ACS SyntheticBiology

pubs.acs.org/synthbio

Research Article

RBS and Promoter Strengths Determine the Cell-Growth-Dependent Protein Mass Fractions and Their Optimal Synthesis Rates

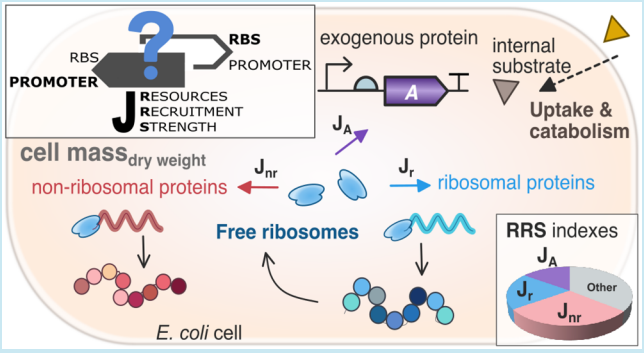
Fernando N. Santos-Navarro, Alejandro Vignoni, Yadira Boada, and Jesús Picó*

Cite This: *ACS Synth. Biol.* 2021, 10, 3290–3303

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Models of gene expression considering host−circuit interactions are relevant for understanding both the strategies and associated trade-offs that cell endogenous genes have evolved and for the efficient design of heterologous protein expression systems and synthetic genetic circuits. Here, we consider a small-size model of gene expression dynamics in bacterial cells accounting for host−circuit interactions due to limited cellular resources. We define the cellular resources recruitment strength as a key functional coefficient that explains 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 lab-accessible gene expression characteristics, such as promoter and ribosome binding site (RBS) strengths, capturing their interplay with the growth-dependent flux of available free cell resources. Despite its simplicity, the model captures the differential role of promoter and RBS strengths in the distribution of protein mass fractions as a function of growth rate and the optimal protein synthesis rate with remarkable fit to the experimental data from the literature for *Escherichia coli*. This allows us to explain why endogenous genes have evolved different strategies in the expression space and also makes the model suitable for model-based design of exogenous synthetic gene expression systems with desired characteristics.



KEYWORDS: *gene expression, burden, resources allocation, growth rate, RBS strength, promoter strength, protein synthesis mass fractions*

1. INTRODUCTION

The interrelations among the cell environment from which the cell uptakes substrates, its metabolism, and the engagement of cell resources needed for gene expression result in host−circuit interactions between gene circuits and their cell host. These interactions induce competition for common shared cell resources affecting gene expression and cell growth. Endogenous genes have evolved different strategies to deal with the problem of optimal protein expression under different needs and cell conditions.[1] As for exogenous genes, one of the fundamental problems in the rational design of synthetic genetic circuits of increasing complexity, partly explaining the current disparity between the ability to design biological systems their actual experimental performance, is the lack of systematic design methods considering the host−circuit interaction.[2] Cells have reached an optimal use of their resources during evolution. The overexpression of exogenous genes by a genetically modified microorganism as well as the production of metabolites by the addition and/or modification of their metabolic pathways introduce a metabolic load that takes the microorganism off its natural state.[3] The resulting competition for common shared cell resources affects cell growth and introduces spurious dynamics,[4] leading to problems of malfunctioning of the synthetic circuit. It also triggers its elimination by evolutionary mechanisms trying to restore the natural optimal state.[5]

Mathematical models of gene expression accounting for cellular resources competition can be used to decipher the mechanisms underneath gene expression strategies that have evolved to optimize different criteria. This is not only useful to understand natural systems but also addresses the rational design of synthetic genetic circuits. Therefore, in the last years, there has been an increasing interest in the development of models and methods for model-based design of synthetic gene circuits considering host−circuit interactions.[6]

The simplest burden-aware models deal with the interactions among genes in a gene network and consider shared cell resources as an external source, without considering the host behavior. This approach has proved very useful to deal with the so-called retroactivity,[7] the loading interaction among circuit modules and host originated from mass exchange.

