this improvement, some chemical reactions must be added:

$$\begin{aligned} \mathrm{P} + \mathrm{m}_p &\xrightarrow{\alpha_p} \mathrm{P} : \mathrm{m_p} \\ \mathrm{P} : \mathrm{m_p} &\xrightarrow{\gamma} \mathrm{P} \\ \mathrm{P} : \mathrm{m_p} &\xrightarrow{\mu} 0 \end{aligned}$$

All the mass conservation laws remain unchanged except for m_p^{TOT} (3.31), that becomes:

$$m_p^{TOT} = m_p + c_p + P : m_p \tag{3.50}$$

The system of differential equations must also be modified. Starting from the one formulated for the second model, the ODEs 3.34, 3.38, 3.40 have been adjusted:

$$\dot{m_p} = \omega_p D_p - a_p m_p R + d_p c_p + \beta_p c_p - \alpha_p m_p P - \gamma m_p \tag{3.51}$$
$$\dot{P} = \beta_p c_p - \sigma_p r P + \gamma P : m_p - \alpha_p m_p P - \mu P \tag{3.52}$$
$$\dot{m}_p^{TOT} = \omega_p D_p - \gamma m_p - \gamma P : m_p - \gamma c_p - \mu c_p - \mu P : m_p \tag{3.53}$$

and the following must be added:

$$\dot{P : m_p} = \alpha_p m_p P - \gamma P : m_p - \mu P : m_p \tag{3.54}$$

Again, in 3.53 the approximation is possible considering the relation between γ and μ and leads to:

$$\dot{m}_p^{TOT} \approx \omega_p D_p - \gamma m_p^{TOT} \tag{3.55}$$

Concerning the study of the equations in *Steady State*, nothing changes compared to the second model.

3.4.1 Choice of Parameters' Value

As with model *M2*, the estimation of the parameters in the maximal growth conditions is not possible. Once more, the lack of experimental data that measure the number of ribosomes per species (the addition of the feedback does not influence or reduce the number of the ribosomal species) and the cell growth prevents to define the value of these parameters. For this reason, all the assumptions made with the second model have been preserved and the parameters ω_p, σ_p and β_p are assumed unvaried with respect to $M2$.
The only parameter that must be set is the association rate α_p between the *R-Protein* and its *mRNA*, which can be interpreted as strength of the feedback effect.
The rate α_p must be lower or equal than σ_p (synthesis rate of ribosomes) because otherwise the process of formation of free ribosomes would be overcome by the process of inhibition of themselves, which is inconsistent with the aim of the feedback. Thus, this parameter can assume values in the following interval[10]:

$$0 \leq \alpha_p \leq \sigma_p \tag{3.56}$$

Since it is not possible to derive its precise value from the literature nor from data, it has been set as $0.005 \cdot \sigma_p$[11].
In order to improve this estimate, further analysis should be performed. For example, a sensitivity analysis can be executed for quantifying the contribution of the feedback on the system. Nevertheless, the best solution is to design some laboratory experiments in absence of metabolic load that can measure the actual value of this parameter.

[10] The rates can not be negative.
[11] The feedback should have a limited influence and strength compared to the synthesis of ribosomes.

3.4.2 Analysis of Load effect

Differently from the previous two cases (*M1* and *M2*), the analysis of the distribution of R^{TOT} among the ribosomal species is no more available. This is a direct consequence of the new formulation of the mass conservation law of m_p^{TOT} (3.50), that has been made necessary after the addition of the feedback. Indeed it involves the quantities m_p and c_p, but also $P : m_p$ and this does not allow to derive the manifolds as explained in Appendix C.
In particular, equation C.8 must be replaced by:

$$m_p^{TOT} = \frac{R}{K_p} m_p + m_p + P : m_p \tag{3.57}$$

which leads to:

$$c_p = \frac{R}{R + K_p}(m_p^{TOT} - P : m_p) \qquad \text{and } m_p = \frac{K_p}{R + K_p}(m_p^{TOT} - P : m_p) \tag{3.58}$$

and these equations depend not only on R, but also on $P : m_p$[12]. Thus, an eventual study of the evolution of 3.58 should involve also the differential equation of $P : m_p$ (3.54), which in turn should consider the one of P (3.52). Consequently, the analysis on the ribosomal competition is no more feasible.
Therefore, the analysis of the load effect is here presented.
As already introduced before, the only growth rate that has been tested with this model is *Linear* and it depends on R^{active} (Panel *A* of Figure 3.12). The plotted results are coherent with the biological evidences and considerations similar to the case of *M2* can be made.
Rather, it may be interesting to observe the behavior of the components directly involved in the feedback, namely m_p, $P : m_p$ and P (Panel *B* of Figure 3.12). Before to proceed with the analysis, it is worth to state that the complex $P : m_p$ is directly related to the efficiency of the feedback: the higher the concentration of $P : m_p$, the stronger the feedback. This statement can be derived directly from the chemical reactions that describe the feedback. Indeed, the concentration of $P : m_p$ depends on the amount of *mRNA* that is inhibited.
First of all, it can be noticed that the feedback is efficiently working just for certain values of the load and this is highlighted by the relation between m_p and $P : m_p$.
The condition $P : m_p > m_p$ [13] means that the feedback is currently inhibiting the production of R. In fact, the eventual free *mRNA* that could be used for increasing the quantity of *R-Protein* (and consequently synthesizing more ribosomes) is instead bound and progressively degraded by P itself [14].
Instead, $P : m_p < m_p$ indicates that the effect of the feedback is low. Indeed, the amount of *R-Protein* in the cell is used for synthesizing new ribosomes, according to the chemical reaction:

$$\mathrm{P} + \mathrm{r} \xrightarrow{\sigma_p} \mathrm{R} \tag{3.59}$$

[12] The quantity m_p^{TOT} is constant as shown in 3.43, which remains valid also for *M3*.

[13] m_p stands for the concentration of free *mRNA*, means the quantity of *mRNA* not bound with ribosome yet.

[14] This can be easily proved by noticing that the complex $P : m_p$ that is directly related to the feedback has higher concentration than m_p.

instead of forming the complex $P : m_p$, according to the chemical reaction:

$$\mathrm{P} + \mathrm{m}_p \xrightarrow{\alpha_p} \mathrm{P} : \mathrm{m}_\mathrm{p} \tag{3.60}$$

In other words, since P is used for the formation of both $P : m_p$ and R, in case of limited amount, it must be shared between the two components, which compete against each other for it. Since the bond with the ribosomes is stronger than with the complex $P : m_p$, the formation of R prevails to the detriment of the feedback action.
These considerations are also coherent with the graph of the P protein. The activation phase of the feedback coincides with an abundance of the protein (i.e., the ribosomes can be accumulated due to the high availability of the resources); the disabled feedback matches with the lack of the protein (i.e., the accumulation of ribosomes can not occur). Even though the fact that the feedback is active with lower values of the metabolic burden could sound counter intuitive, it can be explained considering the metabolic requirement of the cell. For small load values, it is not as significant as for stronger loads. This implies that in the first case, an accumulation of R can occur, causing the activation of the feedback, while in the latter one the amount of R is totally spent to accomplish the ribosomal needs of the cell.

3.4.3 Final Remarks

The performed analyses aimed to identify the best model for the cell growth. The cell system have been described by using three different models in order of complexity and some growth rate functions have been tested. Starting from the basic model, four different functions have been explored. This study has revealed that not all the growth rate functions proposed by the literature are suitable. In particular, the *Linear* growth rate dependent on R^{active} has been turned out to be the most adequate. According to that, it has been tested with the second model and, for sake of completeness, one of the rejected functions has been analysed as well, in order to confirm its deficiency. The analysis has supported the previous guess. Finally, the most complex model has been investigated with the chosen growth rate function. Also in this case, it has produces coherent results with the biological evidences. Therefore, it has been possible to affirm that the *Linear* growth rate function dependent on R^{active} is the most suitable for the description of the cell growth, independently on the complexity of the model.

