Resources allocation explains the differential roles of RBS and promoter strengths in cell mass distribution and optimal protein expression productivity

Fernando Nóbel[a], Jesús Picó[a,1]

[a] *Synthetic Biology and Biosystems Control Lab, I.U. de Automática e Informática Industrial (ai2)*
Universitat Politècnica de València,
Camí de Vera S/N, 46022, València, Spain.

Abstract

Design of synthetic genetic circuits without considering the impact of host–circuit interactions results in an inefficient design process and lengthy trial-and-error iterations to appropriately tune the expression levels. Microorganisms have evolved to reach an optimal use of cellular resources. This balance is perturbed by circuit-host interactions resulting from the interaction among the cell environment from which the cell takes substrates, its metabolism, and the needs of exogenous synthetic genetic circuit introduced in the cell host. The resulting competition for common shared cell resources introduces spurious dynamics leading to problems of malfunctioning of the synthetic circuit due to lack of enough cellular resources. Therefore, there is an increasing interest in development of methods for model-based design of synthetic gene circuits considering host-circuit interactions. Here we present a small-size model of gene expression dynamics in bacterial cells accounting for host-circuit interactions. For each gene, the model defines the cellular resources recruitment strength as the key functional coefficient that allows to explain the distribution of resources among the host and the genes of interest and the relationship between the usage of resources and cell growth. This functional coefficient explicitly takes into account the availability of resources and lab-accessible gene characteristics, such as promoter and ribosome binding site (RBS) strengths and capture their interplay with the availability of free cell resources. In spite of its simplicity, the model is able to explain the differential role of promoter and RBS strengths in the distribution of protein mass and the optimal protein expression productivity with remarkable fit to the experimental data from the literature for *E. coli*. This makes the model amenable for model-based circuit design purposes. Moreover, the model also allows to understand why endogenous ribosomal and non-ribosomal genes have evolved different strategies in the expression space.

Keywords: Synthetic biology, Protein expression, Growth rate, Metabolic load,

*Corresponding author: J. Picó
Email address: jpico@ai2.upv.es (Jesús Picó)

November 19, 2020

Contents

1 Introduction **3**

2 Modelling protein expression **4**

3 Dynamics of the number of mature available ribosomes **11**

4 Relating free and available ribosomes **12**

5 Protein expression dynamics as a function of the unbound fraction of available mature ribosomes **13**

6 Obtaining the cell specific growth rate **14**

7 Cell specific growth rate and population dynamics **16**

8 Relationship between growth rate and cell mass **19**

9 The relative resources recruitment strength provides the relative mass fraction at steady state. **21**

10 Average host dynamics and steady state. **22**

11 Productivity of exogenous proteins depends on host-circuit interaction **24**

12 Model parameters **25**

13 Estimation of the fraction Φ_b **25**

14 Estimation of the average resources recruitment strength. **29**

15 Estimation of the number of free ribosomes in the cell **31**

16 Sensitivity of the resources recruitment strength to RBS and promoter can explain the variation of the cell mass distribution with growth rate **36**

17 The differential role of RBS and promoter strengths. **41**

18 Host-circuit interaction shapes the optimal productivity of proteins **45**

19 Conclusions **46**

20 Acknowledgements **49**

21 Authors contributions **49**

1. Introduction

The rational design of synthetic genetic circuits of increasing complexity, their experimental characterization and robust tuning, and their industrial scaling are some important milestones towards the integration of complex synthetic genetic circuits into microorganisms for a variety of practical applications. Within this context, one of the fundamental problems in designing synthetic genetic circuits that explain the current disparity between the ability to design biological systems and to synthesize them is the lack of systematic design methods considering the interaction with the host cell [7].

Circuit-host interactions result from the interrelations among the cell environment from which the cell takes substrates, its metabolism, and the needs of the synthetic genetic circuit introduced in the cell host. The over-expression of exogenous proteins by a genetically modified microorganism as well as the production of metabolites by the addition and/or modification of their metabolic pathways introduces a metabolic load that removes the microorganism from its natural state, which has evolved to reach an optimal use of resources [40]. The resulting competition for common shared cell resources affects cell growth and introduces spurious dynamics [31] leading to problems of malfunctioning of the synthetic circuit due to lack of enough cellular resources. It also triggers its elimination by evolutionary mechanisms trying to restore the natural optimal state [10], or both.

Design of synthetic genetic circuits without considering the impact of host–circuit interactions results in an inefficient design process and lengthy trial-and-error iterations to appropriately tune a circuit's expression levels [43]. Therefore, in the last years the has been an increasing interest in development of methods for model-based design of synthetic gene circuits considering host-circuit interactions [33].

One strategy is to consider the shared resources as a variable source, without considering the host behavior. This approach has proved very useful to deal with the so-called retroactivity [21], the loading interaction among circuit modules and host originated from mass exchange. Retroactivity poses problems when predicting the behavior of a large network from that of the composing modules. It is a problem analogous to modeling the coupling between electrical circuits connected to a real energy source. Thus, the models accounting for it somewhat resemble Ohm's law [6, 31]. As these models do not explicitly consider the host behavior they cannot be easily used within a multiscale framework integrating the synthetic circuits of interest, the host, and the cell environment at the macroscopic level.

Alternatively, one may develop models relating substrates uptake, cell growth rate, and availability of free resources as a function of the gene circuits demand. These range from very coarse-grain ones [35, 16, 4] to semi-mechanistic ones with varied degrees of granularity [43, 18, 26]. In this last case, the interplay between circuit, host and environment can be directly incorporated into the circuit model of interest to capture the impact of cellular trade-offs and resource competition on the circuit function.

Construction of a large-scale mechanistic model of *E. coli* in [26] enabled the authors to integrate and cross-evaluate a massive, heterogeneous dataset integrating measurements reported by various groups over decades. On the other hand, medium-size detailed mechanistic models like the one developed in [43] have been used to study behavioral modulations of a gene switch [2] or a feedforward circuit [3, 15]. These medium and large-scale models, though very useful, are most often over-parametrized and cannot easily be integrated within a user-friendly lite and agile computational framework for model-based circuit design.

The main goal of the small-size model presented here is to provide enough granularity so as the model to provide good predictions of the dynamics of the host cell, the expression of the genes of interest and their interactions while having a small number of differential equations and parameters. An additional goal is to provide a model amenable for model-based circuit design purposes. To this end, the model considers explicitly lab-accessible gene characteristics such as promoter and RBS strengths and degradation rates of mRNA and proteins.

The rest of the paper is organized as follows. In Section 2 we derive the expression for the dynamics of protein expression considering the process of parallel translation of the same transcript by several ribosomes the resources recruitment strength functional coefficient. In Sections 3 and 4 we derive the dynamics of the number of mature available ribosomes and its relationship with the number of free ribosomes. This allows to derive the dynamics of protein expression as a function of the unbound fraction of available mature ribosomes In Section 5. These results are used in section 7 to derive an expression of the cell growth rate as a function of the number of ribosomes actively translating. The relationships between cell and population growth rates and between growth rate and cell mass are considered in sections 7 and 8 respectively. This first part of the paper finishes with the derivation of the equivalence between the relative resources recruitment strength and the relative mass fraction of a given protein at steady state, the average host dynamics and steady state and the evaluation of protein mass productivity in sections 9, 10 and 11 respectively. Next, we use experimental data from the literature and apply the previous developments to *E. coli*. We estimate the fraction of ribosomes being used in translating complexes relative to the mature available ones in section 13 and the average resources recruitment strength for both ribosomal and non-ribosomal *E. coli* proteins in section 14. These allow to estimate the number of free ribosomes that explain the experimental translation efficiencies per mRNA in section 15. Finally, in the last part of the paper we show how the sensitivity of the resources recruitment strength to RBS and promoter can explain the variation of the cell mass distribution with growth rate and the differential roles of RBS and promoter strengths in sections 16 and 17, and in section 18 we show how host-circuit interaction shapes the optimal productivity of both endogenous and exogenous proteins in the expression space.

2. Modelling protein expression

In our model, we consider a set of basic assumptions:

1. Transcription dynamics is fast enough as compared to translation so it can be considered at quasi-steady state (QSS).
2. The main resources-dependent process in protein expression is translation. Therefore:
   (a) only ribosomes are considered as limiting shared resource required for protein expression. RNA polymerase is not considered explicitly.
   (b) the effective translation rate is assumed to depend on the availability of intracellular substrate. This is implicitly considering that building the polypeptide protein chain is the limiting energy-consuming process in the cell. We do not explicitly consider the catabolic conversion of substrate into aminoacid building blocks.
3. Several ribosomes may translate a single messenger RNA (mRNA) simultaneously.
4. We identify a transcriptional unit by its promoter.

With these assumptions in mind, we model the expression of a given protein p_k by means of the set of pseudo-reactions 1.

$$
\begin{aligned}
\mathrm{g_k} &\xrightarrow{\omega_k(T_f)} \mathrm{g_k} + \mathrm{m_k} \\
\mathrm{m_k} + \mathrm{r} &\underset{K_u^k}{\overset{K_b^k}{\rightleftharpoons}} \mathrm{C_{k^0}} \\
\mathrm{C_{k^0}} &\xrightarrow{K_e(s_i)} \mathrm{C_{k^1}} + \mathrm{m_k} \\
\mathrm{C_{k^1}} &\xrightarrow{K_k(s_i)} \mathrm{p_k} + \mathrm{r} \\
\mathrm{C_{k^1}} &\xrightarrow{K_e(s_i)} \mathrm{C_{k^2}} \\
&\vdots \\
\mathrm{C_{k^{j-1}}} &\xrightarrow{K_k(s_i)} \mathrm{p_k} + \mathrm{r} \\
\mathrm{C_{k^{j-1}}} &\xrightarrow{K_e(s_i)} \mathrm{C_{k^j}} \\
&\vdots \\
\mathrm{C_{k^{n_r-1}}} &\xrightarrow{K_e(s_i)} \mathrm{C_{k^{n_r}}} \\
\mathrm{C_{k^{n_r}}} &\xrightarrow{K_k(s_i)} \mathrm{p_k} + \mathrm{r} \\
\mathrm{m_k} &\xrightarrow{d_{mk}+\mu} \emptyset \\
\mathrm{p_k} &\xrightarrow{d_k+\mu} \emptyset \\
\mathrm{C_{k^j\, j=0\ldots n_r}} &\xrightarrow{\mu} \emptyset
\end{aligned}
\tag{1}
$$

where the species involved are:

g_k : free copy of the promoter, ie. of the gene transcriptional unit.

m_k : free ribosome binding site (RBS), ie. mRNA copy with its RBS free.

r : free ribosome.

C_{k^0} : complex formed by a ribosome bound to the RBS in a mRNA

C_{k^j} : j-th translating complex formed by j ribosomes simultaneously translating a mRNA copy with freed RBS.

p_k : protein copy number

s_i : intracellular substrate (molecules)

and the specific reaction rates stand for:

$\omega_k(T_f)$: transcription rate. It may be a function of one or several transcription factors (TF) and may include the gene copy number.

K_b^k : association rate between a free ribosome and the RBS.

K_u^k : dissociation rate between a free ribosome and the RBS.

$K_e(s_i)$: translation initiation rate.

$K_k(s_i)$: translation elongation rate.

d_{mk} : mRNA degradation rate.

d_k : protein degradation rate.

μ : cell specific growth rate.

For a protein of length l_{pk} aminoacids, we denote:

$$l_{pk} = l_e + l_{xk} \tag{2}$$

where $1/l_e$ is the ribosomes density, with l_e expressed as equivalent number of codons. We consider the effective RBS length to be the same as l_e. The remaining length of the protein is denoted as l_{xk}. Thus, up to n_r ribosomes can simultaneously be translating a single copy of mRNA, with $n_r = l_{xk}/l_e$, and an additional ribosome is bound to the RBS.

We consider the effective rates $K_e(s_i)$ and $K_k(s_i)$ at which the ribosome glides through the RBS and the remaining mRNA nucleotides respectively. Thus, we consider:

$$K_e(s_i) = \frac{\nu_t(s_i)}{l_e} \ , \ K_k(s_i) = \frac{\nu_t(s_i)}{l_{xk}} \tag{3}$$

with

$$\nu_t(s_i) = \frac{\nu s_i}{K_{sc} + s_i} \tag{4}$$

where ν is the maximum translation rate and K_{sc} is a Michaelis-Menten parameter related to the cell substrate uptake and catabolic capacity. As a first approximation, we consider that ν is organism dependent but does not depend on the sequence of nucleotides, and K_{sc} is organism and substrate dependent but does not depend on the nucleotides sequence either.

The translating complexes $C_{k^{j-1}}$ are pseudo-species modeling the process of parallel translation. They can loosely be identified with each of the chains of aminoacids under formation. The first one, C_{k^0}, represents the ribosome bound to the RBS. Notice with rate $K_e(s_i)$ the ribosome bound to the RBS, forming C_{k^0}, advances to the next ribosome occupancy slot, generating the translating complex C_{k^1} and freeing the RBS so a new ribosome can enter in the cue. In turn, the displacement of the ribosomes in the cue by one occupancy generates the translating complexes C_{k^j} from the previous complex $C_{k^{j-1}}$. Finally, each parallel translating complex C_{k^j} generates a protein with rate $K_k(s_i)$ and frees its bound ribosome. Recall the translating complexes can be identified with each of the chains of aminoacids under formation. This way, we decouple the cue dynamics of the ribosomes advancing along the mRNA from the protein building ones, thus getting a continuous approximation of the parallel process of translation.

Next, we apply mass action kinetics to obtain the dynamic balances for the copy number of each species in the model. This way, we have:

$$\begin{aligned}
\dot{m}_k &= \omega_k(T_f) - K_b^k m_k r + K_u^k C_{k^0} + K_e(s_i)C_{k^0} - (d_{mk} + \mu)m_k \\
\dot{C}_{k^0} &= K_b^k m_k r - K_u^k C_{k^0} - K_e(s_i)C_{k^0} - \mu C_{k^0} \\
\dot{C}_{k^1} &= K_e(s_i)C_{k^0} - K_e(s_i)C_{k^1} - K_k(s_i)C_{k^1} - \mu C_{k^1} \\
&\vdots \\
\dot{C}_{k^{j-1}} &= K_e(s_i)C_{k^{j-2}} - K_e(s_i)C_{k^{j-1}} - K_k(s_i)C_{k^{j-1}} - \mu C_{k^{j-1}} \\
&\vdots \\
\dot{C}_{k^{n_r}} &= K_e(s_i)C_{k^{n_r-1}} - K_e(s_i)C_{k^{n_r}} - K_k(s_i)C_{k^{n_r}} - \mu C_{k^{n_r}} \\
\dot{p}_k &= K_k(s_i)\sum_{j=1}^{n_r} C_{k^j} - (d_k + \mu)p_k
\end{aligned} \tag{5}$$

Recall we assume that transcription dynamics is fast enough as compared to translation so it can be considered at quasi-steady state (QSS). We also assume the binding-unbinding dynamics to form the translation complexes C_{k^j} are fast enough so that we can also consider the copy number of each of the complexes quickly reaches steady state. Therefore, from $\dot{m}_k = 0$ and $\dot{C}_{k^j, j=1\ldots n_r} = 0$ we get:

$$C_{k^0} = \omega_k(T_f)\frac{1}{\frac{d_{mk}+\mu}{\frac{K_b^k}{K_u^k + K_e(s_i) + \mu}} + \mu r} r \tag{6}$$

and

$$C_{k^j} = \frac{K_e(s_i)}{K_e(s_i) + K_k(s_i) + \mu} C_{k^{j-1}} \ , \ j = 1 \ldots n_r \tag{7}$$

In practice, $d_{mk} \gg \mu$ and $k_u^k \gg \mu$ (see Table 12). Therefore the magnitude of the specific growth rate μ can be neglected with respect to both the mRNA degradation rate d_{mk} and the sums $K_u^k + K_e(s_i)$ and $K_e(s_i) + K_k(s_i)$ respectively so that we can approximate:

$$C_{k^0} \approx \omega_k(T_f) \frac{1}{\frac{d_{mk}}{K_{C^0}^k} + \mu r} r \tag{8}$$

and

$$C_{k^j} \approx \frac{K_e(s_i)}{K_e(s_i) + K_k(s_i)} C_{k^{j-1}} \ , \ j = 1 \ldots n_r \tag{9}$$

where we have defined:

$$K_{C^0}^k(s_i) \triangleq \frac{K_b^k}{K_u^k + K_e(s_i)} \tag{10}$$

Notice $K_{C^0}^k$ is directly related to the RBS strength.

Now, the dynamics for the copy number of the protein p_k can be obtained from the equations (5) and (9) as:

$$\begin{aligned} \dot{p}_k &= K_k(s_i) \sum_{j=1}^{n_r} C_{k^j} - (d_k + \mu) p_k \\ &= K_k(s_i) a \left[1 + a + \ldots + a^{n_r - 1} \right] C_{k^0} - (d_k + \mu) p_k \end{aligned} \tag{11}$$

with $a \triangleq \frac{K_e(s_i)}{K_e(s_i) + K_k(s_i)}$.

Notice the geometric sum $1 + a + \ldots + a^{n_r - 1}$ converges, as $a < 1$, so that:

$$\begin{aligned} K_k(s_i) a \left[1 + a + \ldots + a^{n_r - 1} \right] &= K_k(s_i) a \frac{1 - a^{n_r}}{1 - a} \\ &= K_e(s_i) \left[1 - \left(\frac{K_e(s_i)}{K_e(s_i) + K_k(s_i)} \right)^{n_r} \right] \end{aligned} \tag{12}$$

Using the definitions (3), notice $a = \frac{l_{xk}}{l_e + l_{xk}} = \frac{l_{xk}}{l_{pk}}$ and we get the expression:

$$K_k(s_i) a \left[1 + a + \ldots + a^{n_r - 1} \right] = K_e(s_i) \left[1 - \left(\frac{l_{xk}}{l_{pk}} \right)^{\frac{l_{xk}}{l_e}} \right] \tag{13}$$

where we have taken into account that the maximum number of ribosomes bound to active translating complexes $C_{k^j, j=1 \ldots n_r}$ simultaneously translating a mRNA molecule is $n_r = l_{xk}/l_e$, and we assume this maximum is always reached (see section XX for a discussion on the validity of this approximation).

Recall we have assumed that transcription is not a limiting process, so we can express the effective transcription rate:

$$\omega_k(T_f) = \frac{\eta}{l_{pk}} F_k(T_f) \tag{14}$$

where η (aa/min) is the maximum transcription speed and $F_k(T_f)$ is the transcription characteristic function that may depend on one or several transcription factors. By default, we assume the gene copy number c_{nk} is one. If this is not the case, the effective transcription rate $\omega_k(T_f)$ must be multiplied by c_{nk}. Notice in this case the transcription characteristic function $F_k(T_f)$ may depend on the gene copy number as described in [41].

As commented in our preliminary assumptions, we do not model competition for RNA polymerases, which would also affect the effective transcription rate. Yet, notice that, even if there are no cognate transcription factors associated, the term $F_k(T_f)$ can be used to account for competition for RNA polymerases preventing the effective transcription from proceeding at its maximum rate $\frac{\eta}{l_{pk}}$, sequence-dependent affinity of the promoter for the RNA polymerases (promoter strength) and the effect of nucleotides usage on the transcription speed. In summary, the term $F_k(T_f)$ can be used to accommodate aspects affecting transcription so that $\omega_k(T_f)$ is the effective transcription rate.