Received: March 31, 2021
Published: November 12, 2021

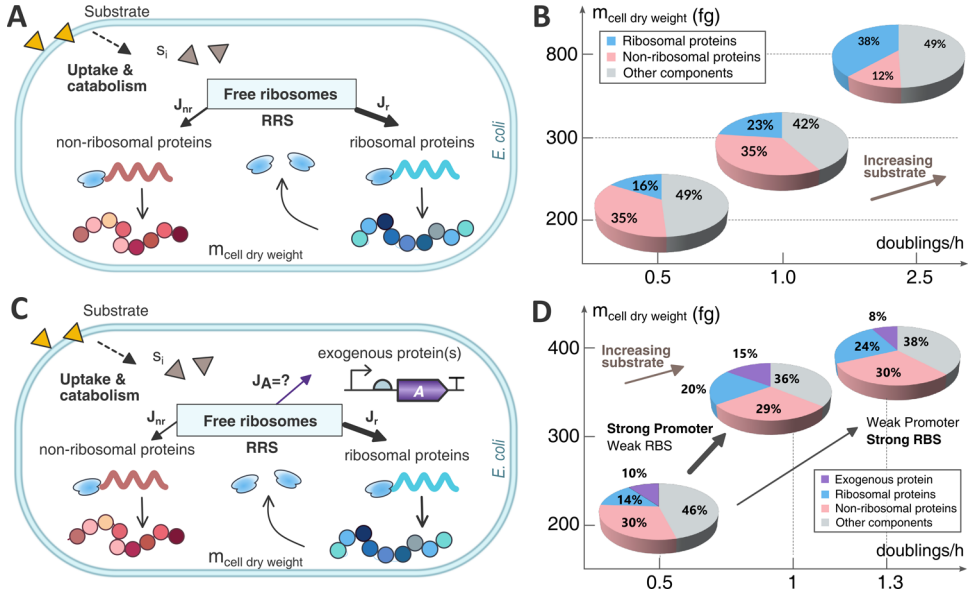


Figure 1. (A, C) Schematic view of the partitioned use of cell resources (ribosomes) to synthesize both ribosomal and nonribosomal proteins. The ribosomal proteins generate functional ribosome molecules to serve as the fundamental resource for protein synthesis. The resources recruitment strength (RRS) coefficient explains how the cell resources are allocated among the host endogenous genes (A) and both endogenous and exogenous genes of interest (C) (e.g., exogenous protein(s) expressed by a synthetic genetic circuit). (B) Resource allocation in a host cell in terms of the fractions of cell dry weight for the ribosomal proteins, nonribosomal ones, and other components. The pie charts represent different resources allocation scenarios, with increasing growth rates when the available substrate is increasing in the *x*-axis, and the resulting cell dry weight in the *y*-axis. (D) Resource allocation for a host cell expressing an exogenous protein. Two strategies were used for expressing the exogenous protein: strong promoter with weak RBS and weak promoter with strong RBS. The pie charts show the resource allocation distribution for both strategies (cell dry weight in the *y*-axis) for different growth rates (*x*-axis) caused when the availability of substrate is increased. Both strategies start from the same mass distribution at 0.5 doublings·h^{-1} (they share the same pie chart). The substrate was increased in the same quantity for both strategies. The arrows point to the resulting mass distribution pie chart for each strategy.

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.[4,8] As these models do not explicitly consider the host behavior, they cannot be easily used within a multiscale framework integrating synthetic circuits of interest, host, and 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[9−12] to semimechanistic ones with varied degrees of granularity.[13−15] 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.

The construction of a large-scale mechanistic model of *Escherichia coli* has enabled us to integrate and cross-evaluate a massive, heterogeneous dataset integrating measurements reported by various groups over decades.[15] On the other hand, medium-size detailed mechanistic models like the one developed in Weiße et al.[13] have been used to study behavioral modulations of a gene switch[16] or a feed-forward circuit.[17,18] These medium- and large-scale models, though very useful, are most often overparametrized and cannot easily be integrated within a user-friendly and lightweight computational framework for model-based circuit design.

The small-size model presented here has enough granularity to provide good predictions of the host dynamics, the expression of the genes of interest, and their interactions while having a small number of parameters to be estimated. We derived the dynamics of gene expression as a function of the fraction of free ribosomes relative to available mature ones considering protein synthesis on polyribosomes. We also defined the gene resources recruitment strength (RRS) as the key functional coefficient that allowed us to explain the distribution of resources among the host and the genes of interest and the relationship between the use of resources and cell growth. An additional goal was to provide a model useful for model-based circuit design purposes. To this end, the model considers explicitly lab-accessible gene expression characteristics such as promoter and ribosome binding site (RBS) strengths. We derived the equivalence between the relative resources recruitment strength and the relative mass fraction of a given protein at steady state. From this equivalence, the protein synthesis rate can be easily evaluated using the average host dynamics at steady state. We used experimental data from the literature to estimate the average resources recruitment strength for both ribosomal and nonribosomal proteins in *E. coli*. This allowed us to evaluate how the sensitivity of the resources recruitment strength to RBS and promoter can explain the variation of the cell protein mass fractions with growth rate and the differential roles they play. These data also can show how host−circuit interaction shapes the optimal abundance rates of both endogenous and exogenous proteins in the expression space.