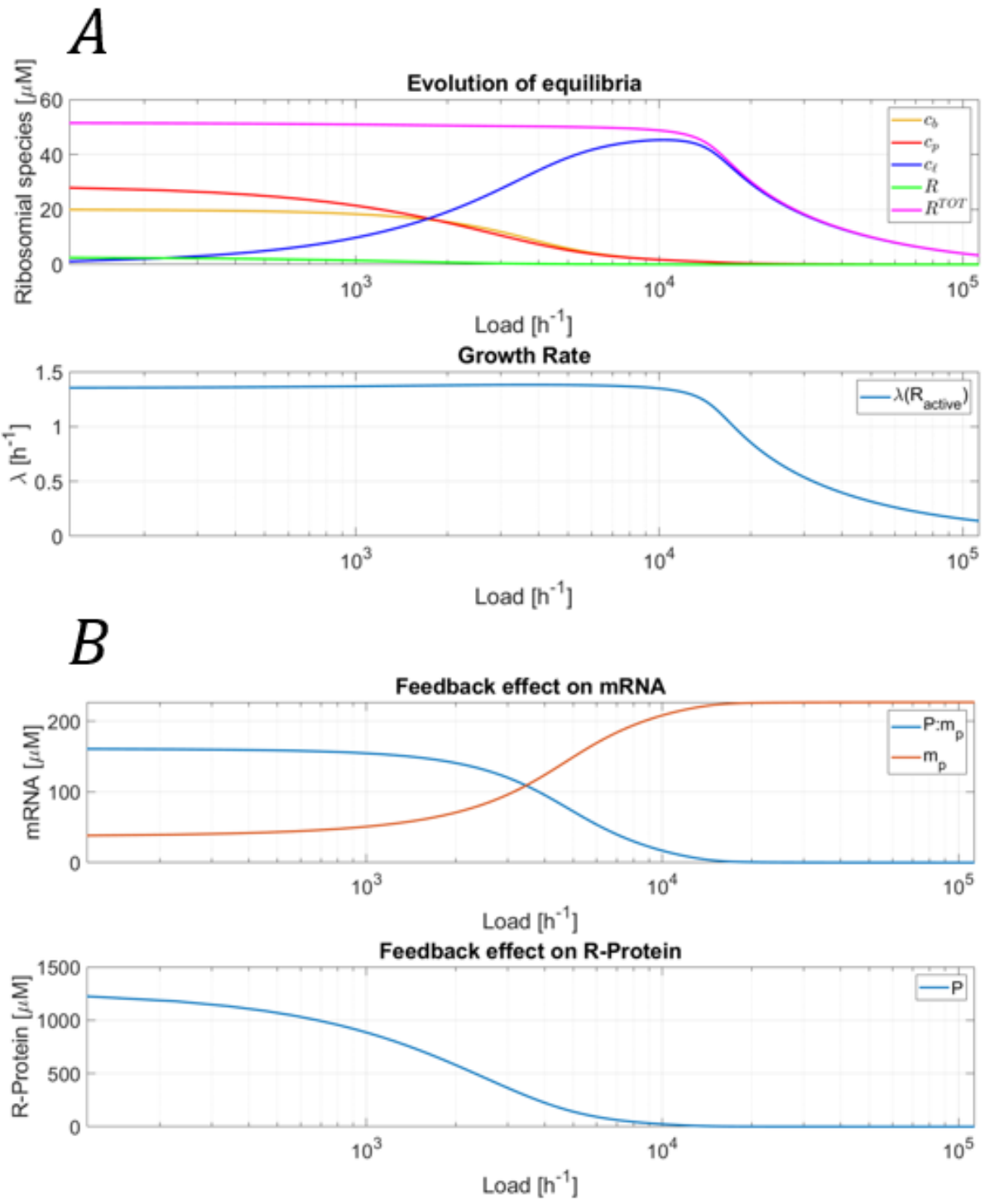

Figure 3.12: Analysis of *M3* with a *Linear* growth rate dependent on R^{active}. *A* shows the evolution of the equilibrium points and the growth rate. *B* displays the feedback effect on $P : m_p$, m_p and P.

Chapter 4

Discussion and Results

In this chapter, the results obtained by the study of the equilibrium points are resumed in order to justify once more the chosen growth rate function. Afterwards, the study of the dynamics of the model *M3* will be discussed, in comparison with the only article from the literature that has analyzed the effects of the load on the cell system.

4.1 Best Choice considering the Load

In the previous chapter, three different models in order of complexity have been introduced and tested considering various growth rate functions.
The growth of the first model *M1* has been analyzed taking into account four various functions: *Linear* dependent on c_b, *Linear* dependent on R^{active}, *Hill* from Del Vecchio and Murray assumption [6] ($Hill_{DV}$) and finally *Hill* from the fitting of the data in [1]. By varying the strength of the metabolic load, the evolution of the equilibrium points (in particular referred to c_b, c_ℓ, R and R^{TOT}) and the growth rate have been examined (Figure 3.6).
At first sight, it can be noticed that the evolution of R^{TOT} is similar in three cases (Panel A, C and D). In these simulations, the stronger the load, the higher the amount of total ribosomes in the cell. This evidence has been used as reason to discard the growth rate functions that have led to this result. Indeed, it is not reasonable nor feasible to have a (potentially) endless concentration of ribosomes in the cell. In other words, the reinforcement of the load can not cause the increment of the ribosomal component. In fact, if this would be the case, the cell would have the possibility to grow endless and this not biological plausible.
Instead, the *Linear* function dependent on R^{active} (Panel B) has shown the opposite result, namely for stronger loads, the concentration of the ribosomes decreases. This outcome is coherent with the cell system. Indeed, since the load is an external component that had been added to the system and it needs resources (i.e., ribosomes) to be synthesized, it must limit the activity of the cell, namely it must negatively influence its growth. Consequently, an increment of the ribosomal amount is not consistent with the growth decline. Hence, this allows to state that the supposed best growth rate function is *Linear* dependent on R^{active}.
However, this conclusion must be verified also with the other two models, since the final aim is to find the best option independently on the complexity of the structure.

The simulation with model *M2* is shown in Figure 3.10, Panel *B*. The results confirm the previous statement. Again, the selected growth rate model has produced coherent outcomes: the stronger the load, the slower the growth.
Finally, it has been tested with the third model *M3* (Figure 3.12, Panel 3). Once more, it has been confirmed as best option over the four different types of growth rate proposed at the beginning. This can be justified as in the case of *M1* and *M2*.
Thus, this leads to the final statement that the best function for modeling the growth of a single bacterial cell, independently on the complexity of the model structure, is the *Linear* growth rate dependent on R^{active}, where R^{active} is the summation between all the ribosomal complexes or alternatively the difference between the total number of ribosomes and the free ribosomes.

4.2 Dynamics of the system

As stated in the previous paragraph, the best choice for the modeling of the cell growth is the *Linear* growth rate dependent on R^{active}.
The articles presented in Paragraph 2.2.2 do not allow to make a comparison with the achieved results, since all of them have not considered the addition of the metabolic load. However, *Nikolados et al.* [2] have studied how the resources of the cell are re-assigned after an inducible gene (i.e., the gene responsible for the synthesis of the load) has been added.

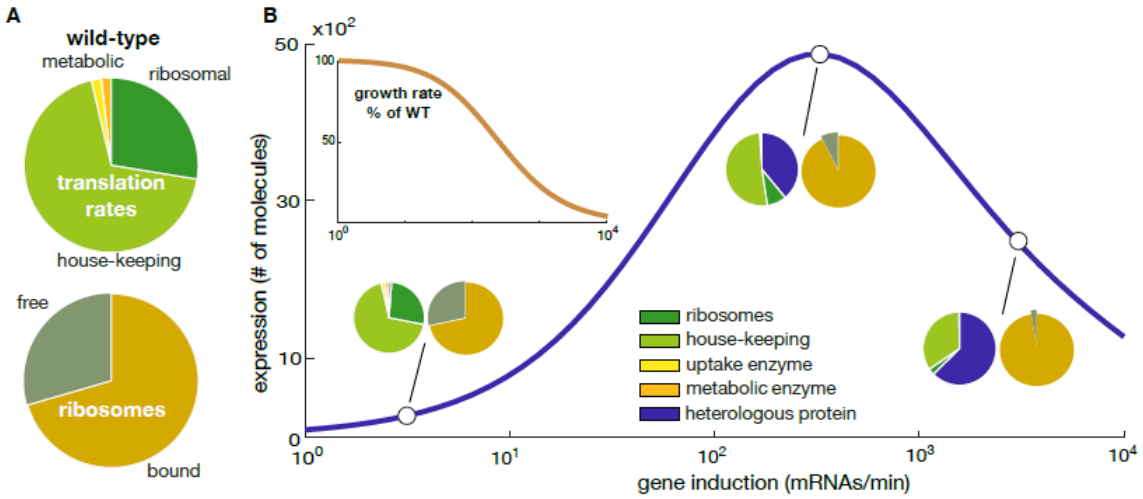

Figure 4.1: Simulation of an inducible gene [2].

In Figure 4.1, the performed simulation is reported. The authors have divided the ribosomal content into five categories: the ribosomes (which corresponds to the R-protein species in *M2* and *M3*), the house-keeping (species responsible for the basic functions of the cell), the uptake enzyme (which is responsible for the nutrients transport inside the cell), the metabolic enzyme (that is responsible for the translation of the nutrients into energy) and the heterologous protein (namely the metabolic load).
This classification is similar to the one performed and explained in the paragraph of biological scenario (Paragraph 2.1.3): different species have been gathered based on their

purposes. Rather than five, the model *M3* presents four species: the free ribosomes, the basal, the R-Protein and the load species.
The starting configuration of the system (i.e., the distribution of the resources among all the different categories or species and the ratio between free and bound ribosomes) is shown Panel *A*. In this initial configuration, the load gene has not been inserted yet and the cell is in equilibrium.
Afterwards, the burden has been added: the evolution of its expression and the distribution among all the species have been studied as function of the rate of the gene induction (Panel *B*).
It is interesting to compare this simulation with the one performed for model *M3*. For a better visualization, the evolution of the load shown in Figure 3.12, Panel *A*, is here reported singularly.