Thus, the dynamics for protein expression become:

$$\begin{aligned}
\dot{p}_k &= K_e(s_i)\left[1-\left(\frac{l_{xk}}{l_{pk}}\right)^{\frac{l_{xk}}{l_e}}\right] C_{k^0} - (d_k+\mu)p_k \\
&= K_e(s_i)\left[1-\left(\frac{l_{xk}}{l_{pk}}\right)^{\frac{l_{xk}}{l_e}}\right] \frac{\eta}{l_{pk}} F_k(T_f) \frac{1}{\frac{d_{mk}}{K^k_{C^0}}+\mu r} r - (d_k+\mu)p_k \\
&= \frac{\nu}{l_e}\frac{s_i}{K_{sc}+s_i}\left[1-\left(\frac{l_{xk}}{l_{pk}}\right)^{\frac{l_{xk}}{l_e}}\right] \omega_k(T_f) \frac{1}{\frac{d_{mk}}{K^k_{C^0}}+\mu r} r - (d_k+\mu)p_k
\end{aligned} \tag{15}$$

We now define the risosomes density related term:

$$E_{mk}(l_{pk}, l_e) \triangleq \frac{l_{pk}}{l_e}\left[1-\left(\frac{l_{xk}}{l_{pk}}\right)^{\frac{l_{xk}}{l_e}}\right] = \frac{l_{pk}}{l_e}\left[1-\left(1-\frac{l_e}{l_{pk}}\right)^{\left(\frac{l_{pk}}{l_e}-1\right)}\right] \tag{16}$$

Figure 1 shows the values of $E_{mk}(l_{pk}, l_e)$ as a function of the protein length l_{pk} for different values of l_e. As, seen $E_{mk}(l_{pk}, l_e)$ can be accurately approximated as a linear function of l_{pk}:

$$E_{mk}(l_{pk}, l_e) \approx 0.62\frac{l_{pk}}{l_e} \tag{17}$$

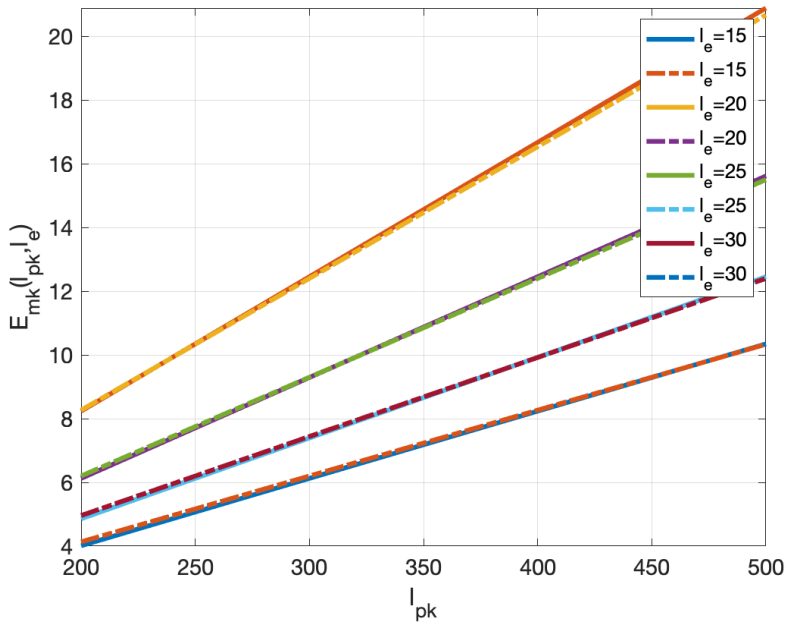

Figure 1: Function $E_{mk}(l_{pk}, l_e)$ as a function of the protein length l_{pk} for different values of l_e and their corresponding linear approximations $E_{mk}(l_{pk}, l_e) = 0.62\frac{l_{pk}}{l_e}$.

We further define the *resources recruitment strength* functional parameter $J_k(\mu, r)$:

$$J_k(\mu, r) \triangleq E_{mk}(l_{pk}, l_e)\omega_k(T_f)\frac{1}{\frac{d_{mk}}{K^k_{C^0}} + \mu r} \tag{18}$$

Notice the *resources recruitment strength* is a dimensionless function that expresses the capacity of the $k-$th gene to recruit cellular resources to get expressed. It explicitly depends on:

- the gene characteristics:
  - mRNA transcription and degradation rates $\omega_k(T_f)$ and d_{mk}
  - RBS strength-related parameter $K^k_{C^0}$
- and the availability of cell resources:
  - flux of free ribosomes μr
  - ribosomes density $\frac{l_{pk}}{l_e}$ (via $E_{mk}(l_{pk}, l_e)$)

Using these definitions in equation (15) we get the expressions for the copy number

dynamics of the $k-$th protein:

$$\begin{aligned}\dot{p}_k &= 0.62\frac{\nu_t(s_i)}{l_e}\frac{r}{\frac{d_{mk}}{K^k_{C^0}(s_i)}+\mu r}\omega_k(T_f) - (d_k+\mu)p_k \\ &= \frac{\nu_t(s_i)}{l_{pk}}J_k(\mu,r)r - (d_k+\mu)p_k \\ &= \frac{1}{\frac{1}{K_e(s_i)}+\frac{1}{K_k(s_i)}}J_k(\mu,r)r - (d_k+\mu)p_k\end{aligned} \tag{19}$$

Note: Using an electrical analogy, notice the term $\frac{1}{\frac{1}{K_e(s_i)}+\frac{1}{K_k(s_i)}}$ corresponds to the equivalent resistance of the circuit formed by the parallel resistances $K_e(s_i)$ and $K_k(s_i)$.

3. Dynamics of the number of mature available ribosomes

Ribosomes are large complexes formed by both ribosomal RNA molecules and a variety of ribosomal proteins, adding up to 55 proteins in *E. coli*. Recall we consider that translation is the main energy and resources limiting process we model. Notice the total copy number of ribosomes in the cell at any one time instant r_t is the sum of the mature (r_a) and inmature (r_i) ribosomes. In turn, the mature ribosomes r_a available for protein translation comprise the free ribosomes r and the ones already bound to translating complexes being used to build proteins. These comprise both the ribosomes r_r bound to translating complexes building the ribosomes themselves, and the ones r_p bound to the translating complexes of non-ribosomal proteins. These last will comprise those associated to the gene circuits of interest (r_{sys}), and the ones used by the rest of the transcriptional units being expressed in the cell. Under nominal conditions, these last represent the basal needs of the cell and hereafter we will refer to them as the wild-type (wt) bound amount of ribosomes (r_{wt}). Thus, we have $r_a = r + r_{\text{bound}}$, with $r_{\text{bound}} = r_r + r_p$ and $r_p = r_{\text{sys}} + r_{\text{wt}}$.

The copy number of available mature ribosomes is a fraction of the total number of ribosomes, so that $r_a = \Phi_t r_t$. The fraction Φ_t varies little in time, with an average value $\Phi_t = 0.8$ [4, 5] so that the dynamics of the total number of ribosomes and that of the number of available ribosomes are the same but for a scale factor. Next we consider the dynamics of the total copy number of ribosomes in the cell, r_t.

To get the dynamics of r_t we first consider an analogous expression to (19) for each of the proteins forming up a ribosome. If we consider the average ribosomal protein p_r, the total number of ribosomal proteins for a given average ribosome composed of N_r proteins (e.g. $N_r = 55$ for *E. coli*) is defined as $p_{\Sigma r} = N_r p_r$. For the average ribosomal

protein p_r we have the dynamics:

$$\begin{aligned}\dot{p}_r &= \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} E_{m,r}\omega_r \frac{1}{\frac{d_{mr}}{K^r_{C^0}}+\mu r} r - \mu p_r \\ &= \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} J_r(\mu, r) r - \mu p_r\end{aligned} \tag{20}$$

where we have assumed that ribosomal proteins are only subject to dilution caused by cell growth.

Then, the dynamics for the total number of ribosomal proteins can be approximated as:

$$\begin{aligned}\dot{p}_{\Sigma r} &= N_r \dot{p}_r \\ &= \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} N_r J_r(\mu, r) r - \mu p_{\Sigma r}\end{aligned} \tag{21}$$

Since we need all N_r proteins forming up an individual ribosome, and considering the average ribosomal protein p_r, the total copy number of ribosomes is $r_t = p_{\Sigma r}/N_r$. Therefore, the dynamics of the total copy number of ribosomes r_t will be the same as those of p_r. That is:

$$\begin{aligned}\dot{r}_t &= \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} E_{m,r}\omega_r \frac{1}{\frac{d_{mr}}{K^r_{C^0}}+\mu r} r - \mu r_t \\ &= \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} J_r(\mu, r) r - \mu r_t\end{aligned} \tag{22}$$

Therefore, the dynamics of the number of mature available ribosomes is:

$$\dot{r}_a = \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc}+s_i} J_r(\mu, r) \Phi_t r - \mu r_a \tag{23}$$

4. Relating free and available ribosomes

Recall we had $r_a = r + r_{\text{bound}}$. For each protein p_k of interest, and using the previous results, the number of ribosomes bound to complexes involved in its translation at each time instant is given by:

$$\begin{aligned}r_{p_k} = C_{k^0} + \sum_{j=1}^{n_r} C_{k^j} = C_{k^0} \frac{1-a^{n_r}}{1-a} &= \frac{1-\left(\frac{l_{xk}}{l_{pk}}\right)^{n_r}}{1-\frac{l_{xk}}{l_{pk}}} \omega_k(T_f) \frac{1}{\frac{d_{mk}}{K^k_{C^0}}+\mu r} r \\ &= E_{mk}(l_{pk}, l_e)\omega_k(T_f) \frac{1}{\frac{d_{mk}}{K^k_{C^0}}+\mu r} r \\ &= J_k(\mu, r) r\end{aligned} \tag{24}$$

An analogous expression can be obtained for the number r_r of ribosomes bound to

complexes involved in translation of ribosomes themselves:

$$r_r = N_r \left(C_{a^0} + \sum_{j=1}^{n_r} C_{a^j} \right) = N_r E_{mr}(l_{p_r}, l_e) \frac{\eta}{l_{p_r}} \frac{1}{\frac{d_{mr}}{K^r_{C^0}} + \mu r} r = N_r J_r(\mu, r) r \tag{25}$$

where we have taken into account that it requires N_r proteins to build-up a ribosome.

Therefore, the copy number of mature available ribosomes r_a can be obtained from:

$$\begin{aligned} r_a &= r + N_r \left(C_{a^0} + \sum_{j=1}^{n_r} C_{a^j} \right) + \sum_{k=1}^{n_p} \left[C_{k^0} + \sum_{j=1}^{n_r} C_{k^j} \right] \\ &= r + N_r J_r(\mu, r) r + \sum_{k=1}^{n_p} J_k(\mu, r) r \end{aligned} \tag{26}$$

where $n_p = n_{\text{sys}} + n_{\text{wt}}$ is the sum of the quantities of basal wild-type and system-of-interest non-ribosomal proteins in the cell.

From equation (26) we get the number of free ribosomes r as a function of the mature available ones r_a:

$$r = \frac{r_a}{1 + N_r J_r(\mu, r) + \sum_{k=1}^{n_p} J_k(\mu, r)} \tag{27}$$

5. Protein expression dynamics as a function of the unbound fraction of available mature ribosomes

We can define:

$$\Phi_b \triangleq \frac{\sum_{j=p,r} J_j(\mu, r)}{1 + \sum_{j=p,r} J_j(\mu, r)} = \Phi_r + \Phi_p \tag{28}$$

with

$$\Phi_r \triangleq \frac{N_r J_r(\mu, r)}{1 + N_r J_r(\mu, r) + \sum_{j=p} J_j(\mu, r)} = \frac{r_r}{r_a} \tag{29}$$

and

$$\Phi_p \triangleq \frac{\sum_{j=p} J_j(\mu, r)}{1 + N_r J_r(\mu, r) + \sum_{j=p} J_j(\mu, r)} = \frac{r_p}{r_a} \tag{30}$$

where we have used (24), (25) and (27) and $\sum_{j=p}$ stands for the sum over the ensemble of all non-ribosomal proteins $p = 1 \ldots n_p$.

Notice that

$$\Phi_b = \frac{r_r + r_p}{r_a} = \frac{r_{\text{bound}}}{r_a} = \frac{r_a - r}{r_a} \tag{31}$$

is the fraction of ribosomes bound to translating complexes (including both translating complexes for ribosomal and for non-ribosomal proteins) relative to the available mature ribosomes.

Notice each term

$$\frac{J_j(\mu, r)}{1 + \sum_{j=p,r} J_j(\mu, r)}$$

in Φ_p can be understood as the share of cell resources required by the $j-$th protein. Thus, the magnitude of the adimensional coefficient $J_j(\mu, r)$ is a measure of the resources recruited by the $j-$th protein.

Using the definitions (14), (16), (18) and (28) and taking into account the relationships above, we can express the dynamics of protein expression(19) and available ribosomes (22) as:

$$\dot{p}_k = \frac{1}{l_{pk}} \frac{\nu s_i}{K_{sc} + s_i} J_k(\mu, r) \left[1 - \Phi_b\right] r_a - (d_k + \mu) p_k \tag{32}$$

and

$$\dot{r}_a = \frac{1}{l_{p_r}} \frac{\nu s_i}{K_{sc} + s_i} J_r(\mu, r) \left[1 - \Phi_b\right] \Phi_t r_a - \mu r_a \tag{33}$$

respectively.

As for the dynamics of the total number of ribosomes, using $r_a = \Phi_t r_t$ we have:

$$\dot{r}_t = \frac{\nu}{l_{p_r}} \frac{s_i}{K_{sc} + s_i} J_r(\mu, r) \left[1 - \Phi_b\right] r_a - \mu r_t \tag{34}$$

6. Obtaining the cell specific growth rate

Cell growth can essentially be explained as the time variation of the protein fraction of the total cell mass. To deal with protein degradation, we take into account that the protein fraction of cell mass is the sum of the mass of functional and non-functional proteins (ie. proteins undergoing degradation). Though non-functional proteins do not contribute to cell growth, they do to cell mass. Thus, for a protein k we can consider the fraction quantity of functional molecules of the protein, p_k, and the one of non-functional units p_k^{nf} so that the total number of proteins of the species k is $p_k^T = p_k + p_k^{nf}$. Then, considering the dynamics (32) of a generic protein, we will have:

$$\begin{aligned} \dot{p}_k &= \frac{\nu}{l_{pk}} \frac{s_i}{K_{sc} + s_i} J_k(\mu, r) r - (d_k + \mu) p_k \\ \dot{p}_k^{nf} &= d_k p_k - \mu p_k^{nf} \end{aligned} \tag{35}$$

where we have taken into account that the non-functional fraction p_k^{nf} only undergoes dilution due to cell growth, there is a conversion from functional to non-functional fraction caused by protein degradation, and notice that all expressions for protein expression obtained in the previous sections referred to the functional fraction of proteins.

If we consider the average mass of an aminoacid m_{aa}, the mass weight of a protein of length l_{pk} can be approximated as $m_{aa} l_{pk}$. Thus, for p_k^T molecules of the k-th protein, their total mass weight is $m_k = m_{aa} l_{pk} p_k^T$.

The mass of n_p proteins can be approximated as:

$$m_{n_p} = \sum_{k=1}^{n_p} m_{aa} l_{pk} \left(p_k + p_k^{nf} \right) \tag{36}$$

Therefore, the times variation of the protein mass explained by this set of n_p proteins is:

$$\begin{aligned}
\dot{m}_{n_p} &= \sum_{k=1}^{n_p} m_{aa} l_{pk} \left(\dot{p}_k + \dot{p}_k^{nf} \right) \\
&= m_{aa} \sum_{k=1}^{n_p} l_{pk} \left[\frac{\nu}{l_{pk}} \frac{s_i}{K_{sc} + s_i} J_k(\mu, r) r - (d_k + \mu) p_k \right] + m_{aa} \sum_{k=1}^{n_p} l_{pk} \left(d_k p_k - \mu p_k^{nf} \right) \\
&= m_{aa} \frac{\nu s_i}{K_{sc} + s_i} \sum_{k=1}^{n_p} J_k(\mu, r) r - \mu m_{n_p} \\
&= m_{aa} \nu_t(s_i) \Phi_p r_a - \mu m_{n_p}
\end{aligned} \tag{37}$$

where recall $J_k(\mu, r)r$ is the number of ribosomes bound to complexes involved in the translation of each $k-$th protein and we have used the expression (27) relating the free and available ribosomes and definition (30). Notice, in addition, that degradation of proteins does not play a role.

As for the protein cell mass variation explained by the time variation in the total number of ribosomes r_t we will have the analogous expression:

$$\dot{m}_{r_t} = m_{aa} \nu_t(s_i) \Phi_r r_a - \mu m_{r_t} \tag{38}$$

where we have used again the fact that N_r ribosomal proteins are required to form up one ribosome. Notice that here we are only considering the weight of the protein fraction of the ribosomes. This accounts only for approximately one third of the ribosomes mass [28].

Denoting the cell protein weight $m_{n_p+r_t} = m_{n_p} + m_{r_t}$, we reach the expression:

$$\dot{m}_{n_p+r_t} = m_{aa} \nu_t(s_i) \left(\Phi_r + \Phi_p \right) r_a - \mu m_{n_p+r_t} \tag{39}$$

Now, consider the set n_p of proteins being simultaneously expressed equals the sum of the proteins being expressed by the gene circuits of interest n_{sys} plus all remaining cell environment (or context) proteins expressed in the cell n_{wt}. Then, the dynamics of the cell protein mass (m_p) can be approximated as:

$$\dot{m}_p = m_{aa} \nu_t(s_i) \Phi_b r_a - \mu m_p \tag{40}$$

Recall that the specific growth rate μ is a continuous approximation of the discrete event process of cell duplication. Here we consider that the total biomass dry weight

variation (ie. that of the whole population of cells) is mainly caused by cell duplication (i.e. population growth), and the dynamics of cell mass accumulation are much faster than those of cell duplication. Under this assumption, we may consider the protein mass for each cell quickly reaches steady state ($\dot{m}_p \approx 0$). Thus, from equation (40) we get the expression for the cell specific growth rate:

$$\mu = \frac{m_{aa}}{m_p} \nu_t(s_i) \Phi_b r_a \tag{41}$$

Notice $\Phi_b r_a$ is the number of ribosomes actively translating proteins (both ribosomal and non-ribosomal) at a given time instant. Equation (41) allows to predict this number given a specific growth rate, assuming saturation of intracellular substrate (eg. considering a batch experiment and the exponential growth phase) and considering the average values for the amino acids mass and that of the protein fraction of the cell.

Notice (41) can be expressed as a function of the total number of ribosomes r_t as:

$$\mu = \frac{m_{aa}}{m_p} \nu_t(s_i) \Phi_b \Phi_t r_t \tag{42}$$

7. Cell specific growth rate and population dynamics

Now, it will be interesting to relate the expression for the cell specific growth rate μ obtained in (41) with the classical Monod-like expressions for a population of cells growing in a bioreactor.

Recall if we have a population of N cells and we consider the average cell dry mass m_c, the total biomass dry-weight will be $M_p = N m_c$. By taking derivative with respect to time, we get:

$$\dot{M}_p = \dot{N} m_c + N \dot{m}_c \tag{43}$$

Now, consider the continuous approximation of cell duplication:

$$\dot{N} = \mu N \tag{44}$$

where $\mu = \log 2/t_d$, with t_d the duplication time, is the specific growth rate. Then:

$$\dot{M}_p = \mu M_p + N \dot{m}_c \tag{45}$$

from which we get:

$$\dot{m}_c = \frac{\dot{M}_p}{N} - \mu m_c \tag{46}$$

where $\frac{\dot{M}_p}{N}$ is the mean value per cell of the population mass growth.

As done in section 6, we assume the cell mass quickly reaches steady state as compared to the population dynamics. Thus, from equation (46) and assuming $\dot{m}_c \approx 0$

we get:

$$\mu = \frac{\dot{M}_p}{N m_c} = \frac{\dot{M}_p}{M_p} \tag{47}$$

Experimental evidence suggests that the cell densityρ varies little throughout the adult cell life [20], so:

$$\mu = \frac{\dot{M}_p}{M_p} = \frac{\rho \dot{M}_p}{\rho M_p} = \frac{\dot{V}_p}{V_p} \tag{48}$$

The identities (48) allow to relate the specific growth rate obtained in (41) with the one obtained from population-scale macroscopic experiments under the condition of steady-state growth where the rate of total cell-mass growth is identical to the rate of cell number growth [45]:

$$\mu = \frac{\dot{M}_p}{M_p} = \frac{\dot{V}_p}{V_p} = \frac{m_{aa}}{m_p} \frac{\nu s_i}{K_{sc} + s_i} \Phi_b r_a \tag{49}$$

Next, we relate the expressions (49) for growth rate obtained from intracellular considerations with the classical empirical Monod expression for the specific growth rate under a limiting substrate obtained from the experimental analysis of a macroscopic culture:

$$\mu(s) = \frac{\mu_m s}{K_s + s} \tag{50}$$

where μ_m is the maximum specific growth rate, K_s is the Monod affinity constant, and s is the concentration of the limiting substrate in the culture medium.