2. RESULTS

2.1. Burden-Aware Model of Gene Expression Dynamics. Our model considers, on the one hand, the

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

dynamics of the expression of the cell host endogenous protein-coding genes. These are the genes that contribute to cell growth (Figure 1A). On the other hand, the model allows the possibility of considering the expression of protein-coding exogenous genes (Figure 1C). These do contribute to cell mass, but not to cell growth, akin to the consideration of unproductive proteins used in ref 9.

We started by modeling the polysomic gene expression dynamics for a generic kth protein-coding gene in prokaryote cells. We considered that transcription is faster than translation so it can be assumed at steady state, and that ribosomes are the limiting shared resource required for protein expression (see Section SI.1, Supporting Information (SI)). Under these assumptions, we derived the dynamics for the number of molecules of a kth protein as

$$\dot{p}_k = \frac{\nu_t(s_i)}{l_{pk}} J_k(\mu, r) r - (d_k + \mu) p_k \tag{1}$$

where p_k is the number of copies of the kth protein, l_{pk} is the protein length expressed as equivalent number of amino acids, d_k is the protein degradation rate constant, μ is the cell specific growth rate, r is the number of free ribosomes, and $\nu_t(s_i)$ is the substrate-dependent effective peptide chain elongation rate. This one is expressed using the Michaelis−Menten expression

$$\nu_t(s_i) = \nu \frac{s_i}{K_{sc} + s_i} \tag{2}$$

where ν is the maximum attainable peptide synthesis rate and K_{sc} is a Michaelis−Menten parameter related to the cell substrate uptake and catabolic capacity. As a first approximation, we considered that ν is organism-dependent but does not depend on the sequence of nucleotides (i.e., on the particular gene being expressed) and K_{sc} is organism- and substrate-dependent but does not depend on the nucleotides sequence either.

The term $J_k(\mu, r)$, defined as the resources recruitment strength (RRS), is a dimensionless function of the growth rate μ and the number of free mature ribosomes r that quantifies the capacity of a kth gene to engage cellular resources to get expressed (Figure 1A). It is the key functional coefficient in our model that explains the distribution of resources among the host and the genes of interest and the relationship between the usage of resources and cell growth (see its derivation in Section SI.1, SI). Besides depending on the cell growth rate and the availability of cell resources, the RRS is a function of the promoter and RBS strengths. For a generic protein-coding gene, its RRS is defined as

$$J_k(\mu, r) \triangleq E_{mk}(l_{pk}, l_e)\omega_k(T_f)\frac{1}{\frac{d_{mk}}{K^k_{C^0}} + \mu r} \tag{3}$$

On the one hand, the resources recruitment strength $J_k(\mu, r)$ depends on the availability of cell resources: the flux of free ribosomes μr and the ribosomes density-related term $E_{mk}(l_{pk}, l_e)$. This one is obtained (see Section SI.1, SI) as

$$E_{mk}(l_{pk}, l_e) \triangleq \frac{l_{pk}}{l_e}\left[1 - \left(\frac{l_{pk}}{l_{pk} + l_e}\right)^{l_{pk}/l_e}\right] \tag{4}$$

where $1/l_e$ is the specific ribosomes density, with l_e expressed as equivalent number of codons. The ribosomes density can be estimated as an inversely log-linear function of the protein length l_{pk} (see eq S92 in Section SI.13, SI). Interestingly, E_{mk} can accurately be approximated as $E_{mk}(l_{pk}, l_e) \approx 0.62 l_{pk}/l_e$ (i.e., a linear function of the number of ribosomes elongating along the transcript) for a wide range of values of l_{pk} and l_e (see Figure SI.2).

On the other hand, the RRS $J_k(\mu, r)$ also depends on gene expression characteristics: mRNA transcription rate $\omega_k(T_f)$ or the promoter strength, mRNA degradation rate constant d_{mk} and the effective ribosome binding site (RBS) strength $K^k_{C^0}(s_i)$. This one is defined as

$$K^k_{C^0}(s_i) \triangleq \frac{K^k_b}{K^k_u + K_e(s_i)} \tag{5}$$

where K^k_b and K^k_u are, respectively, the association and dissociation rate constants between a free ribosome and the RBS, and $K_e(s_i) = \nu_t(s_i)/l_e$ is the translation initiation rate constant, which depends on the availability of intracellular substrate (see Section SI.1 for details).