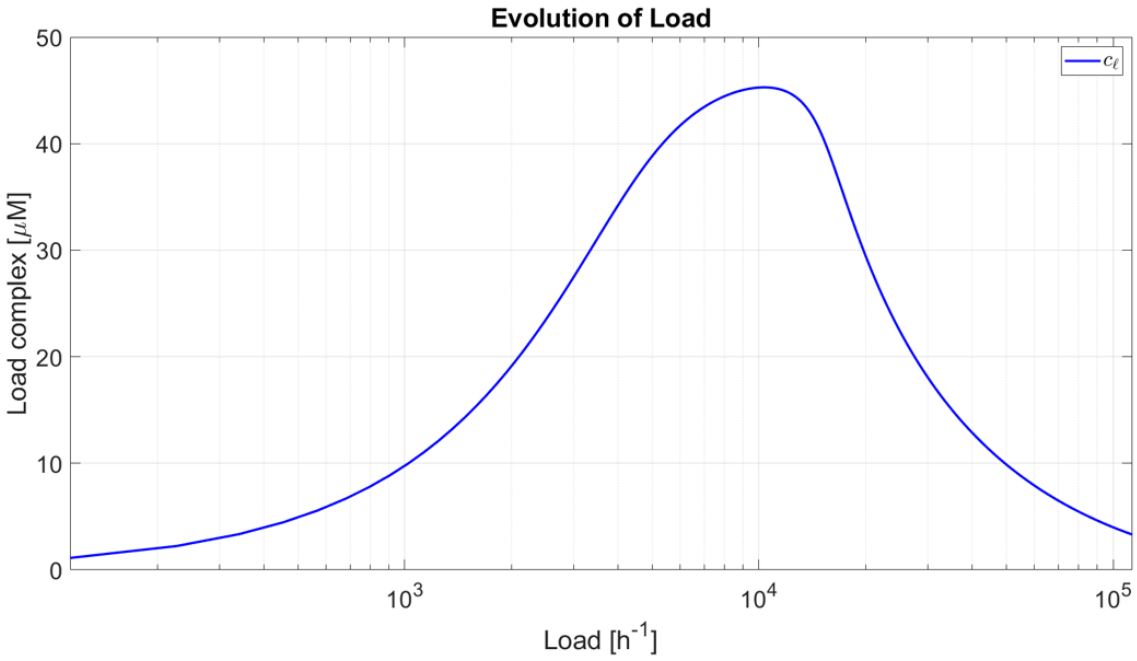

Figure 4.2: Evolution of the load in *M3*.

It can be easily seen that the model result is close to the one proposed in the article [2]: the load expression has a peak for a certain value of the transcription rate and then it starts to decrease. Unfortunately, it is not possible to compare the behavior of the other species since the performed classification is not the same in the two cases. This reveals a sort of conformity between the results of the designed model and the outcomes presented in the article.
A further analysis can be proposed, namely the study of the dynamics of the system (i.e., its evolution over time), based on model *M3*. The dynamics[1] of the system are shown in Figure 4.3 and they have been computed using *Matlab* toolboxes *SymBiology Model Builder* and *SimBiology Model Analyzer* (further details in Appendix G). In the upper

[1]Initial conditions: $R = 23.6\mu M$, $c_b = 10.2\mu M$, $c_p = 10.2\mu M$, $P = 0$, $r = 0.697\mu M$, all the other components were set equal to 0. The transcription rate of the load was set equal to $1.29 \times 10^6 \frac{\mu M}{h}$, the value of the other parameters are reported in Appendix A.

graph, the dynamics of the ribosomal components (R, c_b, c_p and c_ℓ) are plotted.

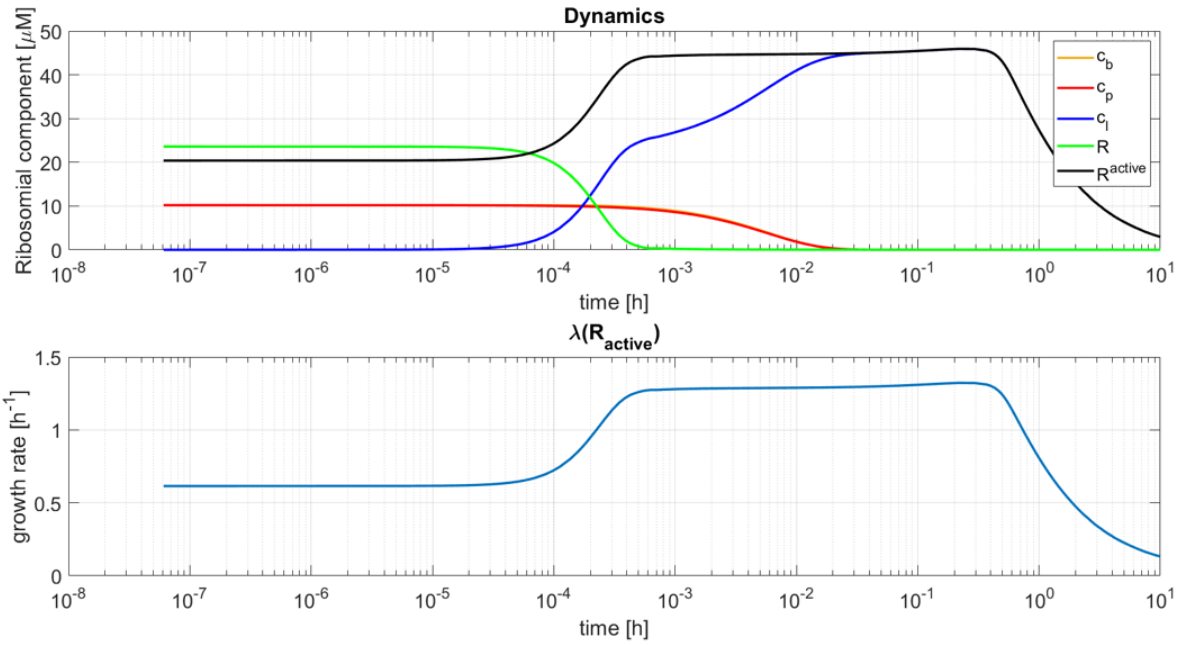

Figure 4.3: Dynamics of model *M3*.

Since the starting value of c_b and c_p depends uniquely on their initial condition ($c_b^0 = c_p^0$) and their evolution is similar (the load will steal the ribosomes equally from c_b and c_p), their curves are overlapped. Furthermore, at the beginning, the only null component is c_ℓ^0 and thus $\lambda^0 \neq 0$ [2]. It is straightforward to notice that the increment of the load coincides with the drop of all the other ribosomal complexes. Additionally, the first component to be affected by the presence of the load is the compartment of the free ribosomes R , which decreases before all the others. Afterwards, when the compartment is empty, c_b and c_p are consumed.
The explanation lies in the amount of needed ribosomes. As already described, the synthesis of the load requires ribosomes. When the related compartment is used up, they must be stolen from somewhere else, namely from the other ribosomal complexes. This justifies the drop (and its order) of the free ribosomes and the complexes. Once the peak of the load expression is reached, it can be maintained by the system for a short interval of time. Later, it will start decreasing due to the unavailability of other ribosomes, its spontaneous decay, the effect of the growth rate and the translation of the complex into protein.
It is interesting to analyze the relation between the dynamics and the growth rate. The maximum growth rate is achieved before the peak of the load expression and it has been maintained until the start of the drop of c_ℓ. This is expected, since the growth rate is a linear function dependent on R^{active} (black curve). Recalling that $R^{active} = R^{TOT} - R = c_b + c_p + c_\ell$, R^{active}_{max} is obtained when $R = 0$.
The dynamics of the cell growth is coherent with the biological evidences, which can be interpreted as another proof that the chosen function is suitable for modeling the cell

[2] λ depends linearly on R^{active}. Since $c_b^0 \neq 0$, $c_p^0 \neq 0$ and $c_\ell^0 = 0$, then $R^{active,0} = c_b^0 + c_p^0 \neq 0$.

growth. Indeed it has been experimentally tested that the addition of a strong load does not lead the cell immediately to the death but instead it grows for a while and then stabilizes. After, the drop of the growth rate is verified, potentially causing the death of the whole system.

Finally, it has been possible to check the goodness of the chosen function also considering the dynamics over time of the cell system. This strengthens the statement that the *Linear* growth rate function is the best option for the description of the cell growth process.