Alternative theoretical approaches to derive the Monod equation exist [24, 44, 42]. Here we follow a reasoning derived from the model developed in [43], where the quantity of intracellular substrate s_i is related to the one of extracellular substrate s through the dynamics of nutrient import and catabolism:

$$\dot{s}_i = e_t \frac{\nu_t s}{\frac{k_t}{V_m} + s} - e_m \frac{\nu_m s_i}{k_m + s_i} - \mu s_i \tag{51}$$

where V_m is the volume of the culture broth, e_t and e_m are transport and catabolism enzymes, and Michaelis-Menten kinetics are assumed (see [43]).

If we assume that nutrient import quickly balances nutrient catabolism and we neglect the dilution term:

$$\dot{s}_i \approx 0 \rightsquigarrow s_i = \frac{\frac{k_m}{c-1} s}{\frac{k_t}{V_m} \frac{c}{c-1} + s} \tag{52}$$

where $c = \frac{e_m \nu_m}{e_t \nu_t}$. Notice if the maximum import and catabolism fluxes are balanced, ie. $c \approx 1$, then there is a linear relationship between the intracellular amount of substrate and its extracellular concentration. Otherwise, if catabolism is more efficient than transport ($c > 1$) the intracellular amount of substrate s_i saturates with increasing

values of s.

Recall in our model the specific growth rate (49) is a function of s_i. Using (52) we obtain the Monod-like expression:

$$\frac{s_i}{K_{sc}+s_i} = \frac{\frac{k_m}{K_{sc}(c-1)+k_m}s}{\frac{K_{sc}k_t c}{V_m[K_{sc}(c-1)+k_m]}+s} \tag{53}$$

If we assume that the Michaelis-Menten constant for substrate catabolism is the same as the constant we defined in (3), that is, $K_{sc} = k_m$, we have:

$$\frac{s_i}{K_{sc}+s_i} = \frac{\frac{1}{c}s}{\frac{k_t}{V_m}+s} \tag{54}$$

Notice that the hypothesis $K_{sc} = k_m$ implicitly implies that the Michaelis-Menten constants for substrate catabolism and transport have similar values ($k_t = k_m$), in agreement with the assumptions in [43].

Also notice the term $1/c = \frac{e_t\nu_t}{e_m\nu_m}$ can be interpreted as the maximum flux yield between nutrient import and its catabolism. If $c \approx 1$, that is, under the hypothesis that the efficiency of nutrient import and catabolism are balanced so the maximum import and catabolism fluxes are similar:

$$\frac{s_i}{K_{sc}+s_i} = \frac{s}{\frac{k_t}{V_m}+s} \tag{55}$$

In case catabolism is more efficient than transport ($c > 1$) we will need an increase in the concentration of the substrate in the extracellular medium (s) to achieve the same value of $\frac{s_i}{K_{sc}+s_i}$ in (54) as compared to the balanced case $c = 1$. Finally, in case transport is more efficient than catabolism ($c < 1$) we will need lower concentrations of the extracellular substrate.

Note: The relationship (54) is valid but in the extreme cases non relevant cases $c = 0$ (there is transport into the cell but nutrients are not metabolised) and $c = \infty$. In these cases dilution cannot be neglected in equation (51) and the equilibria are different.

Then, from (49) ,(50) and (54) we get:

$$\mu(s) = \frac{\frac{1}{c}\frac{m_{aa}}{m_c}\nu\Phi_b r_a s}{\frac{k_t}{V_m}+s} = \frac{\mu_m s}{K_s+s} \tag{56}$$

so we can identify:

$$\begin{aligned} \mu_m &= \frac{1}{c}\frac{m_{aa}}{m_p}\nu\Phi_b r_a \\ K_s &= \frac{k_t}{V_m} \end{aligned} \tag{57}$$

8. Relationship between growth rate and cell mass

Our model accounts for the protein mass distribution but does not consider the relationship between growth rate and the total cell protein mass. Several phenomenological models have been proposed in the literature accounting for the relationship between growth rate and cell dry weight, like the recent ones [37, 45]. In [45] a Monod-like relation between the chromosome replication–segregation period $C + D$ and the cell specific growth rate μ is considered:

$$C + D = \frac{\alpha + \beta\mu}{\mu} \tag{58}$$

We estimated the parameters α and β in (58) using the data in [5]. Figure 2(left) shows the good fit obtained.

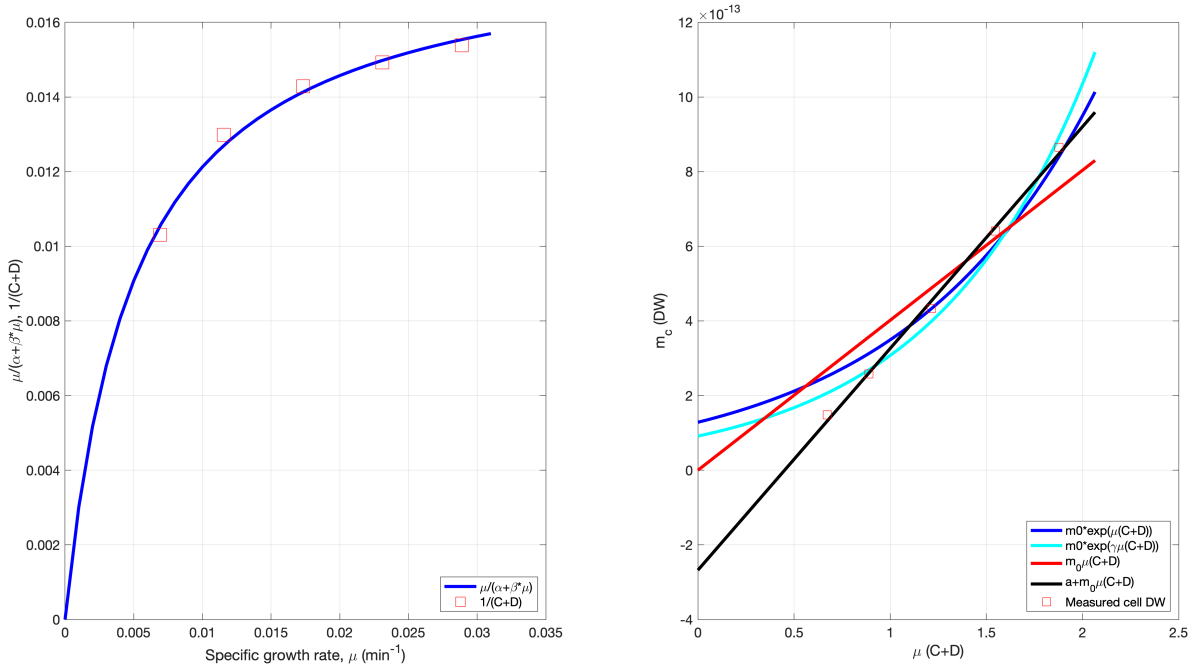

Figure 2: ns.

In addition the authors in [45] propose a linear relation between the cell dry weight and the product of $C + D$ and μ:

$$m_c = m_0\mu(C + D) \tag{59}$$

Therefore, according to this model, there is an affine relationship between the cell dry weight and specific growth rate:

$$m_c = m_0(\alpha + \beta\mu) \tag{60}$$

As shown in Figure 2(right), for the data we used, the relationship (59) gives a very rough approximation. Indeed, as shown in the same figure, and affine relationship gives much better fit.

As an alternative, in [37] an exponential relationship is proposed between the cell volume S_c and the product of $\mu(C+D)$:

$$S_c = S_0 e^{\mu(C+D)} \tag{61}$$

Assuming constant cell density, equation (61) can be expressed as a function of the cell dry weight. Figure 2(right) shows the fit assuming constant cell density, so that equation (61) can be expressed as a function of the cell dry weight. Notice that a better fit can be obtained considering the modified relationship:

$$m_c = m_0 e^{\gamma\mu(C+D)} \tag{62}$$

Now, if we consider expressions (58) and (62), we get:

$$m_c = m_0 e^{\gamma(\alpha+\beta\mu)} = \bar{m}_0 e^{\bar{\gamma}\mu} \tag{63}$$

where $\bar{m}_0 = m_0 e^{\gamma\alpha}$ and $\bar{\gamma} = \gamma\beta$.

Notice equation (63) is the solution of the differential equation:

$$\frac{dm_c}{d\mu} = \bar{\gamma} m_c \tag{64}$$

which expresses that the variation of cell mass (dry weight) relative to the variation of the specific growth rate is proportional to the cell mass. Figure 3 (left) shows the results obtained for (63) and Table 8 lists the best fitted parameters. Better fits can be obtained with alternative phenomenological expressions to (63) at the cost of losing the simple interpretation provided by expression (64).

For the relationship between the cell protein content and the specific growth rate we postulate a relationship analogous to (64). Thus, we consider:

$$\frac{dm_p}{d\mu} = \beta m_p \tag{65}$$

Figure 3 (right) shows the results obtained and Table 8 lists the parameters corresponding to the best fit.

Notice that alternative phenomenological relationships can be used instead, with possibly better fit to the experimental data. Thus, for instance, the power law model $m_p = \bar{m}_0 e^{\bar{\gamma}\mu^\nu}$ with $\bar{m}_0 = 1.2918\,10^{-14}$, $\bar{\gamma} = 14.1089$ and $\nu = 0.389$ is an alternative.

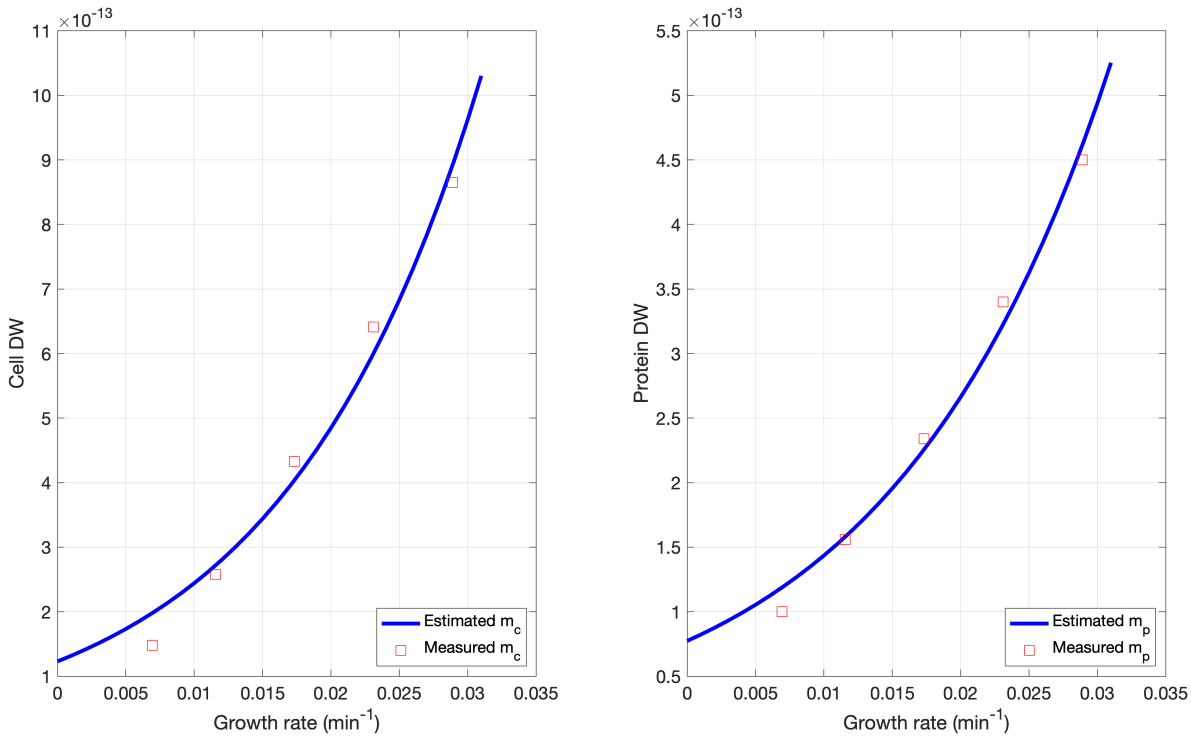

Figure 3: Fit between the cell dry weight (left) and protein content dry weight (right) and the specific growth rate.

Model	**Model parameters**	
$m_c = \bar{m}_0 e^{\bar{\gamma}\mu}$	$\bar{m}_0 = 123.022\, 10^{-15}$ (g)	$\bar{\gamma} = 68.5547$ (min)
$m_p = \bar{m}_{p0} e^{\beta\mu}$	$\bar{m}_{p0} = 77.3748e\, 10^{-15}$ (g)	$\beta = 61.7813$ (min)

Table 1: Phenomenological models relating cell and protein dry mass and specific growth rate and their best fit parameters for the data in [5].

9. The relative resources recruitment strength provides the relative mass fraction at steady state.

the relative resources recruitment strength of a given protein equals its relative mass

Recall the expressions (19) and (23) for the dynamics of given protein and the average ribosome respectively. Now, using (28), (41) and the dynamics of the total $k-$th protein $p_k^t = p_k^{nf} + p_k$ given by (35) we can rewrite:

$$\dot{p}_k^t = \left[\frac{1}{l_{pk}} \frac{m_p}{m_{aa}} \frac{J_k}{\sum_{j=p,r} J_j} - p_k^t \right] \mu \tag{66}$$

and

$$\dot{r}_a = \left[\frac{1}{l_{p_r}} \frac{m_p}{m_{aa}} \frac{\Phi_t J_r}{\sum_{j=p,r} J_j} - r_a \right] \mu \tag{67}$$

where $\sum_{j=p,r} J_j$ stands for the sum of the resources recruitment strengths for all proteins in the cell, including both ribosomal and non-ribosomal ones.

Now, using the expression for the cell protein mass

$$m_p = m_{aa} \sum_{j=p,r} l_{p_j} p_j \tag{68}$$

where the sum includes both the ribosomal and non-ribosomal proteins and we have assumed a common average amino acid mass m_{aa}, we have:

$$\dot{p}_k^t = \left[\frac{\sum_{j=p,r} l_{p_j} p_j}{l_{p_k}} \frac{J_k}{\sum_{j=p,r} J_j} - p_k^t \right] \mu \tag{69}$$

and

$$\dot{r}_a = \left[\frac{\sum_{j=p,r} l_{p_j} p_j}{l_{p_r}} \frac{\Phi_t J_r}{\sum_{j=p,r} J_j} - r_a \right] \mu \tag{70}$$

Notice that the steady state will reached either when the growth rate $\mu = 0$ or at:

$$p_{k=p,r,\mathrm{ss}} = \frac{\frac{J_k}{\sum_{j=p,r} J_j}}{\frac{l_{p_k}}{\sum_{j=p,r} l_{p_j} p_j}} \tag{71}$$

where $p_{k=p,r,\mathrm{ss}}$ stands for any protein, either ribosomal or non-ribosomal. Notice that expression (71) tells us that at steady state the relative resources recruitment strength of a given protein equals its relative mass in the cell:

$$\frac{l_{p_k} p_{k=p,r,\mathrm{ss}}}{\sum_{j=p,r} l_{p_j} p_j} = \frac{J_k}{\sum_{j=p,r} J_j} \tag{72}$$

10. Average host dynamics and steady state.

We were interested in having an average model for the host dynamics and its steady state that can be used as base for analysing host-circuit interactions. To this end we considered, on the one hand, the dynamics of the mass of total ribosomes m_{r_t} and, on the other, the ones of the mass of the ensemble of non-ribosomal proteins m_{wt}. In the later case, we have considered this set as a lumped species with a single average resources recruitment strength $J_{\mathrm{wt}}(\mu, r)$.

Thus, using (67), we have:

$$\dot{m}_{r_t} = \mu \left[m_p \frac{N_r J_r(\mu, r)}{\sum J_{r,\mathrm{wt}}(\mu, r)} - m_{r_t} \right] \quad ; \quad J_r(\mu, r) = 0.62 \frac{l_p^r}{l_e} \omega_r \frac{1}{\frac{d_m^r}{K_{C^0}^r(s_i)} + \mu r} \tag{73}$$

$$\dot{m}_{\mathrm{wt}} = \mu \left[m_p \frac{N_{\mathrm{wt}} J_{\mathrm{wt}}(\mu, r)}{\sum J_{r,\mathrm{wt}}(\mu, r)} - m_{\mathrm{wt}} \right] \quad ; \quad J_{\mathrm{wt}}(\mu, r) = 0.62 \frac{l_p^{\mathrm{wt}}}{l_e} \omega_{\mathrm{wt}} \frac{1}{\frac{d_m^{\mathrm{wt}}}{K_{C^0}^{\mathrm{wt}}(s_i)} + \mu r} \tag{74}$$

with the effective RBS strengths

$$K_{C^0}^k(s_i) = \frac{K_b^k}{K_u^k + \frac{\nu_t(s_i)}{l_e}} \tag{75}$$

for $k = \{r, \mathrm{wt}\}$, where we use the data of $\nu_t(\mu)$ for the effective translation rate $\nu_t(s_i)$ –implicitly assuming the growth rate at steady state is a function of the substrate s_i– and where

$$\sum J_{r,\mathrm{wt}}(\mu, r) = N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}}(\mu, r) \tag{76}$$

and N_r and N_{wt} are the number of ribosomal proteins building up a ribosome (see Section 3) and the number of non-ribosomal proteins in the *wild type* cell respectively, so that the number of proteins and the resulting mass are related by:

$$m_{r_t} = m_{aa} N_r l_p^r r_t = m_{rib} r_t \tag{77}$$

$$m_{\mathrm{wt}} = m_{aa} N_{\mathrm{wt}} l_p^{\mathrm{wt}} p_{\mathrm{wt}} \tag{78}$$

where m_{rib} is the average mass of the protein content of ribosomes, and we have assumed a common average amino acid mass m_{aa} as in (68) and average protein lengths l_p^r and l_p^{wt}.

Notice from (73)–(74) that at steady state we will have:

$$\frac{m_{r_t}}{m_p} = \frac{N_r J_r(\mu, r)}{\sum J_{r,\mathrm{wt}}(\mu, r)} \tag{79}$$

$$\frac{m_{\mathrm{wt}}}{m_p} = \frac{N_{\mathrm{wt}} J_{\mathrm{wt}}(\mu, r)}{\sum J_{r,\mathrm{wt}}(\mu, r)} \tag{80}$$

The number of free ribosomes r is obtained as:

$$r = \frac{\Phi_t r_t}{1 + N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}}(\mu, r)} \tag{81}$$

with r_t obtained from (77), and the specific growth rate as:

$$\mu = \frac{m_{aa}}{m_p} \nu_t(s_i) \Phi_b \Phi_t r_t \tag{82}$$

with

$$\Phi_b = \frac{N_r J_r(\mu, r) + N_{\text{wt}} J_{\text{wt}}}{1 + N_r J_r(\mu, r) + N_{\text{wt}} J_{\text{wt}}} \tag{83}$$

Notice from (73), (81) and (82) that at steady state:

$$\mu_{ss}(s_i) = \frac{\nu_t(s_i)}{N_r l_p^r} \Phi_t \Phi_r = \frac{m_{aa}}{m_{rib}} \nu_t(s_i) \Phi_t \Phi_r \tag{84}$$

that is, the growth rate at steady state depends linearly on the fraction $\Phi_t \Phi_r$ of bound ribosomes being used to build up ribosomes relative to the total number of ribosomes.

At steady state, the flux of free resources for a given intracellular substrate s_i, defined as the number of free ribosomes times the cell growth rate, can be obtained as:

$$\mu_{ss} r_{ss}(s_i) = m_{aa} \nu_t(s_i) \left(\frac{\Phi_r \Phi_r}{m_{rib}} \right)^2 \frac{1 - \Phi_b}{\Phi_b} m_p \tag{85}$$

showing a linear relationship with the cell protein weight m_p.

The cell protein weight m_p depends of the growth rate. To this end, we used both the data available in [5] and the phenomenological relationships obtained in section 8.

11. Productivity of exogenous proteins depends on host-circuit interaction

Product titer and productivity are important measures of performance in biotechnological applications. In this section we consider the expression of a heterologous protein, and we analyse how its copy number at steady state and its specific productivity vary as a function of the RBS and promoter strengths and the host-circuit interaction.