The resources recruitment strength of a given protein-coding gene can be related with its translation rate and number of transcripts. Consider the standard dynamic model for the expression of a protein p[19]

$$\dot{p} = \frac{\beta_p \beta_m}{d_m} - \mu p \tag{6}$$

where β_m (mRNA/t) is the transcription rate, β_p (protein/(mRNA·t)) is the translation one, and d_m is the mRNA degradation rate constant. Comparing with eq 1, we derived the relationship

$$J_k = \frac{l_{pk}}{\nu_t(s_i)} \frac{\beta_p \beta_m}{d_m} \frac{1}{r} \tag{7}$$

The expression eq 7 allows us to calculate the theoretical maximum RRS, $J_k|_{r=1}$, from the available experimental data (see Section SI.12, SI).

The dynamics of the total number of ribosomes can be obtained by considering an analogous expression to eq 1 for each of the N_r proteins forming up a ribosome (see Section SI.2, SI). The total 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) ones. In turn, the mature ribosomes r_a available for translation comprise the free ribosomes r and the ones already bound either to the RBSs or elongating along the transcripts. The number of available mature ribosomes is a fraction of the total number of ribosomes so that $r_a = \Phi_m r_T$. We assumed the fraction Φ_m varies little in time (see Section SI.2, SI) 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. Without loss of generality, we considered average protein-coding endogenous genes with RRSs J_r and J_{nr}, respectively. This allowed us to obtain the relationship between free and total number of ribosomes (see Section SI.3, SI) as

$$r = \frac{r_a}{1 + \mathrm{WSum}(\mu, r)} \tag{8}$$

with

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

$$\mathrm{WSum}(\mu, r) = N_r\left[1 + \frac{1}{E_{mr}}\right]J_r(\mu, r) + N_{nr}\left[1 + \frac{1}{E_{mnr}}\right] J_{nr}(\mu, r) + \sum_{k=1}^{N_{exo}}\left[1 + \frac{1}{E_{mk}(l_{pk}, l_e)}\right]J_k(\mu, r)$$

where N_r and N_{nr} are the number of ribosomal and nonribosomal protein-coding endogenous genes, respectively, and N_{exo} allows for the existence of exogenous genes.

Cell growth can essentially be explained as the time variation of the protein fraction of the total cell mass (Figure 1B). Yet, not all protein mass contributes to cell growth. There are proteins which may be undergoing active degradation while other proteins, e.g., the exogenous ones will not have any active role positively contributing to the cell growth. Therefore, we considered only the endogenous ribosomal and nonribosomal proteins to compute the cell specific growth rate. We used the dynamics eq 1 and assumed an average amino acid mass m_{aa} to obtain the time variation of the total endogenous protein mass content m_h of the native host cell (see Section SI.5, SI)

$$\dot{m}_h = m_{aa}\nu_t(s_i)\Phi_t^h\Phi_m r_T - \mu m_h \tag{9}$$

where

$$\Phi_t^h = \frac{N_r J_r(\mu, r) + N_{nr}J_{nr}(\mu, r)}{1 + \mathrm{WSum}(\mu, r)} \tag{10}$$

is the fraction of ribosomes elongating along endogenous ribosomal and nonribosomal transcripts relative to the total number of mature available ribosomes (see Section SI.4, SI).

Next, we considered that the total biomass dry weight variation (i.e., 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, the protein mass for each cell quickly reaches steady state ($\dot{m}_h \approx 0$). Thus, from eq 9, we obtained the expression for the cell specific growth rate

$$\mu(s_i) = \frac{m_{aa}}{m_h}\nu_t(s_i)\Phi_t^h\Phi_m r_T \tag{11}$$

where note that $\Phi_t^h\Phi_m r_T$ is the number of ribosomes actively elongating along endogenous transcripts at a given time instant (see Sections SI.4 and SI.5, SI).

To relate the growth rate $\mu(s_i)$ obtained from the intracellular dynamics with the extracellular measure of growth rate, $\mu(s)$, derived from cell population dynamics, we followed a reasoning derived from the model developed in ref 13, 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 (see Section SI.6 for details). Under the condition of steady-state growth where the rate of total cell mass growth is identical to the rate of cell number growth[20] and assuming that the maximum import and catabolism fluxes are balanced, we obtained the Monod population growth kinetics

$$\mu(s) = \frac{\frac{m_{aa}}{m_h}\nu\Phi_t^h\Phi_m r_T s}{\frac{k_t}{V_m} + s} \tag{12}$$

where V_m is a parameter related to the effective volume of culture broth for each cell and k_t is the Michaelis−Menten constant for substrate transport into the cell. Note that we recuperate the maximum specific growth rate μ_m as a linear function of the number of ribosomes actively elongating along transcripts at a given time instant. Finally, the Monod constant K_s as a function of the substrate transport capacity and the cell harvesting volume.