Chapter 5

Conclusion and Future Work

5.1 Summary

In this thesis, the cell growth has been analyzed.
First, the system has been modeled, according with the chemical reactions that occur in it. Due to the complexity of the system itself, some approximations and hypotheses have been proposed, in order to simplify the description. Consequently, three models have been developed, in increasing order of complexity; at the same time, different functions that aim to characterize the cell growth have been presented. Starting from the least complex model, the growth have been studied, testing all the functions. Later, just the function that has produced some coherent results has been tested with the second and the third model.
Finally, it has been possible to state that the best growth rate function is the *Linear* dependent on R^{active}.

5.2 Future Prospective adding new feedback

Possible improvements to this work can be performed, especially regarding the models. Their complexity can be increased, reaching a better level of description. In particular, new feedbacks can be added. It is reasonable to present these improvements starting from $M3$, since it is the best model so far (i.e., the model that is close to the actual system).
The third model presents a biological feedback that aims to avoid the accumulation of the free ribosomes inside the cell. However, this regulates just the synthesis of the *R-Protein* and not of the *rRNA*, which is the other reactant that composes the ribosomes. The explanation behind this choice lies in the chemical reactions that have been identified for describing the cell system. Indeed, the production of *rRNA* has been modeled as a simple reaction $\mathrm{D}_r \xrightarrow{\beta_r} \mathrm{D}_r + \mathrm{r}$, namely the *rRNA* is obtained directly from the *DNA*, which is not possible to influence negatively.
Therefore, once a better description of the process of derivation of *rRNA* will be achieved, it will be also feasible to implement a new regulatory feedback on the production of the ribosomes that accounts both the components (*R-Protein* and *rRNA*).
Nevertheless, this model does not consider the needed amount of amino acids. Indeed, even though the translation was modeled as in Paragraph 2.1.3, namely it just requires the *mRNA* and the ribosome, this process is an approximation and the real one involves

more components. In fact, it would need also amino acids and the *tRNA*, which is the *RNA* responsible for the transport of the amino acids in the site of the translation. Consequently, once this improvement will be performed (namely some chemical reactions and components must be added to the system in order to model also the presence of another type of *RNA* and the amino acids), other feedback loop can be added. In particular, this can be inspired by the paper [11], where it has been already assumed that there should be a flux of amino acids in the cell and it should influence the growth rate.
In general, all the future improvements must start from the augmentation of the chemical reactions and/or species, which allows to reach a better description of the cell system.
However, the main limitation of all this dissertation is the unavailability of experimental data. Future work may also regard the design of some experiments in order to validate the model $M3$ (it is reasonable to test just the most complete model) and the results obtained with the identified best growth rate function.

Bibliography

[1] Marr AG. "Growth rate of Escherichia coli". In *Microbiol Rev.* 1991 Jun;55(2):316-33. DOI: 10.1128/mr.55.2.316-333.1991. PMID: 1886524; PMCID: PMC372817.

[2] Evangelos-Marios Nikolados, Andrea Y.Weiße and Diego A.Oyarzún. "Prediction of Cellular Burden with Host-Circuit Models". In *Synthetic Gene Circuits.* In *Synthetic Gene Circuits* (2021), pp.267-291. DOI: 10.1007/978-1-0716-1032-9_13.

[3] Chenhao Wu, Rohan Balakrishnan, Nathan Braniff, Matteo Mori, Gabriel Manzanarez, Zhongge Zhang and Terence Hwa. "Cellular perception of growth rate and the mechanistic origin of bacterial growth rate". In *Proceedings of the National Academy of Sciences* (2022), Vol.119, No. 20. DOI: 10.1073/pnas.2201585119.

[4] Bruce R. Levin, Ingrid C. McCall, Véronique Perrot, Howard Weiss, Armen Ovesepian, Fernando Baquero. "A Numbers Game: Ribosome Densities, Bacterial Growth, and Antibiotic-Mediated Stasis and Death". In *mBio*2017 Feb 7;8(1):e02253-16. DOI: 10.1128/mBio.02253-16. PMID: 28174311; PMCID: PMC5296603.

[5] Eliza Atkinson, Zoltan Tuza, Giansimone Perrino, Guy-Bart Stan and Rodrigo Ledesma-Amaro. "Resource-aware whole-cell model of division of labour in a microbial consortium for complex-substrate degradation". In *Microb Cell Fact* 21, 115 (2022). DOI: 10.1186/s12934-022-01842-0

[6] Domitilla Del Vecchio and Richard M.Murray. "Biomolecular Feedback Systems". Princeton University Press (2014)

[7] Bremer H, Dennis PP. Modulation of Chemical Composition and Other. "Parameters of the Cell at Different Exponential Growth Rates". In *EcoSal Plus.* 2008 Sep;3(1). DOI: 10.1128/ecosal.5.2.3. PMID: 26443740.

[8] Condon C, Squires C, Squires CL. "Control of rRNA transcription in Escherichia coli". In *Microbiol Rev.* 1995 Dec;59(4):623-45. DOI: 10.1128/mr.59.4.623-645.1995. PMID: 8531889; PMCID: PMC239391.

[9] Srivastava AK, Schlessinger D. "Mechanism and regulation of bacterial ribosomal RNA processing". In *Annu Rev Microbiol.* 1990;44:105-29. DOI: 10.1146/annurev.mi.44.100190.000541. PMID: 1701293.

[10] Nomura M, Gourse R, Baughman G. "Regulation of the synthesis of ribosomes and ribosomal components". In *Annu Rev Biochem.* 1984;53:75-117. DOI: 10.1146/annurev.bi.53.070184.000451. PMID: 6206783.

[11] Scott M, Klumpp S, Mateescu EM, Hwa T. "Emergence of robust growth laws from optimal regulation of ribosome synthesis". In *Mol Syst Biol.* 2014 Aug 22;10(8):747. DOI: 10.15252/msb.20145379. PMID: 25149558; PMCID: PMC4299513.

[12] Ewart Carson, Claudio Cobelli. "Modelling Methodology for Physiology and Medicine". Elsevier (2013)

[13] Cerretti DP, Dean D, Davis GR, Bedwell DM, Nomura M. "The spc ribosomal protein operon of Escherichia coli: sequence and cotranscription of the ribosomal protein genes and a protein export gene". In *Nucleic Acids Res.* 1983 May 11;11(9):2599-616. DOI: 10.1093/nar/11.9.2599. PMID: 6222285; PMCID: PMC325911

[14] Alexander Johnson, Bruce Alberts, David Morgan, Karen Hopkin, Keith Roberts, Martin Raff, Peter Walter. "Essential Cell Biology". Garland Science (2009).

[15] Verma SC, Qian Z, Adhya SL. "Architecture of the Escherichia coli nucleoid". In *PLoS Genet.* 2019 Dec 12;15(12):e1008456. DOI: 10.1371/journal.pgen.1008456. Erratum in: PLoS Genet. 2020 Oct 21;16(10):e1009148. PMID: 31830036; PMCID: PMC6907758.

[16] Lynch M, Marinov GK. "The bioenergetic costs of a gene". In *Proc Natl Acad Sci U S A.* 2015 Dec 22;112(51):15690-5. DOI: 10.1073/pnas.1514974112. Epub 2015 Nov 2. PMID: 26575626; PMCID: PMC4697398.

Appendix A

Summary Tables

Two summary tables are here reported, in order to simplify the reading of the models.

Parameter	**Description**	**Model**
m_b	Basal *mRNA*	1,2,3
m_p	R-Protein *mRNA*	2,3
m_ℓ	Load *mRNA*	1,2,3
c_b	Basal complex	1,2,3
c_p	R-Protein complex	2,3
c_ℓ	Load complex	1,2,3
B	Basal protein	1,2,3
P	R-Protein protein	2,3
L	Load protein	1,2,3
R	Free Ribosomes	1,2,3
R^{TOT}	Total Ribosomes	1,2,3
r	Free *rRNA*	2,3
r^{TOT}	Total *rRNA*	2,3
P^{TOT}	Total R-Protein content	2,3
m_b^{TOT}	Total basal *mRNA* content	1,2,3
m_p^{TOT}	Total R-Protein *mRNA* content	2,3
m_ℓ^{TOT}	Total load *mRNA* content	1,2,3
$P : m_p$	Complex between R-Protein and its *mRNA*	3
λ	Growth rate	1,2,3
γ	Spontaneous decay of *mRNA* and *rRNA*	1,2,3
δ	Spontaneous decay of proteins, ribosomes and complexes	1,2,3
μ	Sum of growth rate and spontaneous decay	1,2,3