We define the specific productivity a protein A, Π_A, as the steady-state rate of expression of A per cell. Thus, if we consider a population of N cells and the continuous approximation of cell duplication (44), the total quantity of the protein A, P_A^N, will increase as the population of cells does as:

$$\dot{P}_A^N = p_{A,\text{ss}} \dot{N} = p_{A,\text{ss}} \mu N \tag{86}$$

where $p_{A,\text{ss}}$ is the quantity of the protein A at steady state in a single cell. Therefore:

$$\Pi_A = \frac{\dot{P}_A^N}{N} = p_{A,\text{ss}} \mu \tag{87}$$

For the host endogenous dynamics we considered the model as described in section 10. We extend the model by adding the dynamics of the exogenous protein of interest A as:

$$\dot{m}_A = \mu \left[m_p \frac{J_A(\mu, r)}{\sum J_{r,\text{wt},A}(\mu, r)} - m_A \right] \quad ; \quad J_A(\mu, r) = 0.62 \frac{l_p^A}{l_e} \omega_A \frac{1}{\frac{d_m^A}{K_{C^0}^A} + \mu r} \tag{88}$$

where

$$\sum J_{r,\mathrm{wt},A}(\mu, r) = N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}} + J_A(\mu, r) \tag{89}$$

and the number of molecules of protein A and the resulting mass are related by:

$$m_A \quad = m_{aa} l_p^A p_A \tag{90}$$

The number of free ribosomes r is obtained as:

$$r = \frac{\Phi_t r_t}{1 + N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}} + J_A(\mu, r)} \tag{91}$$

and the specific growth rate using (82) with

$$\Phi_b = \frac{N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}} + J_A(\mu, r)}{1 + N_r J_r(\mu, r) + N_{\mathrm{wt}} J_{\mathrm{wt}} + J_A(\mu, r)} \tag{92}$$

12. Model parameters

Table 12 shows the set of parameters used in the model.

13. Estimation of the fraction Φ_b

Next we evaluate the fraction Φ_b of ribosomes being used in translating complexes relative to the mature available ones as a function of the free ribosomes r (see Section 5). We expect a very low number of free ribosomes. If this was not the case, there would be no real competition to recruit them. To set an initial upper limit, we use the estimation $r = \mathbf{N}(350, 35)$ in [23]. In addition, having too many free ribosomes in excess would imply a superfluous us of energy for the cell. Considering this hypothesis, we evaluated equation (31) as a function of the mature available ribosomes r_a and the free ones r. Figure 4(left) shows the values estimated. Notice the values of Φ_b close to $\Phi_b = 1$ indicating that the cell is always at the edge of its maximum capacity for using the available resources.

From the estimation of the fraction Φ_b we can obtain that of the dimensionless sum $\sum_{j=k,r} J_j(\mu, r)$ reflecting the resources recruitment load generated by the whole set of ribosomal and non-ribosomal proteins being expressed at a given moment in the cell (see Figure 4(right)) using the definition (28).

To validate the estimations above, we used the data in [19] to evaluate the sum $\sum_{j=k,r} J_j(\mu, r)$ for all genes being expressed in *E. coli*. We took into account that the data in [19] was obtained for fast growing cells, with $t_d = 21$ minutes. Under this conditions, it is sensible to consider that the intracellular substrate will be saturated and the cells are growing at its maximum growth rate. Then, we took into account that the number of active available mature ribosomes $\Phi_b r_a$ is related to the cell specific growth rate as described in Section 6. Recall, from equation (41) and considering substrate

Model parameters for *E. coli*				
Parameter	**Description**	**Value**	**Unit**	**Reference**
t_d	duplication time	$[20, 60]$	min	[28]
μ	specific growth rate	$[0.01, 0.035]$	min^{-1}	$\mu = \log 2/t_d$
$y_{x/s}$	yield biomass on glucose	0.5	gDW/g glucose	
$\bar{l}_p$	average non-ribosomal protein length	333	aminoacids (aa molec^{-1})	calculated from data in [19]
$\bar{l}_{p,a}$	average ribosomal protein length	195	aminoacids (aa molec^{-1})	calculated from data in [19]
l_e	ribosome occupancy length	$[20, 30]$	aa molec^{-1}	estimated ([38, 14, 13, 30])
$l_{e,\text{min}}$	ribosome occlusion length	11	aa molec^{-1}	[36]
l_a	average total number aa per ribosomal protein	165	aa molec^{-1}	estimated from protein content ([39]) and $N_r = 55$
ν	maximum translation speed per ribosome	1260	$\text{aa} \cdot \text{min}^{-1}$	[28, 35]
k_u^k	dissociation rate RBS-ribosome	[6, 135]	min^{-1}	[11, 23]
k_b^k	association rate RBS-ribosome	[3, 15]	$\text{molec}^{-1} \cdot \text{min}^{-1}$	[11, 34, 23]
$t_{\frac{1}{2},\text{mRNA}}$	average mRNA half-life	$[2, 6]$	min	[1, 8, 19, 17]
$d_{m,\text{ss}}$	median mRNA degradation rate (steady phase)	0.17	min^{-1}	[1]
d_m	median mRNA degradation (exponential growth)	0.28	min^{-1}	[8]
η	maximum transcription speed	$[1500, 3000]$	$\text{aa} \cdot \text{min}^{-1}$	[28, 25]
m_{aa}	average amino acid mass	110	Da	[39]
m_{aa}	average amino acid mass	$182.6\,10^{-24}$	g	[39]
m_p	geometric average of cell protein weight	$224\,10^{-15}$	gDW	[28, 5]

Table 2: Model parameters.

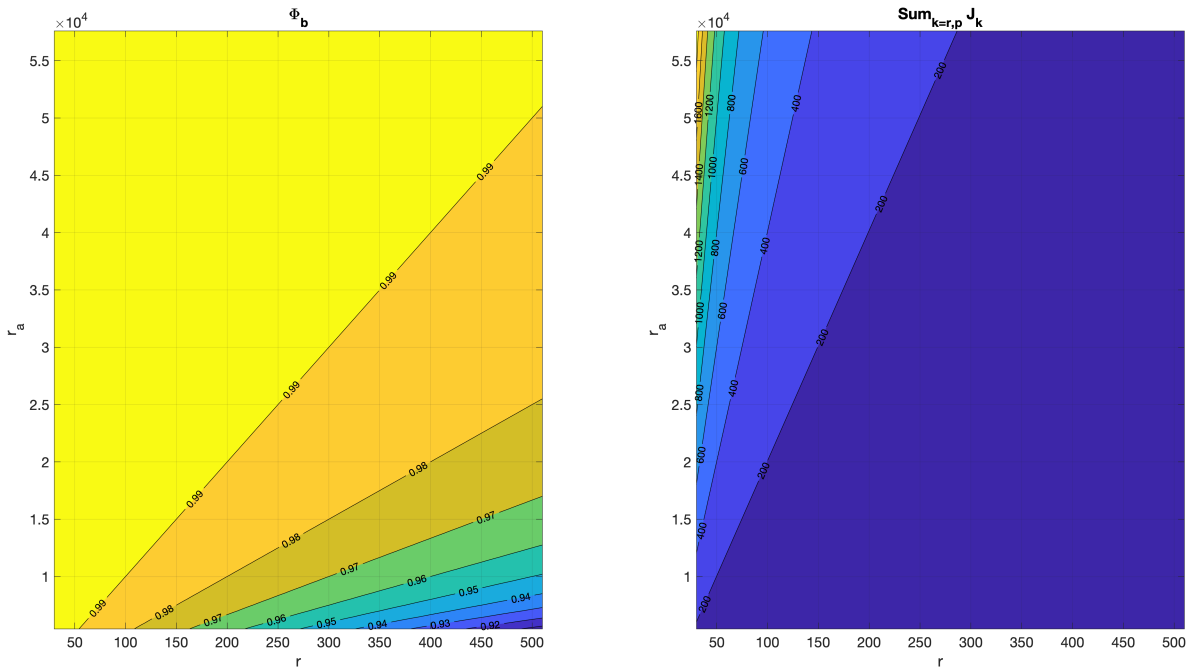

Figure 4: (Left:) Fraction Φ_b of ribosomes actually being used in translating complexes as a function of the available functional r_a and the free ones r. (Right:) Corresponding estimation of the sum $\sum_{j=k,r} J_j(\mu, r)$

saturation, we have:

$$\mu = \mu_m = \frac{m_{aa}}{m_p} \nu \Phi_b r_a$$

We estimated the cell protein mass m_p as the total copy number of all proteins in the cell obtained in [19] times the average amino acid mass. The dynamic model for the expression of protein p in [19] considers:

$$\dot{p} = \frac{\beta_p \beta_m}{d_m} - \mu p \tag{93}$$

where β_m (mRNA/t) is the transcription rate, β_p (protein/(mRNA·t)) the translation one and d_m the mRNA degradation rate. For those transcripts without information for d_m in [19] we used the value shown in Table 12.

Then we used equation (19) to derive the relationship:

$$J_k = \frac{l_{pk}}{\nu} \frac{\beta_p \beta_m}{d_m} \frac{1}{r} \tag{94}$$

We evaluated (93) considering the set of all non-ribosomal proteins, the ribosomal ones, and the full set of proteins.

Figure 5(right) shows the very good agreement between the estimation obtained using equation (31) and the values calculated using equation (94) and the data in [19].

The difference is below 5.6% for all r. Is is important to stress that equation (31) is a purely theoretical expression. Only the expression (41) relating the number of active available mature ribosomes $\Phi_b r_a$ with the cell specific growth rate was used to estimate $\Phi_b r_a$ from the maximum specific growth rate corresponding to the data in [19].

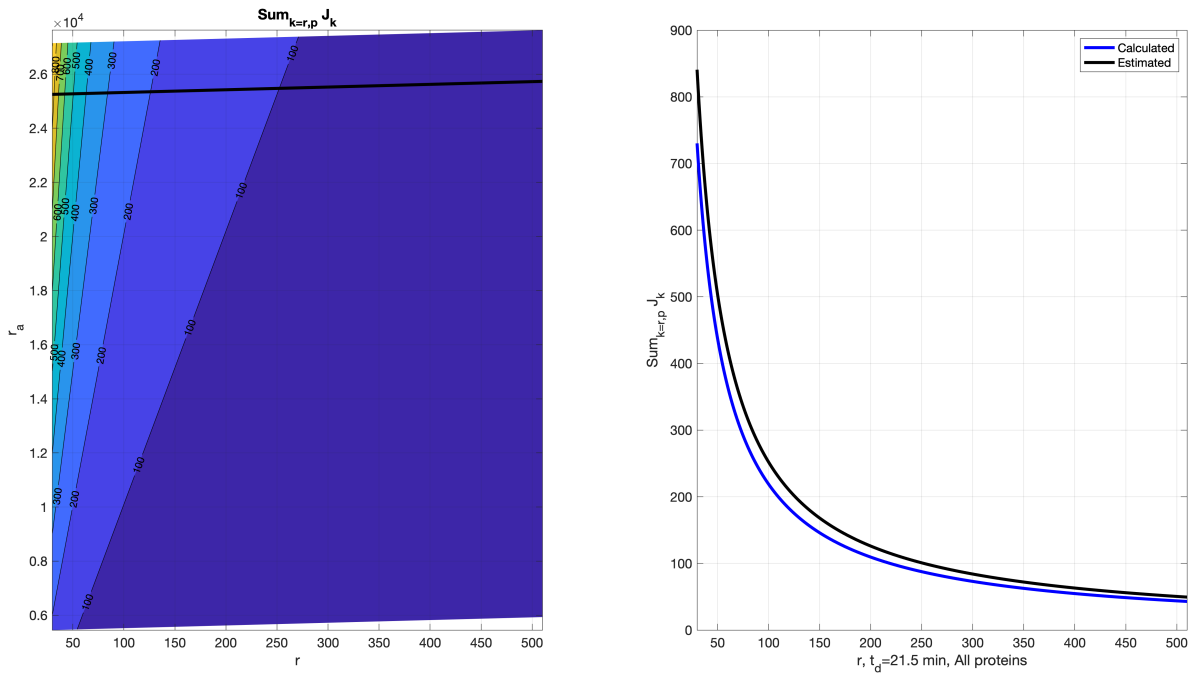

Figure 5: (Left:) Sum $\sum_{j=k,a} J_j(\mu, r)$ as a function of the available functional r_a and the free ones r. The black line depicts the value of mature available ribosomes r_a as a function of the free ones r assuming the number of bound ribosomes that correspond to the specific growth rate considered in [19]. (Right:) Corresponding estimation of the sum $\sum_{j=k,a} J_j(\mu, r)$ (black) and value obtained by evaluating the sum using (94) with the data in [19].

NOTE:. It is interesting to notice that from (41) and (31) we can derive the expression:

$$\Phi_b = \frac{\frac{m_{aa}}{m_p}\frac{\mu_m}{\nu}}{\frac{m_{aa}}{m_p}\frac{\mu_m}{\nu} + r} \tag{95}$$

from which we obtain:

$$\sum_{j=k,a} J_j(\mu_m, r) = \frac{m_{aa}}{m_p}\frac{\mu_m}{\nu}\frac{1}{r} \tag{96}$$

Now, considering (94), we get the relationships

$$\mu_m = \frac{m_{aa}}{m_p}\sum_{j=k,a} l_{pj}\frac{\beta_{pj}\beta_{mj}}{d_{mj}} \tag{97}$$

and

$$\Phi_b r_a = \frac{1}{\nu} \sum_{j=k,a} l_{pj} \frac{\beta_{pj}\beta_{mj}}{d_{mj}} \tag{98}$$

relating the maximum growth rate μ_m and the bound ribosomes $\Phi_b r_a$ respectively with the transcription and translation rates, the transcripts degradation rates and the length of the proteins being expressed.

14. Estimation of the average resources recruitment strength.

In Section 2 we derived equation (19) for the expression of a given protein and we defined the dimensionless function $J_k(\mu, r)$ (see equation (18)) that quantifies the capacity of the k-th gene to recruit cellular resources to get expressed. In the previous section, we obtained the resources recruitment strength J_k for each gene in *E. coli* estimated from the experimental data in [19], and the corresponding sum $\sum_{j=k,r} J_j(t_d = 21.5, r)$ reflecting the total resources recruitment load in the cell. Next we evaluate the average resources recruitment strength J_{avg} as a function of the number of free ribosomes r for both ribosomal and non-ribosomal proteins.

As a first step we used the data in [19] and expression (94) to calculate the individual values of maximum resources recruitment strength J_k evaluated at $r = 1$ for the set of non-ribosomal and ribosomal protein expression genes and sorted them by magnitude. As seen in Figure 6 (top) the values of J_k span several orders of magnitude. We also evaluated the ration between the sorted values of J_k and the corresponding protein lengths. The results do not sensibly change with respect to the previous ones (see Figure 6 (bottom)). That is, the resources recruitment strength of *E. coli* genes is not fundamentally determined by the lengths of the proteins they code. This suggests, as expected, that factors such as the effective transcription and translation rates are more relevant. The fact that the ribosomal proteins, essential for the cell and continuously being expressed, have much higher resources recruitment strength values than the non-ribosomal ones. Moreover, the range of variation of J_k over the ribosomal proteins is much lower than for non-ribosomal ones. This is consistent with the fact that to great extent all ribosomal proteins are equally important for the cell.

Not all genes are expressed all the time. As a proxy to estimate how many genes are active at a given time we calculated the cumulative sum of the maximum resources recruitment strength and obtained how many genes being expressed are required to explain both 95% and 99% of the total cumulative sum. We did this independently for both ribosomal and non-ribosomal proteins. Figure 7 shows the results obtained. *E coli* has around 4225 protein-coding genes [28, 22]. We had information obtained from [19] to get the values of $J_k(r = 1)$ for 3551 non-ribosomal and 68 ribosomal ones. That is 3619 genes, around 86% of all *E coli* genes. Therefore, the results obtained are enough representative of the cell.

Our results show that out of the 68 ribosomal genes, 49 of them (72%) explain 95% of the cumulative sum of the maximum resources recruitment strength of the ribosomal genes. To explain 99% we need 57 ribosomal genes (84% of them). On the other hand,

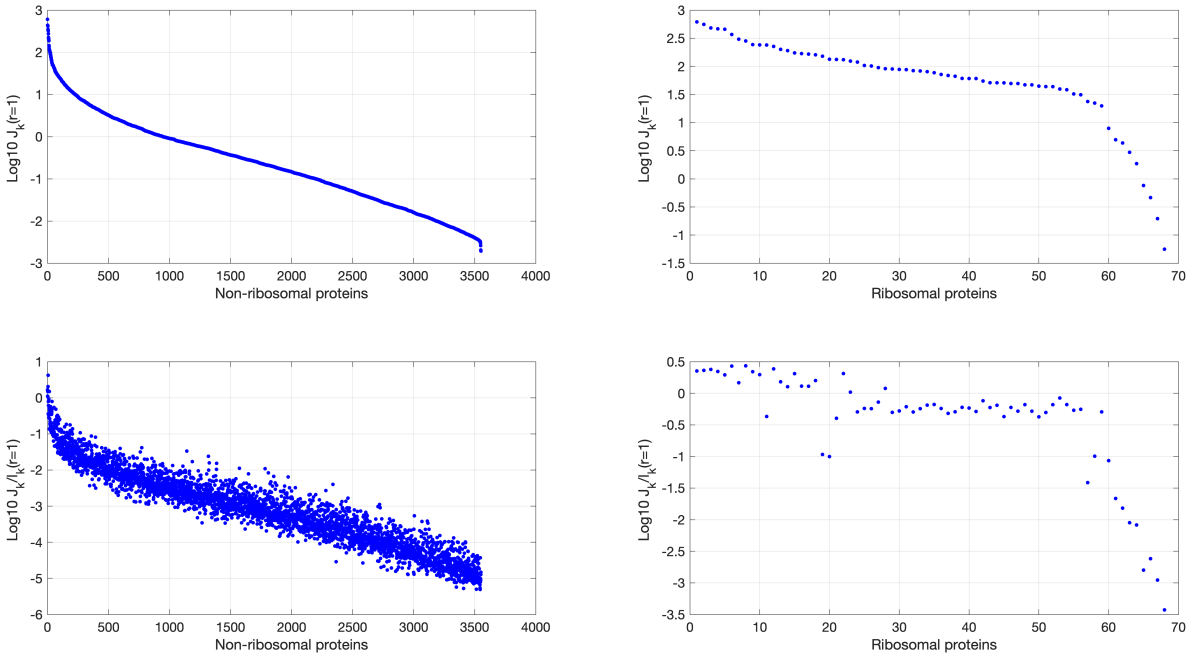

Figure 6: (Top:) Maximum resources recruitment strength $J_k(r = 1)$ for the set of non-ribosomal (left) and ribosomal protein expression genes (right) sorted by magnitude in logarithmic scale. (Bottom:) Ratio between the sorted maximum resources recruitment strength $J_k(r = 1)$ and the corresponding protein length (aa).

for non-ribosomal genes we need 875 out of 3551 genes (25%) to explain 95% of the cumulative sum and 1735 (49%) to explain the 99%. These results show again that ribosomal genes are continuously needed for the cell and thus are continuously expressed. On the contrary, non-ribosomal genes are not required at all times. Since transcription and translation are energetically expensive processes, these genes are regulated to be expressed only when required. This explains the very low values obtained for the resources recruitment strength for most of them. This reflects the fact that most of the time they are "switched-off".

The results above suggest that to get a reliable average value of the resources recruitment strength J_{avg} of a gene as a function of the number of free ribosomes r we must consider, on the one hand ribosomal and non-ribosomal genes as two differentiated sets and, on the other, we must consider some estimation of the number of active genes. To this end, a conservative estimation will consider as active the number of genes that explain 99% of the total cumulative sum of the maximum resources recruitment strength. Thus, we used the estimation of the sum $\sum_{j=p,r} J_j(\mu, r)$ obtained in Section 13 (see Figure 5(left)) and assumed that the fraction of $\sum_{j=p,r} J_j(\mu, r)$ allocated for ribosomal and non-ribosomal proteins keep constant with the growth rate. Then, using the value of this fraction obtained from the data [19] for $t_d = 21.5$ minutes, we considered 57 ribosomal and 1735 non-ribosomal active genes respectively to get the average resources recruitment strength J_{avg} for both ribosomal and non-ribosomal genes.

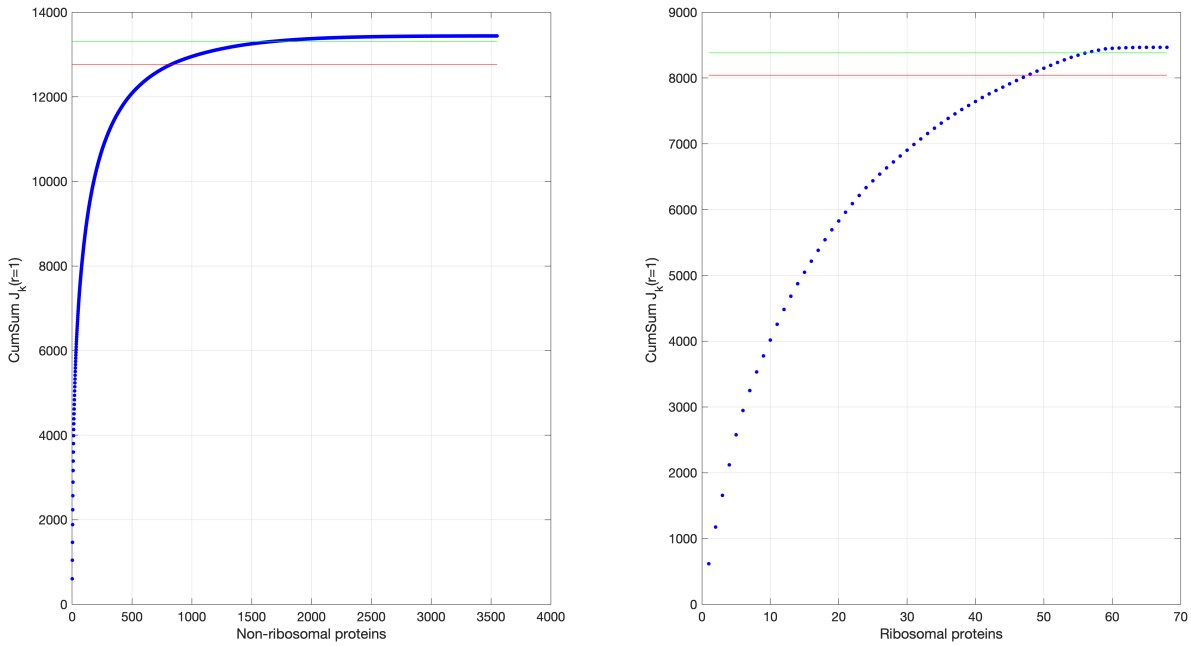

Figure 7: Cumulative maximum resources recruitment strength for the set of non-ribosomal (left) and ribosomal protein expression genes (right). The horizontal lines correspond to the 95% (red) and 99% (green) levels respectively.

Figures 8 and 9 show the values obtained for $J_{\text{avg}}(\mu, r)$ for an *E. coli* protein as a function of the number of free ribosomes r and both the quantity of available mature ribosomes r_a and the duplication time t_d respectively.

15. Estimation of the number of free ribosomes in the cell

Estimation of the number of free ribosomes in the cell, r, is key for assessing the competition among the cell circuits for cellular resources. The results in the previous sections suggest that an extremely low number of free ribosomes, with order of magnitude in the range 10^1, gives a high sensitivity of the total cumulative sum of the maximum resources recruitment strength with respect to variations in the amount of free ribosomes (see Figures 4 and 5) while an order of magnitude in the range 10^2 is enough to keep a robust value of the total amount of recruited resources with respect to variations in the number of free ribosomes. That is, by not expressing superfluous resources, the cell forces a competition for them that induces a high sensitivity of the total amount of recruited resources with respect to variations in the number of free ribosomes, while a small surplus of superfluous resources induces robustness.

To evaluate the range of expected values of r we used experimental data of the translation efficiency per mRNA. Notice from the dynamics (19) for a protein p_k we can define:

$$Y_{p/\text{mRNA}} \triangleq \frac{\nu}{l_{pk}} \frac{J_k r}{\omega_k} = 0.62 \frac{\nu}{l_e} \frac{r}{\frac{d_{mk}}{K^k_{C0}(s_i)} + \mu r} \frac{s_i}{K_{sc} + s_i} \tag{99}$$

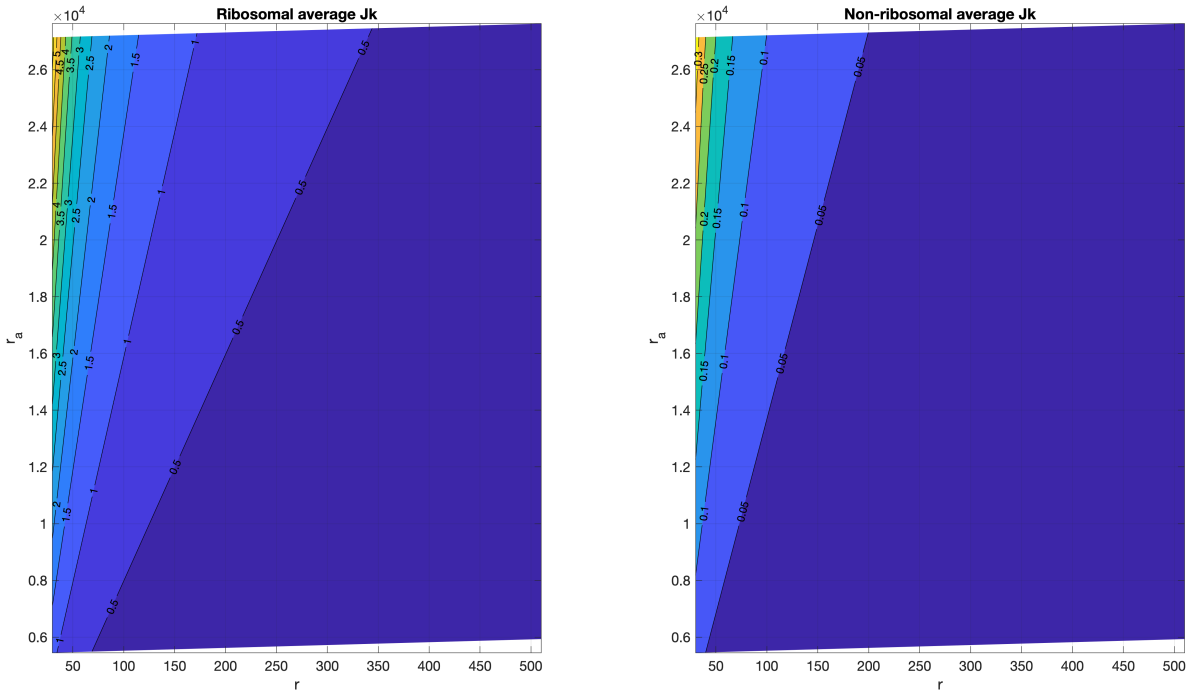

Figure 8: Estimated values obtained for $J_{\text{avg}}(\mu, r)$ for an *E. coli* protein as a function of the number of free ribosomes r and that of available mature ribosomes r_a. (Left:) ribosomal proteins. (Right:) non-ribosomal proteins.

where s_i is the copy number of molecules of intracellular substrate, d_{mk} is the mRNA degradation rate, recall $K^k_{C^0}(s_i)$ is a substrate dependent parameter essentially related to the RBS strength and we consider that substrate availability will only affect translation and not transcription (see Section 2). Notice $Y_{p/\text{mRNA}}$ is the copy number of protein produced per amount of transcript.

We used the data from [19] to estimate an upper bound for the number of free ribosomes r using (99) and the values for $Y_{p/\text{mRNA}}$ obtained from expression (94):

$$Y_{p/\text{mRNA}} = \frac{\beta_{pk}}{d_{mk}} \tag{100}$$

Then, the relationship between the RBS strength-related term $K^k_{C^0}$ and the free ribosomes r becomes:

$$0.62\frac{\nu}{l_e}\frac{r}{\frac{d_{mk}}{K^k_{C^0}(s_i)} + \mu_m f(s_i) r} f(s_i) = \frac{\beta_{pk}}{d_{mk}} \tag{101}$$

where $f(s_i) = s_i/(K_{sc} + s_i)$ and we have used equations (55), (56) and (57) relating the specific growth rate μ with the maximum one μ_m and the availability of intracellular substrate. Notice that, for any given protein and intracellular substrate availability, the number of free ribosomes will determine the required value of the RBS strength-related term $K^k_{C^0}(s_i)$ to attain the experimental value of the translation efficiency per mRNA

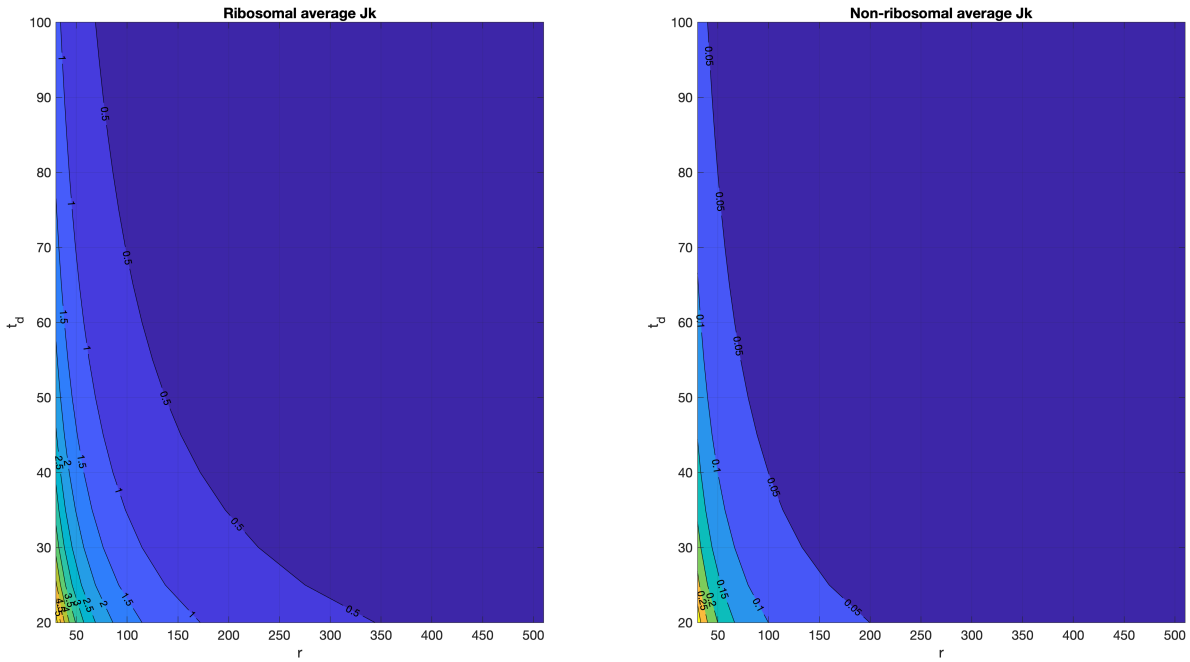

Figure 9: Estimated values obtained for $J_{\text{avg}}(\mu, r)$ for an *E. coli* protein as a function of the amount of free ribosomes r and the duplication time t_dt. (Left:) ribosomal proteins. (Right:) non-ribosomal proteins.

$Y_{p/\text{mRNA}}$.

The translation efficiency given by expression (101) depends on the ribosomes density $1/l_e$. An average ribosomes density around 4.2 ribosomes per 100 codons in optimal growth conditions has been reported in the literature for the prokaryote *L. lactis* [30]. This value is the same we obtained from the data in (94) by considering the total number of available active ribosomes $\Phi_b r_a$ obtained in Section 14 (see Figure 5) and dividing it by the sum of the lengths of all proteins weighted by a factor 0.5 to account for the estimation that 50% of the genes are active (explain 99% of the cumulative sum of the maximum resources recruitment strength). Similar values are found for other organisms [14]. For *E. coli* a value of 3.5 is given in [28]. The ribosomes density is inversely log-linearly related to the length of the coding sequence, with a slope quite consistent for a variety of organisms [14]. To account for this, we approximated a power law consistent with the findings in [14] and resulting in 4 ribosomes per 100 codons for an average protein length of 330 codons. We obtained the relationship:

$$\frac{1}{l_e} = \frac{0.0703}{l_{pk}^{0.097}} \tag{102}$$

This gives a range $l_e \subset [18, 31]$, with a value $l_e = 25$ for the average protein length. The minimum value is consistent with the shortest protein length (18 codons) in the database we used.

Figure 10 shows the results obtained for the set of all ribosomal and non-ribosomal

proteins and their average values.

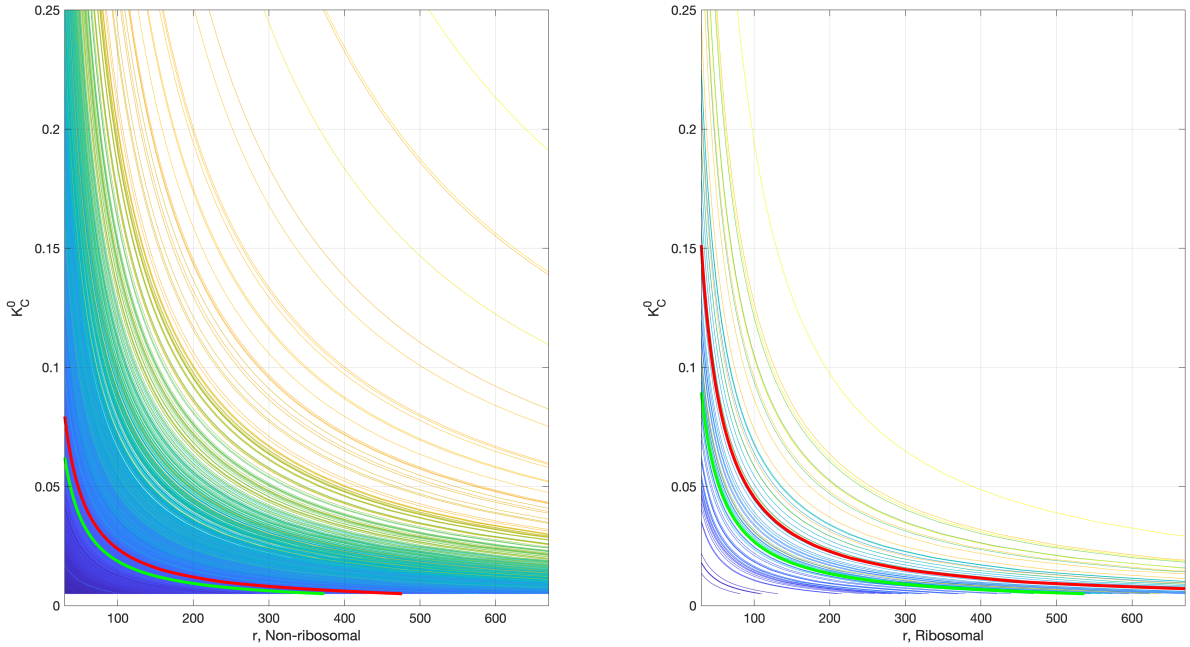

Figure 10: Relationship between the RBS strength-related term $K^k_{C^0}$ and the number of free ribosomes r obtained using the experimental data in [19] for non-ribosomal (left) and ribosomal (right) proteins in *E. coli*. Thin lines correspond to the experimental value of the translation efficiency per mRNA $Y_{p/\mathrm{mRNA}}$ for each protein. The red thick line corresponds to the mean for all proteins in the corresponding non-ribosomal and ribosomal sets. The green thick line corresponds to the approximated mean when the term associated to the maximum specific cell growth rate is neglected in the expression (101).

From the results shown in Figure 10, notice that the number of free ribosomes r required to explain the experimental value of the translation efficiency per mRNA $Y_{p/\mathrm{mRNA}}$ for each protein increases as the value of the RBS strength-related term $K^k_{C^0}$ decreases. Indeed, the more free ribosomes are available, the less competition for shared resources. The number of free ribosomes r is an indicator of the level of competition for resources. Thus, expression (101) implies that a gene producing short-living transcripts will require, for the same level of competition, a stronger RBS to achieve the same translation efficiency per mRNA $Y_{p/\mathrm{mRNA}}$ as one with long-living transcripts.

To estimate an upper limit for the copy number of free ribosomes required to achieve the experimental translation rates per mRNA, we considered an upper bound for the RBS strength-related term $K^k_{C^0}$.

Recall from equation (10) that $K^k_{C^0}$ term is a function of the intracellular substrate availability. Since the data we used from [19] was obtained for fast growing cells, we can consider intracellular substrate saturation. Under this condition we get:

$$K^k_{C^0} = \frac{K^k_b}{K^k_u + \frac{\nu}{l_e}} \tag{103}$$

Using the value of ν in Table 12 and the range of l_e given above, we estimate $\frac{\nu}{l_e} \subset [40,\ 70]$ (molec$^{-1} \cdot$ min^{-1}). On the other hand, the values of the association and dissociation rates of the ribosome to the RBS, K_b^k and K_u^k, may vary in a large range. Values $K_b^k \subset [3,\ 15]$ (molec$^{-1} \cdot$ min^{-1}) are found in the literature (see Table 12). We use a conservative upper bound $K_b^{\max} = 10$ (molec^{-1}) considering binding is diffusion controlled. From the literature, we consider a range for the dissociation rate $K_u^k \subset [3,\ 135]$ (min^{-1}). Overall, these estimates give us a range (under the assumption of intracellular substrate saturation) $K_{C^0}^k \subset [0.02,\ 0.2]$ (molec^{-1}).

From the results shown in Figure 10, notice that a maximum number of free ribosomes $r \approx 350$ can confidently explain the translation efficiencies per mRNA $Y_{p/\text{mRNA}}$ for almost all proteins while maintaining the value of the RBS strength-related term $K_{C^0}^k < 0.2$. This estimation for the amount of free ribosomes is in complete agreement with the estimation $r = \mathbf{N}(350, 35)$ in [23].

With the upper limit $r \approx 350$ we could explain the translation efficiencies per mRNA $Y_{p/\text{mRNA}}$ calculated as $\frac{\beta_{pk}}{d_{mk}}$ (see equation (101)) using the data in [19] but for a small set of 80 non-ribosomal proteins out of 3551 (2.25%). In 52 of them, this could be attributed to their extremely long-living transcripts. In the remaining 28 ones, to their very high translation efficiency per mRNA $Y_{p/\text{mRNA}}$ expected from their values of β_{pk} and d_{mk} given in [19]. This can be explained by rewriting (101) as:

$$Y_{p/\text{mRNA}} = 0.62\frac{\nu}{l_e}\frac{f(s_i)\frac{K_{C^0}^k(s_i)}{d_{mk}}r}{1+\mu_m f(s_i)\frac{K_{C^0}^k(s_i)}{d_{mk}}r} = \frac{\beta_{pk}}{d_{mk}} \tag{104}$$

Figure 11 shows a plot of the function (104), as a function of its argument $f(s_i)K_{C^0}(s_i)r/d_m$ for two mRNA degradation rates corresponding to short and long-living mRNAs and ribosomes densities in the range $l_e = [18, 31]$. Notice $Y_{p/\text{mRNA}} = \frac{\beta_{pk}}{d_{mk}}$ saturates at the maximum attainable value:

$$Y_{p/\text{mRNA,max}} = 0.62\frac{\nu}{l_e}\frac{1}{\mu_m} \tag{105}$$

Saturation is reached for lower values of the argument $f(s_i)K_{C^0}(s_i)r/d_m$ as the transcripts have longer half-lives, i.e. smaller values of the degradation rate d_m.

The few cases our model could not predict all have values of $\frac{\beta_{pk}}{d_{mk}}$ above $Y_{p/\text{mRNA,max}}$ for the values of ν and l_e used in the model.