Table A.1: Table of variables

Parameter	Description	Value	Ref.	Model
γ	Decay of *mRNA* and *rRNA*	100 h^{-1}	[6]	1,2,3
δ	Decay of proteins, complexes and ribosomes	0.05 h^{-1}	[16]	1,2,3
a_b	Association rate of c_b	2.94 $(\mu Mh)^{-1}$	E	1,2,3
d_b	Dissociation rate of c_b	1 h^{-1}	E	1,2,3
a_p	Association rate of c_p	0.294 $(\mu Mh)^{-1}$	E	2,3
d_p	Dissociation rate of c_p	1 h^{-1}	E	2,3
a_ℓ	Association rate of c_ℓ	29.4 $(\mu Mh)^{-1}$	D	1,2,3
d_ℓ	Dissociation rate of c_ℓ	1 h^{-1}	D	1,2,3
β_b	Translation rate of B	69.7 h^{-1}	A	1,2,3
β_p	Translation rate of P	69.7 h^{-1}	B	2,3
β_ℓ	Translation rate of L	69.7 h^{-1}	D	1,2,3
ω_b	Transcription rate of m_b	2266.7 h^{-1}	A	1,2,3
ω_p	Transcription rate of m_p	2266.7 h^{-1}	B	2,3
ω_ℓ	Transcription rate of m_ℓ	2266.7 h^{-1}	D	1,2,3
σ_p	Association rate between r and P	69.7 $(\mu Mh)^{-1}$	B	2,3
α_p	Association rate between m_p and P	0.3485 $(\mu Mh)^{-1}$	C	3
β_r	Transcription rate of r	69.7 h^{-1}	A	1,2,3
D_b	Basal *DNA*	1 μM	[15]	1,2,3
D_p	R-Protein *DNA*	1 μM	[15]	2,3
D_ℓ	Load *DNA*	1 μM	[15]	1,2,3
D_r	Load *rRNA*	1 μM	[15]	1,2,3

Table A.2: Table of parameters

A: It has been computed as shown in Paragraph 3.2.1.

B: It has been computed as shown in Paragraph 3.3.1.

C: It has been computed as shown in Paragraph 3.4.1.

D: It has been assumed that the added load has transcription and translation rates similar to the other species. Furthermore, the dissociation constant is assumed to be smallest, in order to let it steal the ribosomes from the other species and observe some effects on the system.

E: Considering the dissociation constants in [6], the dissociation rates have been assumed equal independently of the species and the association rates have been set differently. Taking into account the importance of each species (Paragraph 2.1.3), the following relation between basal and R-Protein constant has been established: $K_b << K_p$. Moreover, the constants has been calculated in relation of $R_{max}^{TOT} = 34\mu M$ (i.e., $K_b = 0.01 \times R_{max}^{TOT}$).

Appendix B

Negligibility of RNAP

A generic chemical reaction that describes the transcription can be written as follows:

$$\mathrm{D} + \mathrm{RNAP} \underset{\mathrm{p^-}}{\overset{\mathrm{p^+}}{\rightleftharpoons}} \mathrm{CTR} \xrightarrow{\omega} \mathrm{D} + \mathrm{m} + \mathrm{RNAP} \tag{B.1}$$

where D is the DNA, p^+ and p^- are the association and dissociation rate respectively between DNA and RNA, CTR is pre-$mRNA$ molecule, ω is the transcription rate, m is the $mRNA$.
The relative mass law kinetics are (the degradation is not considered):

$$\begin{aligned}
\dot{D} &= p^- CTR - p^+ D \cdot RNAP + \omega CTR && \text{(B.2)}\\
\dot{m} &= \omega CTR && \text{(B.3)}\\
\dot{CTR} &= -p^- CTR + p^+ D \cdot RNAP - \omega CTR && \text{(B.4)}\\
\dot{RNAP} &= 0 && \text{(B.5)}
\end{aligned}$$

where B.5 is equal to zero under the hypothesis of $RNAP$ constant.
Studying the differential equations at the steady state, CTR is:

$$CTR_{ss} = \frac{RNAP}{K} D \qquad \text{with } K = \frac{p^-}{p^+} \tag{B.6}$$

that depends uniquely on D.
This consideration can lead to a possible reformulation of the chemical reaction:

$$\mathrm{D} \xrightarrow{\omega'} \mathrm{D} + \mathrm{m} \tag{B.7}$$

Now, the relation between ω and ω' must be found.
By the substitution of B.6 in B.3, the dynamics of $mRNA$ becomes:

$$\dot{m} = \omega \frac{RNAP}{K} D \tag{B.8}$$

while the mass action law derived from B.7 is:

$$\dot{m} = \omega' D \tag{B.9}$$

Now it can be easily seen that the relation between the transcription rates is:

$$\omega' = \omega \frac{RNAP}{K} \tag{B.10}$$

From this point on, the $RNAP$ will be considered as constant, leading to the reformulation of all transcription chemical reactions as in B.7.
For sake of simplicity, the new transcription rate ω' will be indicated with the symbol ω as well.

Appendix C

Manifolds formal derivation

The manifold derivation is based on the difference between fast and slow dynamics observed in the differential equations, namely on the *Rapid Equilibrium* assumption.
Here the mathematical calculations are computed for a generic species i. Clearly, the same procedure can be applied for every species.
Considering the first model (the derivation of the manifold is identical for all the models), the fast dynamics are:

$$\dot{m}_i = \omega_i D_i - a_i m_i R + d_i c_i + \beta_i c_i - \gamma m_i \tag{C.1}$$
$$\dot{c}_i = a_i m_i R - d_i c_i - \beta_i c_i - \gamma c_i - \mu c_i \tag{C.2}$$

while the slow dynamics consists on the total amount of *mRNA*:

$$\dot{m_i^{TOT}} = \omega_i D_i - \gamma m_i^{TOT} - \mu c_i \tag{C.3}$$

It is possible to rewrite a_i as $\frac{d_i}{K_i}$:

$$\dot{m}_i = \omega_i D_i - d_i m_i \frac{R}{K_i} + d_i c_i + \beta_i c_i - \gamma m_i \tag{C.4}$$
$$\dot{c}_i = d_i m_i \frac{R}{K_i} - d_i c_i - \beta_i c_i - \gamma c_i - \mu c_i \tag{C.5}$$

Now fixing $d_i = \frac{1}{\epsilon}$ in C.1:

$$\dot{m}_i = \frac{1}{\epsilon}(-m_i \frac{R}{K_i} + c_i) + \omega_i D_i + \beta_i c_i - \gamma m_i \tag{C.6}$$

For the condition of fast dynamics $\epsilon \longrightarrow 0$, means that the dissociation rate tends to zero. Consequently, the obtained manifold is:

$$-m_i \frac{R}{K_i} + c_i = 0 \longrightarrow c_i = \frac{R}{K_i} m_i \tag{C.7}$$

Remind that $m_i^{TOT} = c_i + m_i$, then:

$$m_i^{TOT} = \frac{R}{K_i} m_i + m_i \tag{C.8}$$

and finally:

$$c_i = \frac{R}{R + K_i} m_i^{TOT} \qquad \text{and} \qquad m_i = \frac{K_i}{R + K_i} m_i^{TOT} \tag{C.9}$$

which shows that the evolution of c_i and m_i uniquely depends on the total amount of *mRNA* (m_i^{TOT}) and the strength of the bound *mRNA*-ribosome (K_i).
Since it can be proved that m_i^{TOT} depends on the transcription rate ω_i, the only factor that does not influence these quantities is the translation rate β_i.

Appendix D

General competition

The general competition between ribosomal species can be seen as an input-output system as shown in Figure D.1.

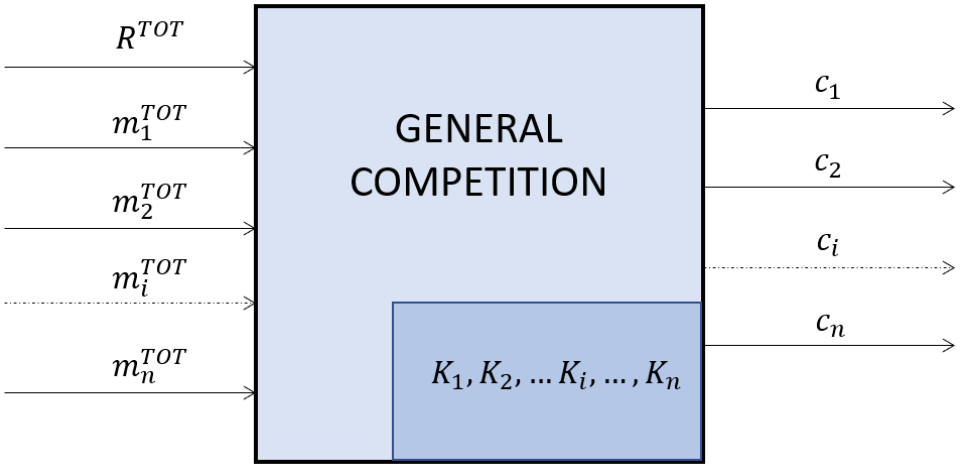

Figure D.1: Input-Output system of general competition

The outputs are of course the ribosomal complexes c_i. Since their evolution depends on the manifold, the system is characterized by the dissociation constant K_i, while the inputs are the total amount of $mRNA$ and ribosomes R^{TOT} available.
Recalling that:

$$c_i = \frac{R}{R + K_i} m_i^{TOT} = \frac{\frac{R}{K_i}}{1 + \frac{R}{K_i}} m_i^{TOT} \tag{D.1}$$

an approximated version can be proposed, in order to simplify the analytical computations (problems related to this estimation will be analyzed in the following Appendix):

$$c_i \approx \begin{cases} \frac{m_i^{TOT}}{K_i} R & , R \leq K_i \\ m_i^{TOT} & , R > K_i \end{cases} \tag{D.2}$$

Without loss of generality, let consider the simple case of just two competitive species with $K_1 < K_2$.