On the other hand, it is interesting to notice that for very low values of $f(s_i)K_{C^0}(s_i)r/d_m$ such that $f(s_i)K_{C^0}(s_i)r \ll \frac{d_m}{\mu_m}$ we can approximate:

$$Y_{p/\text{mRNA}} \approx 0.62\frac{\nu}{l_e}f(s_i)\frac{K_{C^0}^k(s_i)}{d_{mk}}r = \frac{\beta_{pk}}{d_{mk}} \tag{106}$$

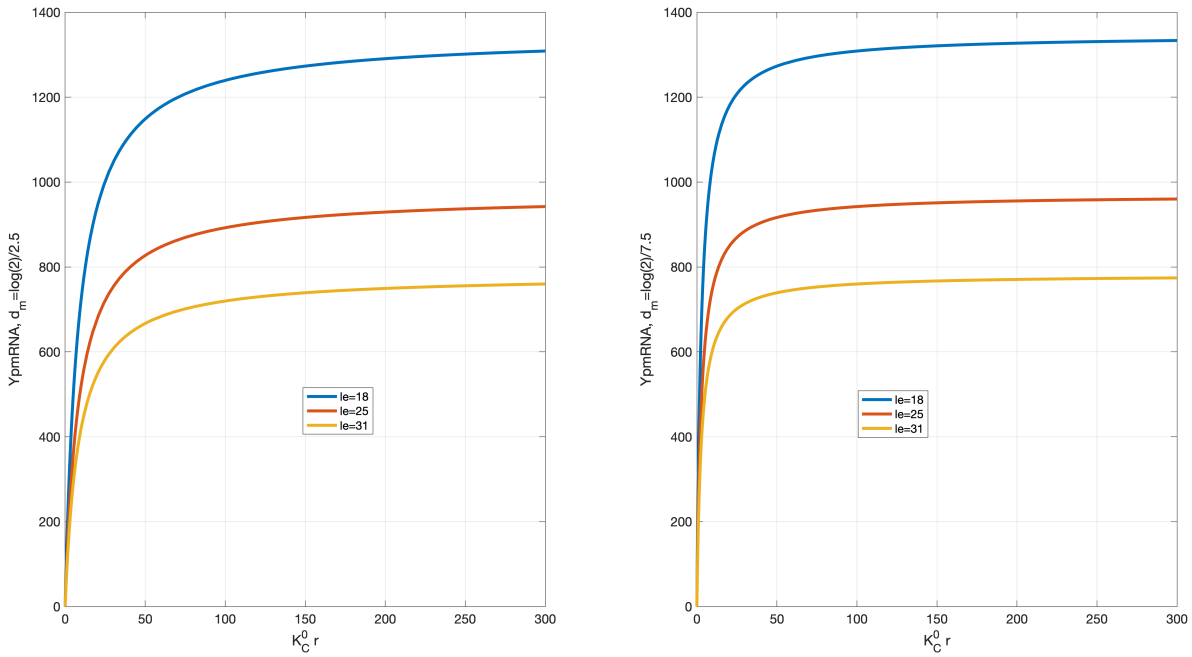

Figure 11: Translation efficiency per mRNA $Y_{p/\mathrm{mRNA}}$ as a function of $f(s_i)K_{C^0}r$ for two mRNA degradation rates corresponding to mRNA half-lives 2.5 minutes (left) and 7.5 minutes (right) for ribosomes densities corresponding to $l_e = \{18, 25, 31\}$ codons.

from which we get:

$$\beta_{pk} \approx 0.62\frac{\nu}{l_e} f(s_i)K^k_{C^0}(s_i)r \tag{107}$$

That is, for a highly competitive scenario where the number of free ribosomes is sufficiently small (e.g. in the order of few tens to few hundreds for typical values of d_{mk}, $\mu_m = 0.032\,\mathrm{min}^{-1}$ and $f(s_i)K_{C^0}(s_i)$ at its maximum estimated value $f(s_i)K_{C^0}(s_i) = 0.2$) the translation rate (proteins per mRNA per time unit) is proportional to the ribosomes density $\frac{0.62}{l_e}$, the effective maximum translation rate per codon attainable for a given substrate availability $\nu f(s_i)$, the RBS strength $K^k_{C^0}$, and the available free ribosomes r.

Notice under this scenario the translation rate will suffer large stochastic fluctuations caused by stochastic fluctuations in the number of free ribosomes. In this case, the transcription rate for a given RBS-strength mainly depends on the competition for cellular resources and, therefore, on the number of free ribosomes r, and it is largely independent of the specific growth rate.

16. Sensitivity of the resources recruitment strength to RBS and promoter can explain the variation of the cell mass distribution with growth rate

The relative distribution of ribosomal and non-ribosomal protein mass in the cell depends on the cell growth rate, so that the ribosome content increases linearly with growth rate [5, 35, 27, 9]. Existing resource allocation models explain this as a result of optimal allocation of cell resources between the ribosomal and protein (non-ribosomal)

fractions balancing the demands of protein synthesis and nutrient influx under the constraint that the sum of both fractions keeps constant [35].

We used the data in [5] of the ribosomal and non-ribosomal protein mass fractions as a function of growth rate and the corresponding effective translation rates to check whether our model is able to describe the linear increase of ribosomes content with growth rate.

In our model, the relative resources recruitment strength of a given protein equals its relative mass fraction in the cell at steady state, according to equation (72). Therefore, the relative distribution of ribosomal and non-ribosomal protein mass must be reflected in their relative distribution of resources recruitment strengths.

As a first step, we considered the lumped resources recruitment strengths (73)–(74) and estimated the RBS-strength related parameters K_b^k, K_u^k, the transcription rates ω_{k} with $k = \{r, \mathrm{wt}\}$ and the fraction Φ_t so that our model provides a good fit of the specific growth rate at steady state. Notice that we did not estimate the parameters in $J_r(\mu, r), J_{\mathrm{wt}}(\mu, r)$ to try and directly fit the experimental relative distribution of resources recruitment strengths. Instead, we were interested in checking whether a good fit of the specific growth rate implies ribosomal and non-ribosomal resources recruitment strengths such that their relative values fit the experimental ones. This, in turn, implies fitting the relative mass fractions in the cell.

We used the model expressions at steady state in Section 10 and the only input information given to the model was the values of the effective maximum translation rate ν_t as a function of growth rate obtained from [5]. We ran 200 instances of the parameter optimization using the optimization global optimization software MEIGO [12] (available at `http://gingproc.iim.csic.es/meigo.html`) and obtained the weighted mean of the 25 runs achieving the best minimum value for the sum over the experimental data points of the absolute growth rate prediction error. We considered $N_r = 57$; and $N_{wt} = 1735$, corresponding to the number of genes that explain 99% of the cumulative sum of the resources recruitments strengths for ribosomal and non-ribosomal proteins respectively (see Section 15). We also considered the average mRNA degradation rates $d_{m,r} = 0.16$ (1/min) and $d_{m,wt} = 0.2$. The resulting average best fit estimated parameters are given in Table 3.

K_u^r	K_u^{nr}	K_b^r	K_b^{nr}	ω_{r}	ω_{nr}	Φ_t
130.74	3.11	5.81	11.80	5.36	0.028	0.78

Table 3: Average best fit estimated parameters of the RBS-strength related parameters K_b^k, K_u^k, the transcription rates ω_{k} with $k = \{r, \mathrm{wt}\}$ and the fraction Φ_t of mature ribosomes with respect to the total number of ribosomes.

Figure 12 shows the results of the optimization. Notice the good agreement between the experimental and the estimated growth rate. Notice also that the estimation for the maximum number of free ribosomes $r = 350$ used in previous sections is consistent with model results for cells growing up to some $t_d = 25$ minutes ($\mu \approx 0.028$. For faster

growing cells, the number of free ribosomes much increases. Notice though, that also the total number of ribosomes (both experimental and estimated) much increases for very fast growing cells. Thus, the fraction of free ribosomes with respect to the total number only increases from 0.08% up to 1.37% for cell duplication times between 100 and 24 minutes respectively even though the number of free ribosomes multiplies by almost 200-fold (see Figure 13). Notice also the logarithmic affine relationship between the number of free ribosomes r and its flux μr, ($\log_{10}(r) \approx 4.11 + 0.79\log_{10}(\mu r)$) reflecting a power-law relationship between growth rate and number of free ribosomes.

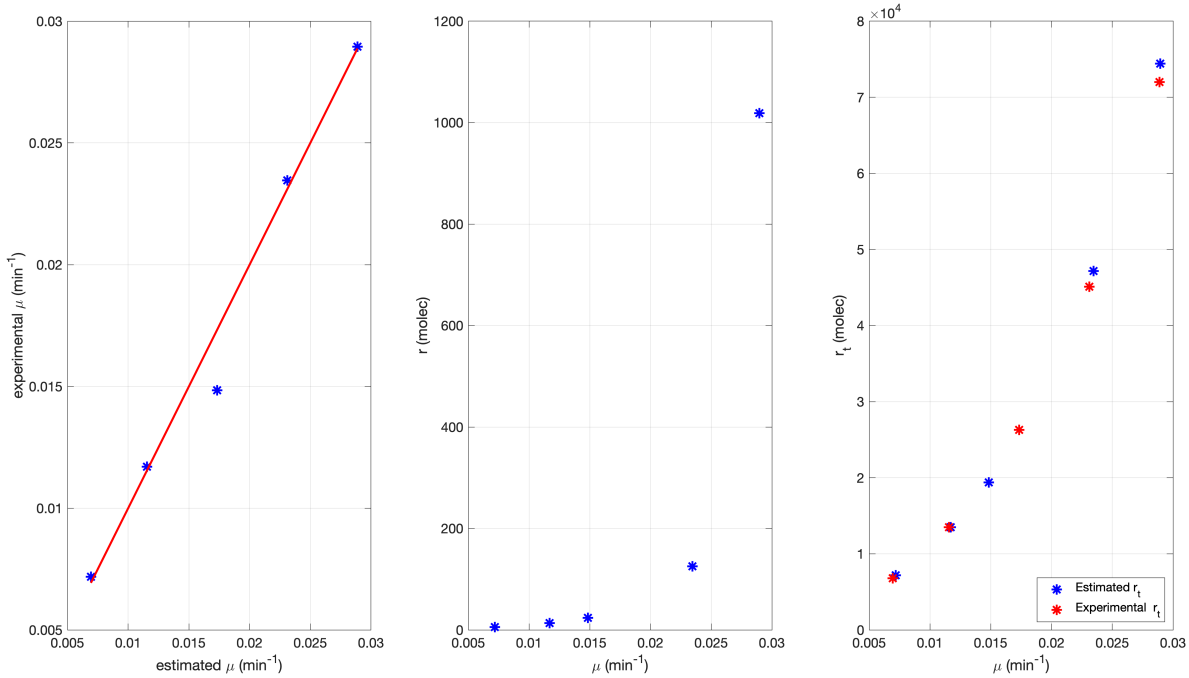

Figure 12: Estimated versus experimental growth rate (Left). Estimated number of free ribosomes (Center). Estimated and experimental number of total ribosomes (Right)

Notice the estimated value of the fraction Φ_t of mature ribosomes with respect to the total number of ribosomes. is very close to the average experimental value $\Phi_t = 0.8$ [4, 5]. Notice also the clear difference between the parameters corresponding to ribosomal proteins as compared to those of non-ribosomal ones. The results suggest that the ribosomal RBSs are much weaker than the non-ribosomal ones. Thus ribosomes have much less affinity for the ribosomal RBSs than for non-ribosomal RBSs. Figure 14 shows the estimated translation initiation rate k_e and the estimated effective RBS strengths $K^r_{C^0}$ and $K^{nr}_{C^0}$ as a function of the specific growth rate μ. As seen, the effective RBS strengths of non-ribosomal proteins is much higher than that of the ribosomal ones. Notice though, that as the growth rate increases –tantamount in our model to increasing intracellular substrate s_i– the ribosomal effective RBS strength $K^r_{C^0}$ keeps almost constant (with a slight decrease around 12%) while the non-ribosomal one decreases by almost a 40%. Notice also the clear difference between the terms $d_{mk}/K^k_{C^0}$

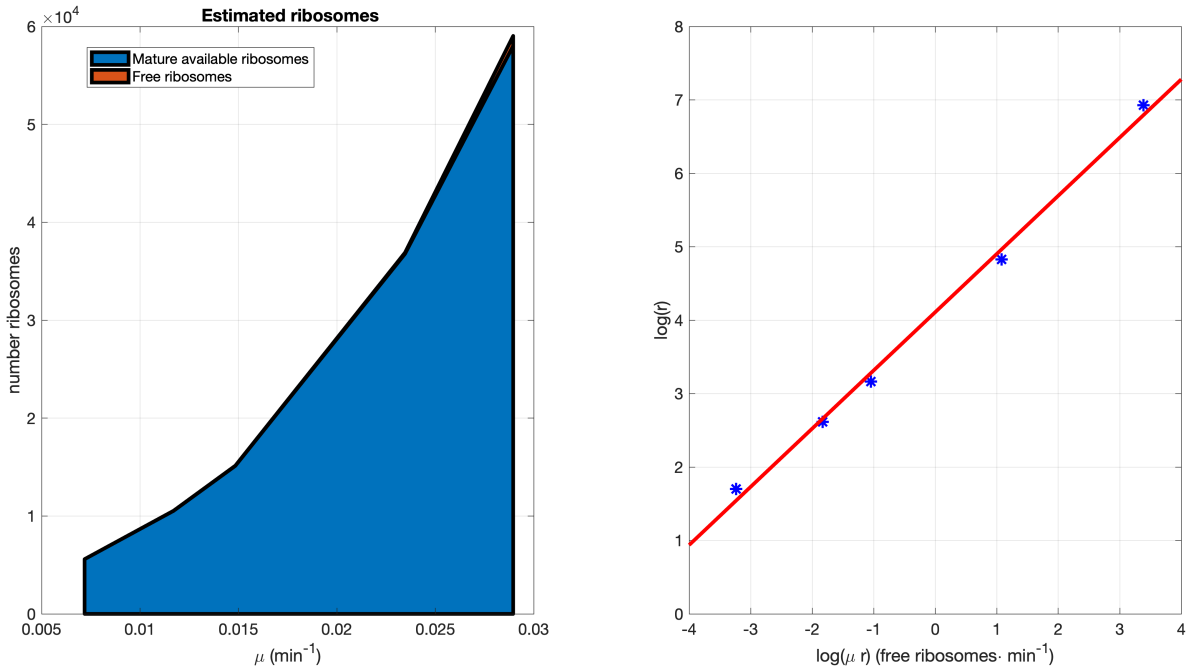

Figure 13: Estimated number of free and total ribosomes (Left) and logarithmic affine relationship between the number of free ribosomes r and its flux μr (Right)

for both ribosomal and non-ribosomal proteins when as a function of the flux of free resources μr (cf. the expression of the resources recruitment strengths given in equations (73) -(74)). The ribosomal proteins keep much higher values of $d_{mk}/K^k_{C^0}$ for all values of the flux of free resources μr. This, considering the monotonous increasing power-law relationship between the growth rate and the number of free ribsosomes, will imply that the value of the ribosomal resources recruitment strength $J_r(\mu, r)$ will decrease much slower than that of the non-ribosomal proteins as the growth rate increases. The plots in Figure 15 confirm this.

The differential behaviour between the ribosomal and non-ribosomal resources recruitment strengths is behind the differential mass distribution between ribosomal and non-ribosomal cell protein content as growth rate increases. Thus, the evaluation of the mass fractions using (79)-(80) provides the results shown in Figure 16 showing a very good agreement between the experimental values and the ones provided by the model.

Notice that the results above suggest that to have a robust expression rate as the growth rate increases, the optimal strategy consists of having a weak RBS and a strong promoter. Interestingly, the values we obtained for the transcription rates are in the same order of magnitude as the mean values obtained from the data in [19] –$\omega_r = 2.4$ and $\omega_{nr} = 0.05$ respectively– and show a much higher value for the average lumped ribosomal protein we considered than for the non-ribosomal one. The optimal productivity will depend on the trade-off between RBS and promoter strengths and their effect on growth rate, as analysed in the next sections.

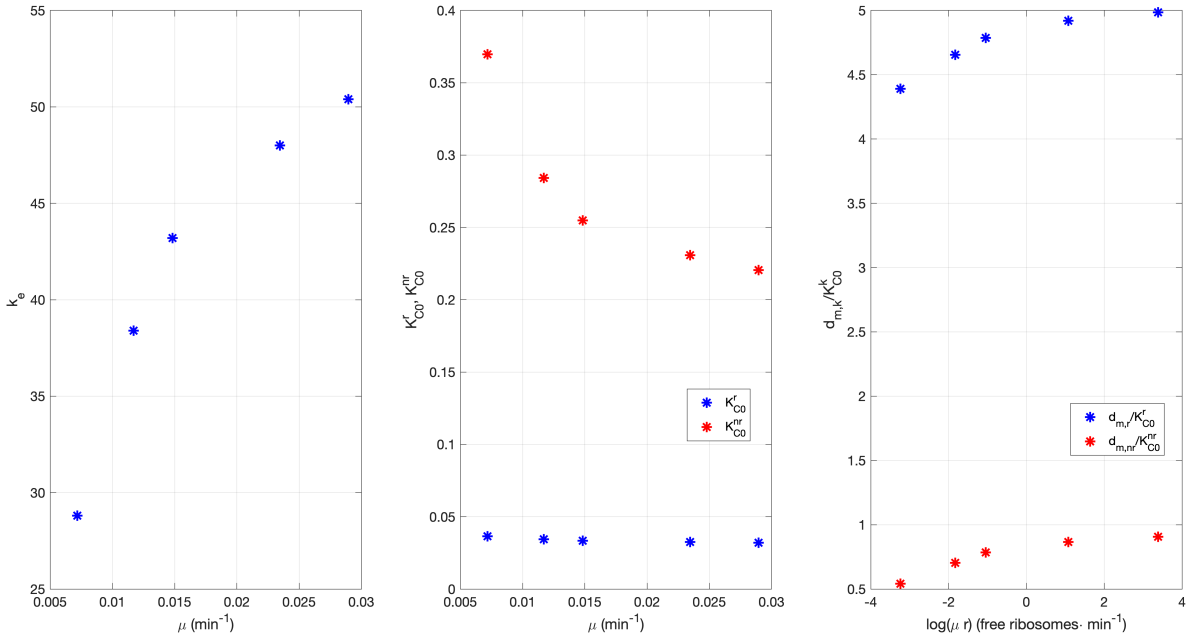

Figure 14: Estimated translation initiation rate k_e (Left) and estimated effective RBS strengths $K^r_{C^0}$ and $K^{nr}_{C^0}$ (Center) as a function of the specific growth rate μ. (Right): plot of $d_{mk}/K^k_{C^0}$ as a function of the flux of free resources μr. A log scale has been used.

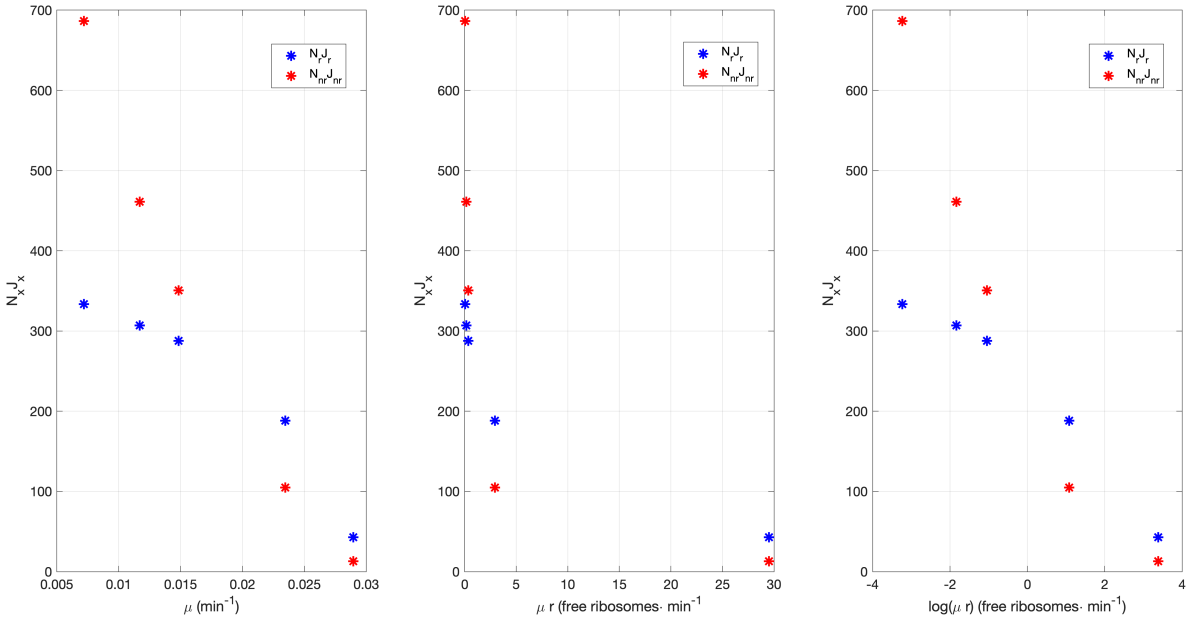

Figure 15: Estimated total resources recruitment strengths $N_r J_r$ and $N_{nr} J_{nr}$ as a function of growth rate μ(Left) and the free resources flux μr using linear (Center) and logarithmic scales (Right).

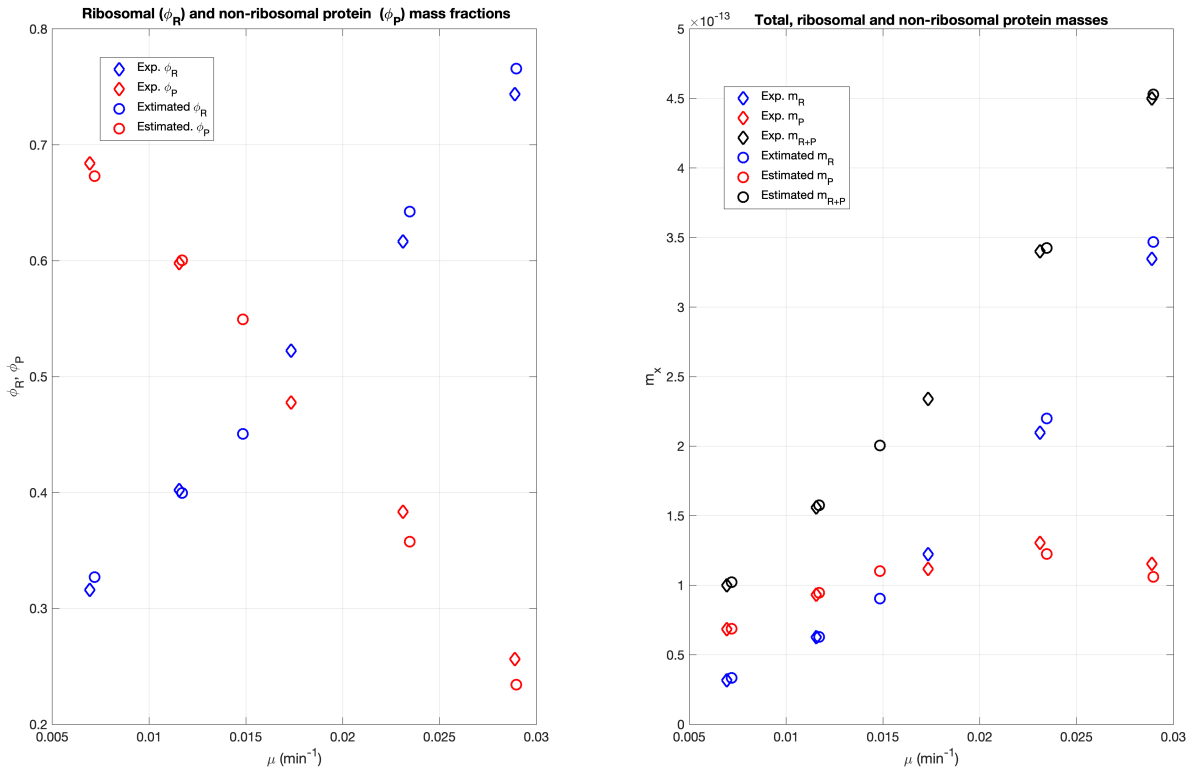

Figure 16: Estimated versus experimental mass fractions of ribosomal and non-ribosomal proteins in *E. coli* (left) and their absolute values along with the total protein cell mass (right).

17. The differential role of RBS and promoter strengths.

It is well known that varying combinations of transcription and translation rates affect the stability of metabolic networks [29] and the trade-off between desired expression levels and noise [19] and between expression of endogenous and synthetic genes and growth [43, 16].

We were interested in analysing the differential role of RBS and promoter strengths for endogenous proteins, that is, for the host ones, assuming there is no loading effect.

To this end, we analysed the protein abundance of an endogenous protein as a function of the of the expression space by varying the transcription rate w_k, and the RBS affinity for the mRNA ($K^k_{C^0}$ in our model). We are now focused on understanding the role of promoter and RBS strengths. Loading effects (that would be expected if overexpressing an exogenous protein, as considered in Section 18) will introduce interactions that will hide the differential effects of varying the transcription rate and the RBS affinity for the mRNA. To avoid this, we considered the case of expression of a protein A under the assumptions that $J_A \ll \sum J_{r,wt,A}(\mu, r)$ and $\sum J_{r,wt}(\mu, r)$ is constant (c.f. equations (88) and (89) in Section 11). This assumption holds for very low values of w_A. Under these conditions, the flux of free resources μr can be assumed to depend only on the constant value of $J_{r,wt}(\mu, r)$.

It is important to note that $J_A(\mu, r)$ does not directly define the expression of the

protein A, that is, the steady state value of the protein mass $\bar{m}_A$. Recall from Section 9 (see also equation(88)) that it is the relative value of the resources recruitment strength $J_A(\mu, r)$ with respect to the total sum $J_{r,wt,A}(\mu, r)$ which defines the fraction of the cell protein mass m_p that will be invested as protein A. Our assumptions imply constant $\sum J_{r,wt,A}(\mu, r)$ and constant flux of free resources μr. Under these conditions, from equation (85), the cell protein mass m_p will also keep constant as a function only of the intracellular substrate s_i. If in addition, we assume that the intracellular substrate s_i keeps constant, then a variation in the value of $J_A(\mu, r)$ will be monotonously related to a corresponding variation of $\bar{m}_A$.

Under the assumptions above, we can express:

$$J_A(\mu r) = \frac{k_A}{h_A} \frac{h_A}{h_A + \mu r} \tag{108}$$

with:

$$k_A = 0.62 \frac{l_p^A}{l_e} w_A \quad \text{and} \quad h_A = \frac{d_m^A}{K_{C^0}^A}. \tag{109}$$

where k_A/h_A is the maximum value that $J_A(\mu r)$ can attain and $h_A/(h_A + \mu r)$ is a Hill-like function which depends on the flow of free resources μr with h_A as half-activation threshold.

Next we analyse the effect of increasing the transcription rate w_A (for our purposes tantamount to increasing the transcription rate times the gene copy number) on the value of $J_A(\mu, r)$. Recall we introduce changes in w_A small enough so that our assumptions hold. As we vary w_A, the term k_A/h_A will vary in the same proportion and h_A will not be affected. Therefore, as the flow of free resources keeps constant under our conditions, the resources recruitment strength $J_A(\mu, r)$ will vary in proportional relationship to the variations in w_A. This, in turn, will affect in proportional relationship to the value of the mass $\bar{m}_A$. Therefore, under the condition of negligible metabolic load, variations in the transcription rate and the resulting variations in the amount of the expressed protein are proportional. Figure 17 (left) shows the agreement between this reasoning and the simulation results obtained with the full model.

As for variations in the RBS strength, consider for instance the effect of increasing $K_{C^0}^A$, that is, increasing the RBS-ribosome association rate k_b^A and decreasing the dissociation one k_u^A respectively. As we increase $K_{C^0}^A$, the term k_A/h_A will increase in the same proportion. Yet, the term $h_A/(h_A + \mu r)$ will decrease. Which one of both opposite effects will determine the evolution of $J_A(\mu r)$ will depend on the relative weights of h_A and the flux of free resources μr. If $h_A \gg \mu r$, an increment in $K_{C^0}^A$ (compatible with the condition $h_A \gg \mu r$) will increase the steady state value of the protein mass m_A, since k_A/h_A will grow and $h_A/(h_A + \mu r)$ will remain mostly constant. However, if we keep $K_{C^0}^A$ increasing we will reach a situation where $h_A \ll \mu r$. In the limit, equation 108 can be approximated as:

$$J_A(\mu r) = \frac{k_A}{\mu r} \tag{110}$$

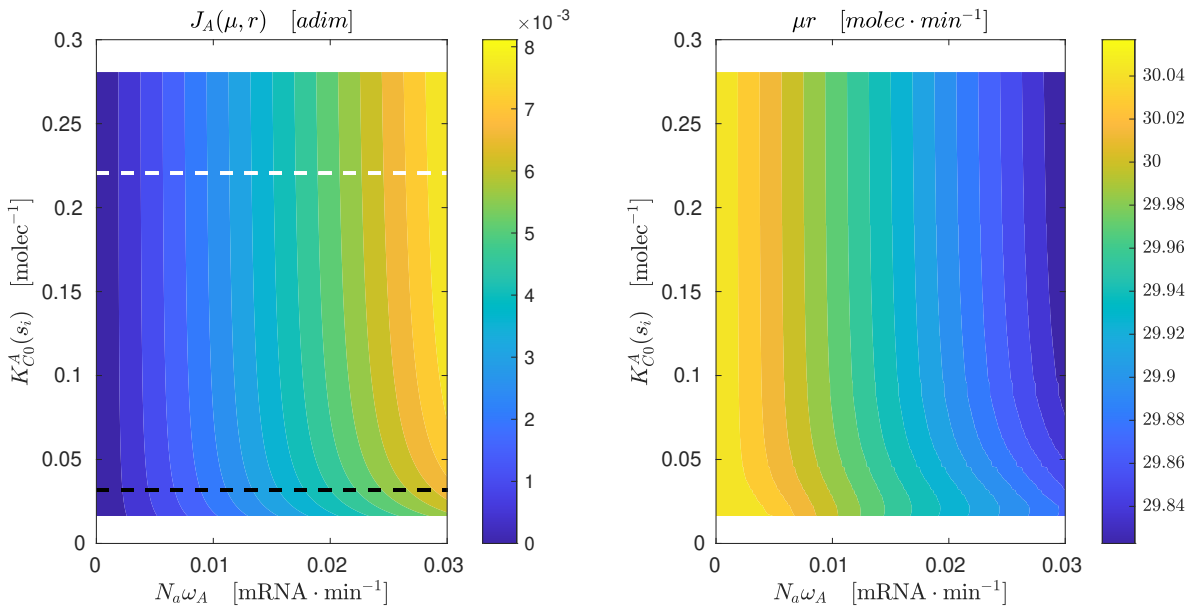

Figure 17: Effect of increasing transcription rate and RBS strength on the value of $J_A(\mu, r)$ (left) and on the value of μr (right). The y-axis shows the full experimental range of $K^A_{C^0}$ for $f(s_i) = 1$, that is, saturated substrate. The average RBS strength value for *E. coli* ribosomal and non-ribosomal proteins are shown as black dashed and white dashed lines respectively. The x-axis shows the transcription rate values around the experimental average value of *E. coli* non-ribosomal genes ($0.028\,\text{min}^{-1}$).

showing that at high RBS strength the resources recruitment strength $J_A(\mu r)$ saturates. When this saturation occurs, the protein mass $\bar{m}_A$ becomes independent of $K^A_{C^0}$. In summary, increasing the RBS strength has a limited effect to increase the expression of the protein A, saturating at high RBS strengths. Figure 17 confirms this general rule in the simulation results obtained with the full model.

In summary, RBS strength (together with mRNA degradation d^A_m) determines the sensitivity of the resources recruitment strength $J_A(\mu, r)$ of a given protein A with respect to the flux of free resources μr. Thus, if protein A is expressed using a weak RBS, the value of $J_A(\mu, r)$ will remain mostly constant with respect to variations of μr. On the contrary, if A is expressed with a strong RBS, the resources recruitment strength $J_A(\mu, r)$ will decrease as μr increases. This result is confirmed in Figure 18 which shows the effect of varying μr on the value of $J_A(\mu, r)$ as a function of the RBS strength.

As observed in Figure 18, there is a trade-off between increasing the expression of a protein by increasing the strength of RBS and increasing robustness with respect to μr by decreasing the RBS strength. It is not possible to achieve high expression and high robustness of the resources recruitment strength simply by adjusting the RBS strength.

This trade-off is consistent with the optimization results obtained in Section 16. Recall the ribosomal and non-ribosomal RBS parameters in Table 3, where ribosomal genes have low RBS strength (i.e. low k^r_b and high k^r_u) and high transcription rate, while the non-ribosomal genes have high RBS strength (i.e. high k^r_b and low k^r_u) and low transcription rate. On the one hand, ribosomal genes have low RBS strength and

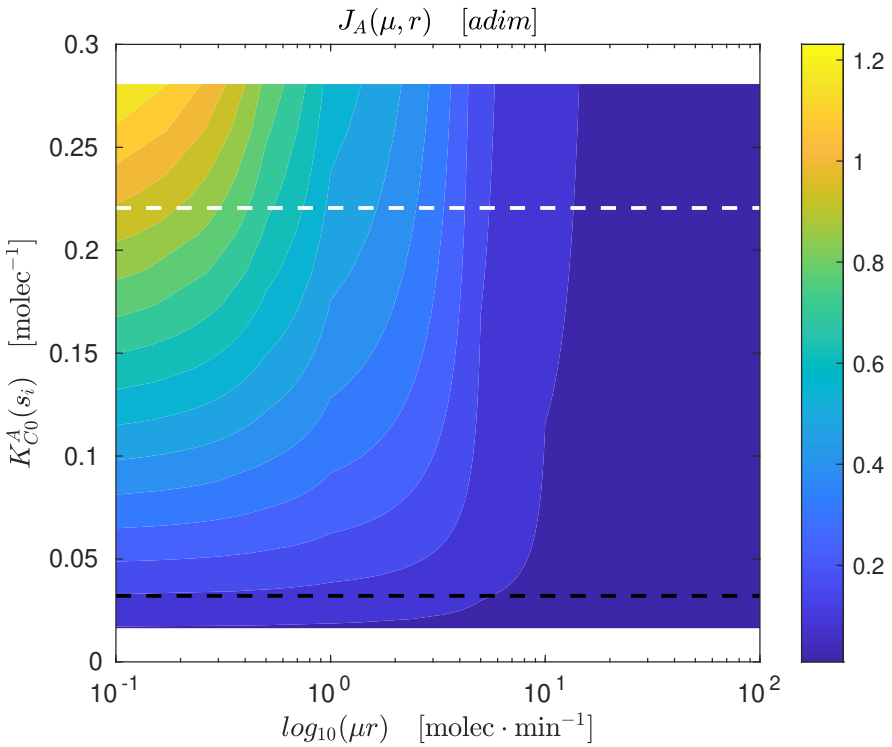

Figure 18: Effect of varying flux of free resources μr on the value of $J(\mu, r)$ with $k_A = 1$ and $d_m^A = 0.2\,\text{min}^{-1}$. The y-axis shows the full experimental range of $K^A_{C^0}$ for $f(s_i) = 1$, that is, saturated substrate. The RBS strenght value of ribosomal (black dashed) and non-ribosomal (white dashed) is shown. The x-axis shows a sweep in μr values for the range obtained in previous simulations.

high transcription rate, so their $J_r(\mu, r)$ will remain mostly constant with respect to changes in the flow of free resources μr. On the other hand, non-ribosomal genes have high RBS strength and a low transcription rate. Therefore their expression will depend on the flow of free resources μr. As μr increases, the value of $J_{nr}(\mu, r)$ will decrease. This difference in the values of RBS strength and transcription rate, long with the equivalence between the relative value of mass fractions $m_{r,nr}$ and the relative value of $J_{r,nr}(\mu, r)$, allow us to explain the relative mass fractions distributions of ribosomal and non-ribosomal proteins we showed in Figure 16 at different growth rates. As seen there, the relative value of $J_{nr}(\mu, r)$ decreases as μ increases, as predicted by the high RBS strength of non-ribosomal proteins. However, the relative value of $J_r(\mu, r)$ increases as μ increases. This is possible because the value of $J_r(\mu, r)$ remains mostly constant with respect to $J_{nr}(\mu, r)$ (or at least it decreases less than $J_{nr}(\mu, r)$ does). Therefore, the relative value of $J_r(\mu, r)$ increases.

In summary, the RBS strength sets the sensitivity of the resources recruitment strength with respect to the flux of free resources. The resources recruitment strength $J_k(\mu, r)$ of a given protein, related to its expression, can be made dependent on the flux of free resources μr (strong RBS) or can be robust with respect to variations of μr (weak RBS). This allows to set how much of a given protein (e.g ribosomal or non-ribosomal)

will be expressed at different levels of μr (i.e. at different growth rates).

This differential expression may have evolved in *E. coli* and other micro-organisms to encode the mass distribution of ribosomal and non-ribosomal proteins as shown in Figure 16 (right). The cell achieves a fairly constant absolute expression of non-ribosomal proteins by using a high RBS strength to express them. This way, as the growth rate increases, (tantamount to μr) the value of $J_{nr}(\mu, r)$ will decrease, compensating the increase of expression induced by the increase of the total cell protein mass. On the other hand, the cell uses much weaker RBSs to express the ribosomal proteins. This way the value of $J_r(\mu, r)$ will remain mostly constant with respect to $J_{nr}(\mu, r)$. As a consequence, the absolute expression of ribosomal proteins increases with growth rate as the value of $J_{nr}(\mu, r)$ decreases, causing the fraction of ribosomes to increase.

18. Host-circuit interaction shapes the optimal productivity of proteins

There are multiple ways to increase the expression of a given exogenous protein, including the choice of the expression vector, optimizing the use of codons, co-expression of chaperones to aid protein folding, etc. [32]. Here we pay attention to the differential role of RBS and promoter strengths in determining protein abundance. We evaluated the mass and specific productivities across the expression space $N_A\omega_A, K^A_{C^0}$ for a given exogenous protein A at different values of intracellular substrate availability using the average host dynamics at steady state and the evaluation of protein mass productivity derived in sections 10 and 11.

First we considered the average values of the RBS-strength related parameters K^A_b, K^A_u obtained in section 16 for a non-ribosomal protein (see Table 3) and we varied the promoter strength times the gene copy number $N_A\omega_A$ in the range $[0.1, 12500]$ times the average value of the non-ribosomal promoter strengths obtained from the data in [19]. This gives maximum value $N_A\omega_A = 375$ (mRNA $\cdot$ min^{-1}) well within the average transcription rate in *E. coli* if we consider a gene copy number around 100 (e.g. a medium-high copy number plasmid). Figure 19 shows the results obtained. As expected, the cell growth rate abruptly decreases as the mass fraction of the exogenous protein increases for increasing values of $N_A\omega_A$. Notice that the mass productivity of the exogenous protein reaches a maximum at $N_A\omega_A \approx 30$. Thus, for a constitutive promoter with high transcription rate $\omega_A \approx 3$ (mRNA $\cdot$ min^{-1}) the maximum productivity is achieved for a gene copy number $N_A \approx 10$ typical of low copy number plasmids. Interestingly, this maximum is independent of the substrate availability, and only depends on the promoter strength and gene copy number.

Next we considered two values $N_A\omega_A = \{3, 150\}$ corresponding to low and high promoter strength times the gene copy number and we varied K^A_b, K^A_u in the ranges considered in section 15 so as to achieve a continuous varying RBS strength $K^A_{C^0}$. Figure 20 shows the results obtained. Again, the cell growth rate abruptly decreases as the mass fraction of the exogenous protein increases for increasing values of the RBS-strength. Notice though that the main affecting factor is the promoter strength times gene copy number $N_A\omega_A$. Thus, for low values of $N_A\omega_A$, increasing RBS strength gives rise to

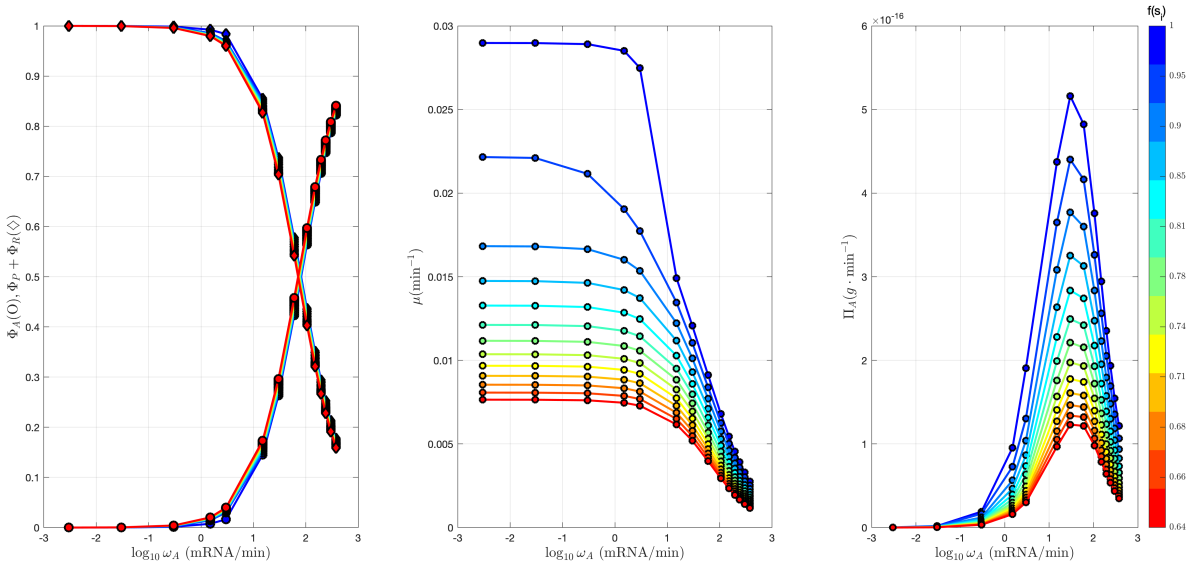

Figure 19: Effect of promoter and gene copy number variation on protein ∘ and cell ♢ mass fractions (Left) cell growth rate (Center) and mass productivity (Right) for varying intracellular substrate concentration.

increasing protein mass productivity. Only for high values of $N_A\omega_A$ there appears a maximum of the mass productivity as a function of the RBS strength. Notice also that in this case, the maximum also varies as a function of the substrate availability.