Then the approximated ribosomal complexes are:

$$c_1 \approx \begin{cases} \frac{m_1^{TOT}}{K_1} R & , R \leq K_1 \\ m_1^{TOT} & , R > K_1 \end{cases} \quad \text{and} \quad c_2 \approx \begin{cases} \frac{m_2^{TOT}}{K_2} R & , R \leq K_2 \\ m_2^{TOT} & , R > K_2 \end{cases} \tag{D.3}$$

Since $R^{TOT} = R + c_1 + c_2$:

$$R^{TOT}(R) = \begin{cases} \left(1 + \frac{m_1^{TOT}}{K_1} + \frac{m_2^{TOT}}{K_2}\right) R & , 0 \leq R \leq K_1 \\ \left(1 + \frac{m_2^{TOT}}{K_2}\right) R + m_1^{TOT} & , K_1 < R \leq K_2 \\ R + m_1^{TOT} + m_2^{TOT} & , R > K_2 \end{cases} \tag{D.4}$$

Now it is possible to reverse the equations to obtain R as function of R^{TOT}:

$$R(R^{TOT}) = \begin{cases} \frac{1}{\left(1 + \frac{m_1^{TOT}}{K_1} + \frac{m_2^{TOT}}{K_2}\right)} R^{TOT} & , 0 \leq R^{TOT} \leq \tilde{R}_1^{TOT} \\ \frac{R^{TOT} - m_1^{TOT}}{\left(1 + \frac{m_2^{TOT}}{K_2}\right)} & , \tilde{RTOT} < R^{TOT} \leq \tilde{R}_2^{TOT} \\ R^{TOT} - m_1^{TOT} - m_2^{TOT} & , R^{TOT} > \tilde{R}_2^{TOT} \end{cases} \tag{D.5}$$

where accordingly to the conditions in the D.4 system:

$$\tilde{R}_1^{TOT} = \left(1 + \frac{m_1^{TOT}}{K_1} + \frac{m_2^{TOT}}{K_2}\right) K_1 \quad \text{and} \quad \tilde{R}_2^{TOT} = \left(1 + \frac{m_2^{TOT}}{K_2}\right) K_2 + m_1^{TOT} \tag{D.6}$$

Thus it is straightforward to rewrite the complexes c_1 and c_2 as function of R^{TOT}:

$$c_1(R^{TOT}) = \begin{cases} \frac{m_1^{TOT}}{K_1} \frac{1}{\left(1 + \frac{m_1^{TOT}}{K_1} + \frac{m_2^{TOT}}{K_2}\right)} R^{TOT} & , 0 \leq R^{TOT} \leq \tilde{R}_1^{TOT} \\ m_1^{TOT} & , R^{TOT} > \tilde{R}_1^{TOT} \end{cases} \tag{D.7}$$

and

$$c_2(R^{TOT}) = \begin{cases} \frac{m_2^{TOT}}{K_2} \frac{1}{\left(1 + \frac{m_1^{TOT}}{K_1} + \frac{m_2^{TOT}}{K_2}\right)} R^{TOT} & , 0 \leq R^{TOT} \leq \tilde{R}_1^{TOT} \\ \frac{m_2^{TOT}}{K_2} \frac{R^{TOT} - m_1^{TOT}}{\left(1 + \frac{m_2^{TOT}}{K_2}\right)} & , R^{\tilde{TOT}}{}_1 < R^{TOT} \leq \tilde{R}_2^{TOT} \\ m_2^{TOT} & , R^{TOT} > \tilde{R}_2^{TOT} \end{cases} \tag{D.8}$$

From D.7 and D.8 the graph of the expected competition between species can be derived.

In Figure D.2, it is clearly shown that the breaking points are the thresholds $\tilde{R}_1^{TOT}$ and $\tilde{R}_2^{TOT}$, which consequently means that the competition depends on m^{TOT} and K (as it was declared at the beginning).
At the same time, m^{TOT} depends on the transcription rate ω:

$$\dot{m}^{TOT} \approx \omega D - \gamma m^{TOT} = 0 \longrightarrow m_{ss}^{TOT} \approx \frac{\omega D}{\gamma} \tag{D.9}$$

In conclusion, it is trivial to extend the reasoning to n competitive ribosomal species.
The complex with the lowest K will have just one breaking point as in the analyzed case; the complex with the highest K will have n breaking points and its evolution will be strongly influenced by all the other complexes.

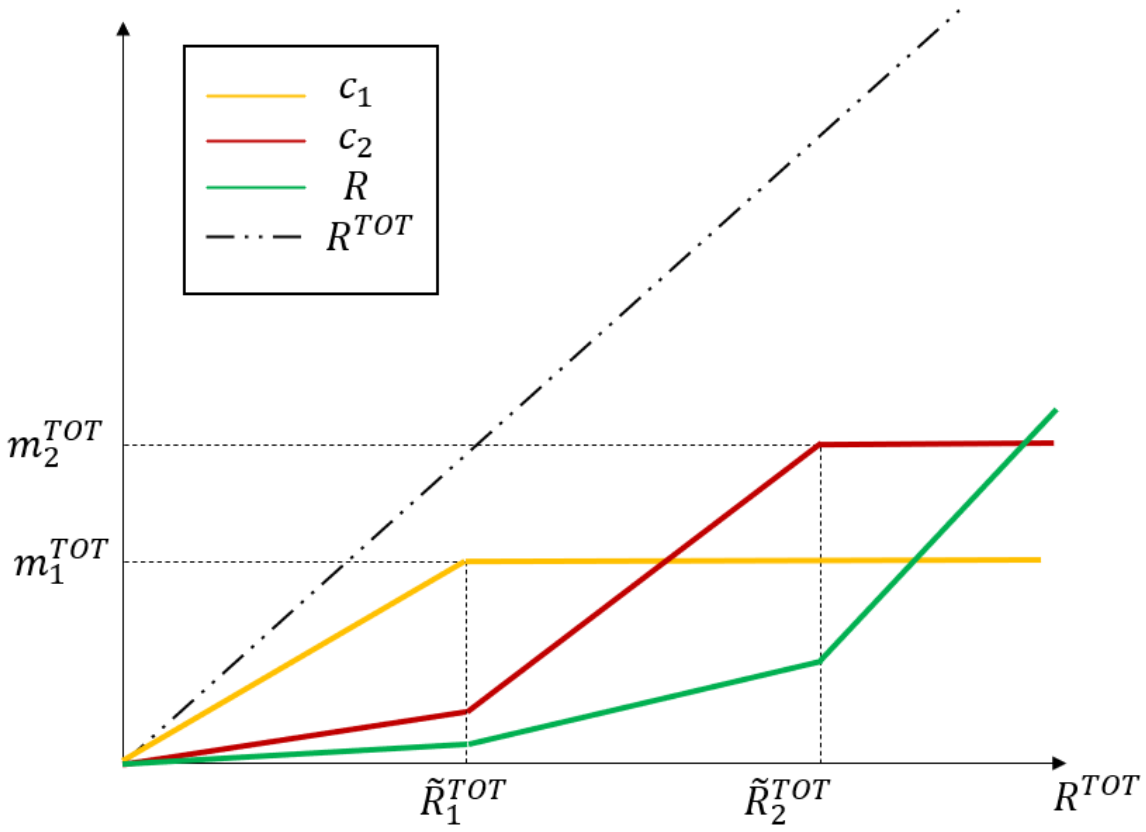

Figure D.2: Expected graph of general competition between two species.

Appendix E

Approximation-simulation differences

In Appendix D an approximation has been made, namely:

$$c_i = \frac{\frac{R}{K_i}}{1+\frac{R}{K_i}} m_i^{TOT} \approx \begin{cases} \frac{m_i^{TOT}}{K_i} R & , R \leq K_i \\ m_i^{TOT} & , R > K_i \end{cases} \tag{E.1}$$

Without loss of generality, assume $m_i^{TOT} = 1$ and $K_i = 1$; then the difference between the approximation and the real Michaelis Menten has been investigated.