Finally, Figure 21 shows the variation of protein mass productivity across the expression space $N_A\omega_A, K^A_{C^0}(s_i)$ for different values of substrate availability. Notice the optimal productivity subspace corresponds to a hyperbola in the $(N_A\omega_A, K^A_{C^0}(s_i))$-space showing a trade-off between promoter and RBS strengths. The figure also shows the average values of $N_x\omega_x, K^x_{C^0}(s_i)$ of the *E. coli* host non-ribosomal and ribosomal proteins respectively. Interestingly, ribosomal and non-ribosomal proteins lay in opposite branches of the optimal subspace. While the host ribosomal proteins have evolved a expression strategy based on having a low RBS strength and high promoter strength, the non-ribosomal ones have evolved the opposite strategy. Notice also that while for ribosomal proteins the values of $N_x\omega_x, K^x_{C^0}(s_i)$ do not appreciably change as a function of substrate, for non-ribosomal proteins the effective RBS strength $K^x_{C^0}(s_i)$ increases as the substrate availability decreases so that their expression keep within the optimal subspace. Therefore, in agreement with our findings in Section 17 the weak promoter-strong RBS strategy versus the strong promoter-weak RBS one is not determined by optimal productivity, but by the strategic decision on the absolute expression should increase with growth rate or not.

19. Conclusions

In this work we have presented a small-size model of gene expression dynamics accounting for host-circuit interactions. The good agreement between the predictions

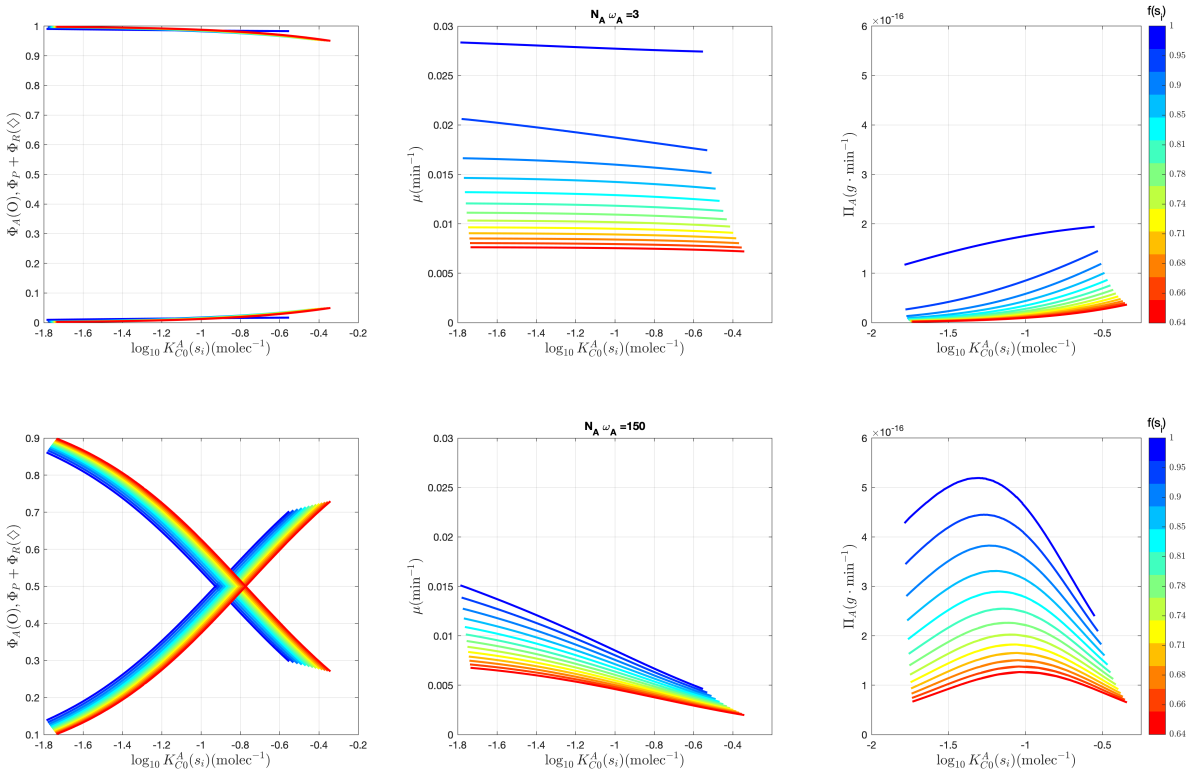

Figure 20: Effect of RBS strength variation on protein and cell mass fractions (Left) cell growth rate (Center) and mass productivity (Right) for varying intracellular substrate concentration and the two values $N_A\omega_A = \{3, 150\}$.

of our model and experimental data highlight the relevance of the cellular resources recruitment strength defined in our model as a key functional coefficient that allows to explain the distribution of resources among the host and the genes of interest and the relationship between the usage of resources, cell growth and protein productivity. This functional coefficient explicitly takes into account the interplay between the flux of available free resources and lab-accessible gene characteristics. In particular, the promoter and RBS strengths.

Though we fitted the model to *E. coli*, our findings can be extrapolated to other microorganisms, and the model can be easily fitted using a small amount of experimental data of the host cell.

Among other predictions, the model provides insights into how the differential role of promoter and RBS strengths in protein expression may have evolved in *E. coli* and other micro-organisms to encode the mass distribution between ribosomal and non-ribosomal proteins as a function of cell growth rate. Weak transcription and strong translation and the complementary strong transcription and weak translation arise a two equally optimal protein productivity strategies in the expression space but with different characteristics from the point of view of the resulting copy number of the expressed protein as a function of growth rate. The capacity of the defined resources recruitment strength functional

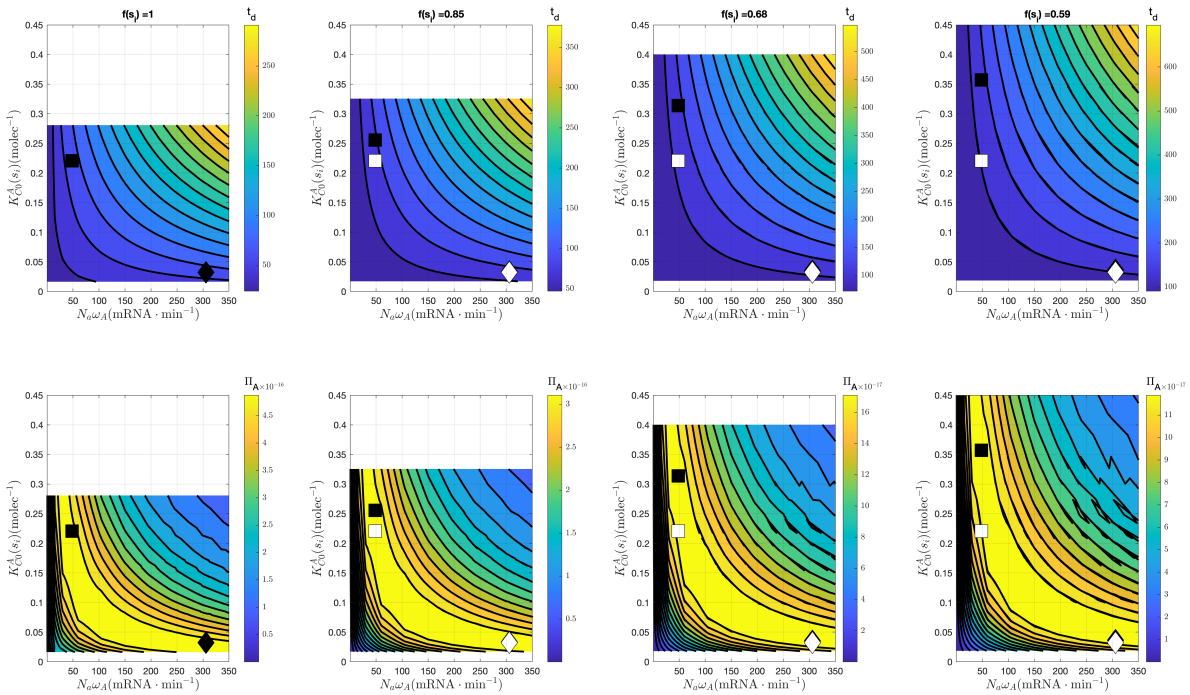

Figure 21: Cell duplication time (top row) and protein mass productivity (bottom row) across the expression space $N_A\omega_A, K^A_{C^0}$ for different values of substrate availability. The black symbols ($\square, \diamondsuit$) correspond to the average values of $N_x\omega_x, K^x_{C^0}(s_i)$ of the *E. coli* host non-ribosomal and ribosomal proteins respectively (see Table 3). The white symbols correspond to the value of $N_x\omega_x, K^x_{C^0}(s_i)$ for saturated substrate.

coefficients to capture the interaction between growth, cell resources and gene expression characteristics is reflected in the fact that we did not estimate their parameters by trying and directly fitting the available experimental relative distribution of resources recruitment strengths. Instead, we sought the model to fit the cell specific growth rate and saw that this implied ribosomal and non-ribosomal resources recruitment strengths such that their relative values fitted the experimental ones.

The model also explains some of the phenomena typically encountered when building protein expression systems in synthetic biology. Thus, for instance, it explains the limited effect that increasing the RBS strength has to increase the expression of a given protein of interest, saturating at high RBS strengths. In this context, our model may also be useful for design purposes in synthetic biology where it can be used to design the proper strategy in the expression space depending on the desired expression behaviour as a function of growth rate.

Further extensions of the model can be easily implemented. Thus, the model explicitly considers the relationship between the cell specific growth rate and the population dynamics, so it can be integrated within a multi-scale framework that considers the macroscopic extracellular dynamics of the substrate and population of cells in a bioreactor. The possibility to consider expression systems using orthogonal ribosomes can also be implemented without much difficulties.

20. Acknowledgements

This work was partially supported by grant MINECO/AEI, EU DPI2017-82896-C2-1-R. F.N. Santos-Navarro is grateful to grant PAID-01-2017 (Universitat Politècnica de València)

21. Authors contributions

J. Picó conceived the study. J. Picó and F.N. Santos-Navarro designed the methods and experiments. F.N. Santos-Navarro implemented the methods and performed the experiments. J. Picó and F.N. Santos-Navarro analysed the results and wrote the manuscript. J. Picó edited and approved the final manuscript.

22. Declaration of interests

The authors declare no competing interests.

References

[1] Jonathan A. Bernstein, Arkady B. Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N. Cohen. Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *Proceedings of the National Academy of Sciences*, 99(15):9697–9702, 2002.

[2] Andrew E. Blanchard, Chen Liao, and Ting Lu. Circuit-host coupling induces multifaceted behavioral modulations of a gene switch. *Biophysical J.*, 114:737–746, 2018.

[3] Y. Boada, A. Vignoni, D. Oyarzún, and J. Picó. Host-circuit interactions explain unexpected behavior of a gene circuit. *IFAC-PapersOnLine*, 51(19):86 – 89, 2018.

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

[5] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. *EcoSal Plus*, 2008.

[6] M. Carbonell et al. Dealing with the genetic load in bacterial synthetic biology circuits: covergences with the ohm's law. *Nucleic Acids Research*, 44(1):496–507, 2016.

[7] S. Cardinale and A.P. Arkin. Contextualizing context for synthetic biology- identifying causes of failure of synthetic biological systems. *Biotechnol. J.*, 7(7):856–866, 2012.

[8] Huiyi Chen, Katsuyuki Shiroguchi, Hao Ge, and Xiaoliang Sunney Xie. Genome-wide study of mRNA degradation and transcript elongation in E scherichia coli . *Molecular Systems Biology*, 11(1):781, 2015.

[9] Xiongfeng Dai and Manlu Zhu. Coupling of Ribosome Synthesis and Translational Capacity with Cell Growth. *Trends in Biochemical Sciences*, 45(8):681–692, 2020.

[10] V. de Lorenzo. Evolutionary tinkering vs. rational engineering in the times of synthetic biology. *Life Sci Soc Policy*, 14(18):1263–1272, 2018.

[11] Maarten H. de Smit and Jan van Duin. Translational standby sites: How ribosomes may deal with the rapid folding kinetics of mrna. *Journal of Molecular Biology*, 331(4):737 – 743, 2003.

[12] Jose A Egea, David Henriques, Thomas Cokelaer, Alejandro F Villaverde, Aidan MacNamara, Diana-Patricia Danciu, Julio R Banga, and Julio Saez-Rodriguez. MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinformatics*, 15(1):136, 2014.

[13] Mette Eriksen, Kim Sneppen, Steen Pedersen, and Namiko Mitarai. Occlusion of the ribosome binding site connects the translational initiation frequency, mrna stability and premature transcription termination. *Frontiers in Microbiology*, 8:362, 2017.

[14] L.D. Fernandes, A.P.S.d. Moura, and L. Ciandrini. Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Sci Rep*, 7:17409, 2017.

[15] T Frei, F Cella, F Tedeschi, J Gutierrez, GB Stan, M Khammash, and V Siciliano. Characterization, modelling and mitigation of gene expression burden in mammalian cells. *bioRxiv*, 2019.

[16] Thomas E. Gorochowski, Irem Avcilar-Kucukgoze, Roel A. L. Bovenberg, Johannes A. Roubos, and Zoya Ignatova. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth. Biol.*, 5(7):710–720, 2016.

[17] Peter Großmann, Anja Lück, and Christoph Kaleta. Model-based genome-wide determination of rna chain elongation rates in escherichia coli. *Scientific Reports*, 7:17213, 2017.

[18] J. Guillaume, A. Goelzer, S. Tebbani, D. Dumur, and V. Fromion. Dynamical resource allocation models for bioreactor optimization. *IFAC PapersOnLine*, 51(19):20–23, 2018.

[19] Jean Hausser, Avi Mayo, Leeat Keren, and Uri Alon. Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*, 10(1):68, 2019.

[20] Kubitschek HE., Baldwin WW, Schroeter SJ., and Graetzer R. Independence of buoyant cell density and growth rate in escherichia coli. *J Bacteriol.*, 158(1):296–299, 1984.

[21] S Jayanthi, KS Nilgiriwala, and D Del Vecchio. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synth. Biol.*, 2(8):431–441, 2013.

[22] M Kanehisa and S Goto. Kegg: kyoto encyclopedi of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[23] Andrzej M. Kierzek, Jolanta Zaim, and Piotr Zielenkiewicz. The Effect of Transcription and Translation Initiation Frequencies on the Stochastic Fluctuations in Prokaryotic Gene Expression. *Journal of Biological Chemistry*, 276(11):8165–8172, 2001.

[24] Yu Liu. Overview of some theoretical approaches for derivation of the monod equation. *Appl Microbiol Biotechnol*, 73:1241–1250, 2007.

[25] Jason Lloyd-Price, Sofia Startceva, Vinodh Kandavalli, Jerome G. Chandraseelan, Nadia Goncalves, Samuel M. D. Oliveira, Antti H'akkinen, and Andre S. Ribeiro. Dissecting the stochastic transcription initiation process in live escherichia coli. *DNA Research*, 23(3):203–214, 03 2016.

[26] Derek N. Macklin, Travis A. Ahn-Horst, Heejo Choi, Nicholas A. Ruggero, Javier Carrera, John C. Mason, Gwanggyu Sun, Eran Agmon, Mialy M. DeFelice, Inbal Maayan, Keara Lane, Ryan K. Spangler, Taryn E. Gillies, Morgan L. Paull, Sajia Akhter, Samuel R. Bray, Daniel S. Weaver, Ingrid M. Keseler, Peter D. Karp, Jerry H. Morrison, and Markus W. Covert. Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science*, 369(6502), 2020.

[27] Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6:e28034, aug 2017.

[28] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(Database issue):D750–3, January 2010.

[29] Diego A. Oyarzún and Guy Bart V. Stan. Synthetic gene circuits for metabolic control: Design trade-offs and constraints. *Journal of the Royal Society Interface*, 10(78), 2013.

[30] F. Picard, P. Loubire, and L. et al. Girbal. The significance of translation regulation in the stress response. *BMC Genomics*, 14(588), 2013.

[31] Yili Qian, Hsin-Ho Huang, José I. Jiménez, and Domitilla Del Vecchio. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.*, 6(7):1263–1272, 2012.

[32] Germán L. Rosano and Eduardo A. Ceccarelli. Recombinant protein expression in escherichia coli: advances and challenges. *Frontiers in Microbiology*, 5:172, 2014.

[33] Renana Sabi and Tamir Tuller. Modeling and measuring intracellular competition for finite resources during gene expression. *J. R. Soc. Interface*, 16:20180887, 2019.

[34] Howard M. Salis. Chapter two - the ribosome binding site calculator. In Christopher Voigt, editor, *Synthetic Biology, Part B*, volume 498 of *Methods in Enzymology*, pages 19 – 42. Academic Press, 2011.

[35] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, 2010.

[36] Gilad Shaham and Tamir Tuller. Genome scale analysis of escherichia coli with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation. *DNA Research*, 25(2):195–205, 2017.

[37] Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou, Amy B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, and Suckjoon Jun. Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Current Biology*, 27(9):1278–1287, 2017.

[38] M Siwiak and P Zielenkiewicz. Transimulation - protein biosynthesis web service. *PLoS ONE*, 8(9):e73943, 2013.

[39] S. Sundararaj, A. Guo, habibi Nazhad B., Rouani M., Stothard P., Ellison M., and Wishart DS. The cybercell database (ccdb): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of escherichia coli. *Nucleic Acids Research*, 32(suppl1):D293–D295, 01 2004.

[40] Benjamin D. Towbin, Yael Korem, Anat Bren, Shany Doron, Rotem Sorek, and Uri Alon. Optimality and sub-optimality in a bacterial growth law. *Nature Communications*, 8:14123, 2017.

[41] Heykel Trabelsi, Mathilde Koch, and Jean-Loup Faulon. Building a minimal and generalizable model of transcription factorised biosensors: Showcasing flavonoids. *Biotechnology and Bioengineering*, 115:2292–2304, 2018.

[42] Pablo Ugalde-Salas, Elie Desmond-Le Quéméner, Jérôme Harmand, Alain Rapaport, and Théodore Bouchez. Insights from microbial transition state theory on monod's affinity constant. *Scientific Reports*, 10(1):5323, 2020.

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

[44] Hong Zeng and Aidong Yang. Bridging substrate intake kinetics and bacterial growth phenotypes with flux balance analysis incorporating proteome allocation. *Scientific Reports*, 10(1):4283, 2020.

[45] Hai Zheng, Yang Bai, Meiling Jiang, Taku A. Tokuyasu, Xiongliang Huang, Fajun Zhong, Yuqian Wu, Xiongfei Fu, Nancy Kleckner, Terence Hwa, and Chenli Liu. General quantitative relations linking cell growth and the cell cycle in Escherichia coli. *Nature Microbiology*, 2020.