Consider a generic approximation $z(x)$:

$$z(x) = \begin{cases} \frac{x}{\alpha} & 0 < x \leq \alpha \\ 1 & x > \alpha \end{cases} \tag{E.2}$$

that has to be compared with the real function $y(x)$:

$$y(x) = \frac{x}{1+x} \tag{E.3}$$

Thus the final aim to find α such that $|w(x,\alpha)| = |y(x) - z(x,\alpha)|$ is minimizes for all x, i.e.:

$$\min_{\alpha} \max_{x} |w(x,\alpha)| \tag{E.4}$$

The optimal choice of α corresponds to the case $w(\bar{x},\alpha^*) = -w(\alpha^*,\alpha^*)$.
$\bar{x}$ is the point where w reaches the maximum positive value:

$$\frac{dw}{dx} = 0 \longrightarrow \frac{1}{1+x} - \frac{x}{(1+x)^2} - \frac{1}{\alpha} = 0 \longrightarrow \bar{x} = \sqrt{\alpha} - 1 \tag{E.5}$$

Consequently:

$$w(\bar{x},\alpha) = \frac{\bar{x}}{\sqrt{\alpha}} - \frac{\bar{x}}{\alpha} = \bar{x}\Big(\frac{\sqrt{\alpha}-1}{\alpha}\Big) \tag{E.6}$$

and then by the substitution of $\bar{x}$ from E.5:

$$w(\bar{x},\alpha^*) = \frac{(\sqrt{\alpha}-1)^2}{\alpha^*} \tag{E.7}$$

At the same time:

$$-w(\alpha^*, \alpha^*) = 1 - \frac{\alpha^*}{1+\alpha^*} = \frac{1}{1+\alpha^*} \tag{E.8}$$

The intersection between E.7 and E.8 can not be found analytically.
Using *Matlab*, the optimal choice results to be:

$$\alpha^* \approx 3.5 \tag{E.9}$$

and finally the best approximation for the Michaelis Menten must be:

$$\hat{y}(x) = \begin{cases} \frac{x}{3.5} & 0 < x \leq 3.5 \\ 1 & x > 3.5 \end{cases} \tag{E.10}$$

Appendix F

Relation between growth rate and load

The analysis on general competition addressed in Appendix D sets the basis for the study of the growth of the cell, since it shows how the ribosomal complexes compete for ribosomes.
In particular, it has been proved that the linear growth rate dependent on R^{active} is the best choice considering the load.
According to this formulation, the growth rate has the form:

$$\mu = \varphi R^{active} \qquad \text{where } \varphi \text{ is the angular coefficient} \tag{F.1}$$

Recalling that

$$R^{active} = c_b + c_p + c_\ell = R^{TOT} - R \tag{F.2}$$

it is not straightforward to understand why the increase of the metabolic load (m_ℓ^{TOT}) should anyway lead to the drop of the growth rate μ.

First of all, it is worth to prove that the amount of free $rRNA$ is not dependent on any species of the system at the steady state, and consequently it is constant.
Consider the equation 3.41. It can be rewritten as follows:

$$\dot{r}^{TOT} = \beta_r D_r - \gamma(r^{TOT} - R^{TOT}) - \mu R^{TOT} \tag{F.3}$$

by using the mass conservation law of r^{TOT}.
Additionally, an approximation is possible (again taking into account the relation between γ and μ):

$$\dot{r}^{TOT} = \beta_r D_r + R^{TOT}(\gamma - \mu) - \gamma r^{TOT} \approx \beta_r D_r + \gamma R^{TOT} - \gamma r^{TOT} \tag{F.4}$$

The value of r^{TOT} at the steady state is then:

$$r_{ss}^{TOT} \approx R_{ss}^{TOT} + \frac{\beta_r D_r}{\gamma} \tag{F.5}$$

but recalling that the mass conservation law sets $r^{TOT} = R^{TOT} + r$, it can be immediately deduced that:

$$r_{ss} = \frac{\beta_r D_r}{\gamma} \tag{F.6}$$

which just depends on the value of the parameters.

The following analysis is limited to a specific region $R^{TOT} < \tilde{R}_1^{TOT}$ [1].

[1] $\tilde{R}_1^{TOT}$ has been already calculated in Appendix D.3.

In this zone, all the complexes are growing linearly with an angular coefficient that depends on the association constant K_i and the total amount of *mRNA* m_i^{TOT}:

$$c_i(R^{TOT}) = \chi_i R^{TOT} \quad \text{where} \quad \chi_i = \frac{m_i^{TOT}}{K_i} \frac{1}{1 + \sum_i \frac{m_i^{TOT}}{K_i}} \qquad i = b, p, \ell \tag{F.7}$$

Consequently, F.2 can be rewritten as function of χ_i ($i = b, p, \ell$):

$$R^{active} = (\chi_b + \chi_p + \chi_\ell) R^{TOT} = \chi_R R^{TOT} \tag{F.8}$$

where

$$\chi_R = \frac{\sum_i \frac{m_i^{TOT}}{K_i}}{1 + \sum_i \frac{m_i^{TOT}}{K_i}} \qquad i = b, p, \ell \tag{F.9}$$

Since $m_i^{TOT} >> K_i$, then $\chi_R \approx 1$, which means that in this region the concentration of free ribosomes is low. This is reasonable, since all the complexes are growing and the needs of ribosomes is high, namely almost all the free ribosomes are coupled to *mRNA*.

The equations 3.42 and 3.39 can also be rewritten as it follows:

$$\dot{P}^{TOT} = \beta_p \chi_p R^{TOT} - \varphi \chi_R R^{TOT} P^{TOT} \tag{F.10}$$

$$\dot{R}^{TOT} = \sigma_p r (P^{TOT} - R^{TOT}) - \varphi \chi_R (R^{TOT})^2 \tag{F.11}$$

At the equilibrium

$$P^{TOT} = \frac{\beta_p \chi_p}{\varphi \chi_R} \tag{F.12}$$

$$P^{TOT} = R^{TOT} \Big(1 + \frac{\varphi \chi_R}{\sigma_p r} R^{TOT} \Big) \tag{F.13}$$

which is constant, independent of R^{TOT}.
Coupling together these equations, an expression function of $R^{active} = \chi_R R^{TOT}$ (as in F.8) is found:

$$\frac{\beta_p}{\varphi} \chi_p = R^{active} \Big(1 + \frac{\varphi}{\sigma_p r} R^{active} \Big) \tag{F.14}$$

In F.14, only χ_p and R^{active} are free to evolve.
In particular, while the second term is independent of the load m_ℓ^{TOT}, χ_p is influenced by it according to F.9. More precisely, if m_ℓ^{TOT} grows, χ_p decreases.
Furthermore, in line with F.14, R^{active} decreases as well, but for equation F.1, the growth rate must drop.
Finally, this proves the initial statement.

Unfortunately, this proof is just limited to values of R^{TOT} lower than $\tilde{R}_1^{TOT}$.
Moreover, this region is very narrow since the equation of $\tilde{R}_1^{TOT}$ is [2]:

$$\tilde{R}_1^{TOT} = \Big(1 + \frac{m_b^{TOT}}{K_b} + \frac{m_p^{TOT}}{K_p} + \frac{m_\ell^{TOT}}{K_\ell} \Big) K_b \tag{F.15}$$

[2]Under the assumption that the relation between the dissociation constants is $K_b < K_p, K_\ell$

where K_b represents an important factor of reduction.
For a better analysis, one should observe the relation between R^{active} and the load in every possible region (in a system with three ribosomal species there are four regions).

Additionally, another possible issue can be found in this proof.
In F.10, μ has been substituted with $\varphi R^{active} = \varphi \chi_R R^{TOT}$ but this is an approximation. Indeed $\mu = \delta + \lambda$ where λ is the true growth rate. The performed substitution is based on the assumption that $\delta << \lambda$.
However, this could not be true, especially in the small considered region.

Finally, the proposed analysis has been made just for the second model. Nevertheless, it is possible to extend it also to the third case, with harder calculations.

Appendix G

Matlab code

G.1 *fsolve* implementation

The study of the equilibrium points have been performed by using the method *fsolve* by *Matlab*, which solves systems of nonlinear equations.
The problem must be specified as:

$$F(x) = 0 \tag{G.1}$$

where $F(x)$ is a function that returns a vector value. Here the implementation of the method considering the ODE system of $M2$ will show[1].
First, F must be defined as follows[2]:

```
function F = equilibria(x,omega,beta_Dr,beta_p,sigma_p,K,p)
% params:
% omega = [omega_Db, omega_Dl, omega_Dp] transcription rates
% K     = [Kb, Kl, Kp] dissociation constants
% beta_Dr - transcription rate of rRNA
% beta_p  - translation rate of R-Protein
% sigma_p - synthesis rate of ribosomes
% p       - parameters of the growth rate function (gr)

% x = [cb cl cp R r P Rtot]

gamma_mRNA = 100;

% constant quantities
mb_tot = omega(1) / gamma_mRNA;
ml_tot = omega(2) / gamma_mRNA;
mp_tot = omega(3) / gamma_mRNA;

% eqns of complexes
F(1) = (x(4)/(x(4)+K(1)))*mb_tot - x(1); % eqn of c_b
F(2) = (x(4)/(x(4)+K(2)))*ml_tot - x(2); % eqn of c_l
F(3) = (x(4)/(x(4)+K(3)))*mp_tot - x(3); % eqn of c_p
% eqn of R_tot
F(4) = sigma_p*x(5)*x(6) - gr(x(1),p)*(x(1)+x(2)+x(3)+x(4));
```

[1]For the analysis of $M1$ and $M3$, the only changes must be referred to the following function F.
[2]The function gr must be defined outside the function F and it computes the growth rate.

```
F(5) = beta_Dr - sigma_p*x(5)*x(6)-gamma_mRNA*x(5); % eqn of r
F(6) = beta_p*x(3) - sigma_p*x(5)*x(6) - gr(x(1),p)*x(6); % eqn of P

end
```

where $F(i)$ is the *i-th* differential equation that has been set equal to 0 (i.e., at the equilibrium). Since the easiest implementation of the method requires to define F and x_0, the starting point is:

```
x0 = [R_tot_max*0.3; 0; R_tot_max*0.3; R_tot_max*0.3; R_tot_max; R_tot_max];
```

It is worth to recall that the analysis of the equilibria has been made for different values of the transcription rate of the load. This implies that the method *fsolve* must be involved in a *for* cycle, where ω_ℓ is modified at every loop.

```
n = 100; % number of samples
omega_Dl_vector = linspace(0,max_threshold,n);
results = zeros(6,n);
% starting point where x = [cb cl cp R r P Rtot]
x0 = [R_tot_max*0.3; 0; R_tot_max*0.3; R_tot_max*0.3; R_tot_max; R_tot_max];

for i=1:n
    omega = [omega_Db omega_Dl_vector(i) omega_Dp];
    fun = @(x)equilibria(x,omega,beta_Dr,beta_p,sigma_p,K,p);
    sol = fsolve(fun,x0);
    results(:,i) = sol.';
    x0 = sol;
end
```

In order to avoid numerical problems, the iteration on the starting point x_0 has been revealed as the best solution. Indeed, fixing a unique point can lead the method to find an equilibrium far away from the previous one (the ODE system can have several equilibrium points for the same parameters values).

G.2 *SimBiology* Toolbox

Matlab has many Toolboxes which aim at helping the analysis. For what concerns systems biology, the *SimBiology Builder Toolbox* can be used as powerful tool for the design of the cell network, since it makes it extremely easy.
It is simply based on the mechanism of drag-and-drop different elements (based on their function) on a white board and afterwards connect them in order to establish the relations between the various components of the cell, creating a graph (Figure G.1).
In particular as important feature, it allows to define the units of measurement of the different components, computing an additional check on them for making sure to obtain a coherent result. It also derives the system of differential equations directly from the graph.
The connection between two components must be specified, namely the reaction that involves them must be declared in the *Property Editor* field. Specifically, the *Kinetic Law* must be chosen between the default options. However, the degradation is not present in

the default list provided by *SimBiology*. Despite that, it is possible to define a new kinetic law in *Matlab* that will be imported in the toolbox with the method *sbioabstractkineticlaw*. It requires to insert the name of the new kinetic law and the equation that rules it. Then it must be specified the elements in the equation that are variables or parameters. Finally, the built rule must be added in the library of the toolbox.

```
% Create a kinetic law definition.
abstkineticlawObj = sbioabstractkineticlaw('growth_rate', '(p*(cb+cp+cl) +
gamma_protein)*element');
% Assign the parameter and species variables in the expression.
set (abstkineticlawObj, 'SpeciesVariables', {'cb','cp','cl','element'});
set (abstkineticlawObj, 'ParameterVariables', {'gamma_protein','p'});
% Add the new kinetic law definition to the user-defined library.
sbioaddtolibrary(abstkineticlawObj);
```

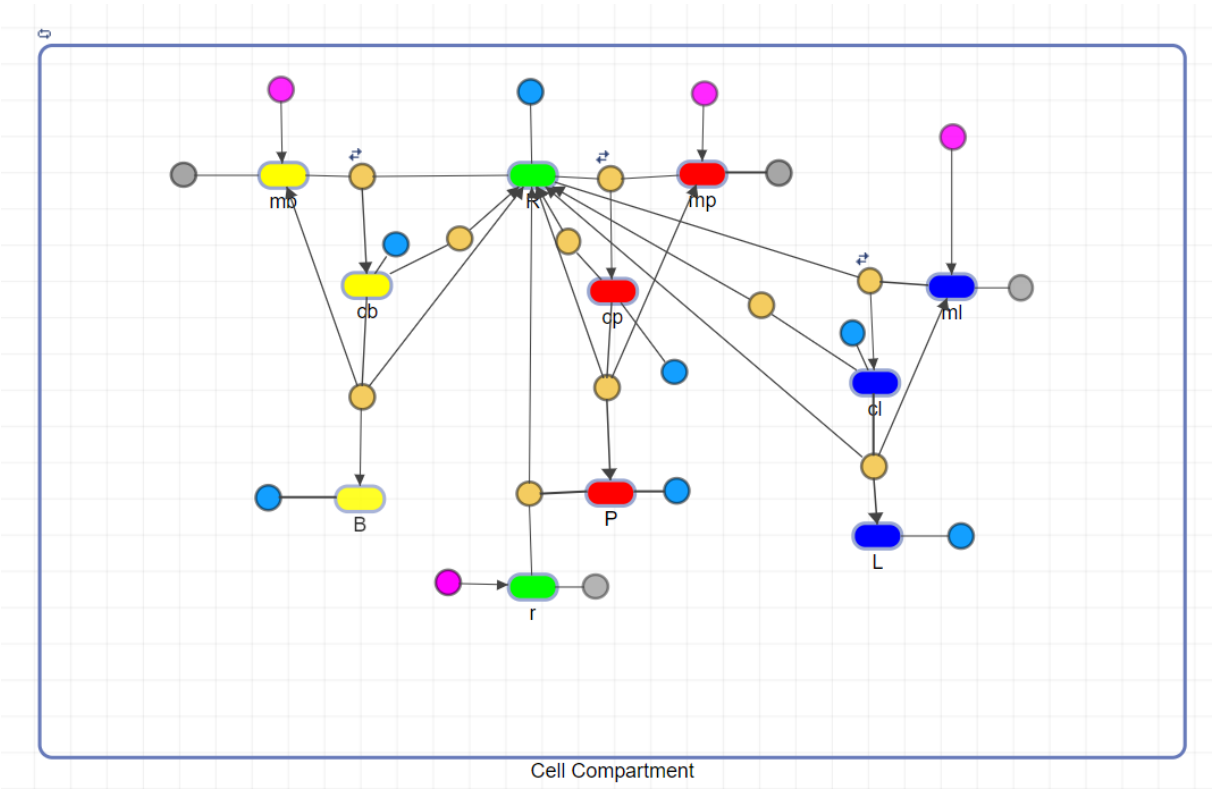

Figure G.1: Graph made with *SimBiology Builder Toolbox* of model $M2$. The yellow, red and blue elements represent the basal, R-Protein and load species. The green components stand for the ribosomes compound (both *rRNA* and free ribosomes). The pink, grey and blue dots represent the transcription process, the degradation with rate γ and the degradation with rate μ respectively. The dark yellow dots represent the reaction that involves the two connected components.

Once the model of the cell system has been defined with *SimBiology Model Builder*, it can be analyzed with another tool, called *SimBiology Model Analyzer*.
It has several features that can be explored, but the most suitable for this thesis is the simulation of the model, that allows to study the dynamics of the system, varying pa-

rameters or initial conditions. For the choosing this type of simulation: *Program* $\longrightarrow$ *Simulate Model*. Afterwards, if several models have been designed and saved in the same project, select the desired one in the field *Model*. It is also possible to set the duration of the simulation by modifying the *Stop Time*.
Additionally, it is possible to change parameters values during the simulation, or its initial conditions. Before running it, drag-and-drop the desired components from the *Browser (Model)* to *Explorer* section. If one wants to visualize all the simulations performed and save the data from every run, tick *Keep results from each run*. Now the simulation can be run, simply using the green button *Run*.
Finally, all the simulations can be imported in the *Workspace* as matfile. From the *SimBiology Model Analyzer*, section *Browser (Project)*, select the running program, *LastRun*, and then right-click on *results* and select *Export Data to MATLAB Workspace*: the data will be saved and ready to be analyzed.