Our model accounts for the protein mass distribution (Figure 1B) but does not consider the relationship between growth rate and the total cell protein mass. Cells growing at faster growth rates are larger and heavier, thus affecting their total protein content.[21] To model the relationship between the cell protein content and the specific growth rate for the native host cell—i.e., $m_h = m_h(\mu)$—we postulated the relationship

$$\frac{dm_h}{d\mu} = \beta m_h \tag{13}$$

with $m_h(0) = 77.375$ fg and $\beta = 61.781$ min as best fits obtained for *E. coli* cells using the data in ref 22. We also considered an analogous model relating the cell dry weight $m_{h,cDW}(\mu)$ with the growth rate (see Section SI.7 for details).

Finally, we structured our model in such a way that it can be used to analyze the resource allocation trade-offs (see Figure 1D) among the endogenous protein-coding genes from the native *E. coli* host cell, and a given set of exogenous ones of interest (e.g., a synthetic gene circuit). In the latter case, we have considered, without losing generality, a single exogenous protein of interest A to exemplify the model expressions and the interaction between the host cell and the exogenous additions.

For the endogenous protein-coding genes, we considered the ensemble of ribosomal and nonribosomal genes as lumped species with average values of $E_{mr}(l_p^r, l_e)$, $E_{mnr}(l_p^{nr}, l_e)$ and $J_r(\mu, r)$, $J_{nr}(\mu, r)$, respectively. Then, we obtained the dynamics of the total mass of ribosomes m_{r_T} and nonribosomal endogenous proteins m_{nr}, and the dynamics of the mass m_A of the exogenous protein (see Sections SI.8 and SI.9 for details) as

$$\dot{m}_{r_T} = \mu\left[m_h(\mu)\frac{N_r J_r(\mu, r)}{N_r J_r(\mu, r) + N_{nr}J_{nr}(\mu, r)} - m_{r_T}\right];$$
$$J_r(\mu, r) = E_{mr}\omega_r\frac{1}{\frac{d_m^r}{K_{C0}^r(s_i)} + \mu r} \tag{14}$$

$$\dot{m}_{nr} = \mu\left[m_h(\mu)\frac{N_{nr} J_{nr}(\mu, r)}{N_r J_r(\mu, r) + N_{nr}J_{nr}(\mu, r)} - m_{nr}\right];$$
$$J_{nr}(\mu, r) = E_{mnr}\omega_{nr}\frac{1}{\frac{d_m^{nr}}{K_{C0}^{nr}(s_i)} + \mu r} \tag{15}$$

$$\dot{m}_A = \mu\left[m_h(\mu)\frac{N_A J_A(\mu, r)}{N_r J_r(\mu, r) + N_{nr}J_{nr}(\mu, r)} - m_A\right];$$
$$J_A(\mu, \mathrm{r}) = E_{mA}(l_{pA})\omega_A\frac{1}{\frac{d_m^A}{K_{C0}^A} + \mu r} \tag{16}$$

where N_A is the gene copy number of A, the number of free ribosomes r is obtained using eq 8, and the specific growth rate μ is calculated using eq 11.

The denominator in the fraction of RRSs only includes the host protein-coding genes. The protein mass $m_h(\mu)$ is that of the native host cell, comprising only the cell endogenous

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

proteins. We defined the mass of the strain $m_s = m_h + m_A$ as the one comprising the mass of the host and that of the exogenous proteins. We obtained the relation between the protein mass of the strain $m_s(\mu)$ and that of the native host $m_h(\mu)$ (see Section SI.9, SI) as

$$m_s(\mu) = \frac{\Phi_t^s}{\Phi_t^h} m_h(\mu) = \frac{N_r J_r + N_{nr} J_{nr} + N_A J_A}{N_r J_r + N_{nr} J_{nr}} m_h(\mu) \tag{17}$$

In addition, we also considered the cell dry weight $m_{cDW}(\mu)$, comprising the mass $m_h(\mu)$ of the endogenous ribosomal and nonribosomal proteins, the mass of the exogenous proteins $m_A(\mu)$ and the mass of other constituents of the cell, denoted as $m_Q(\mu)$. Thus, $m_{cDW}(\mu) = m_h(\mu) + m_A(\mu) + m_Q(\mu) = m_s(\mu) + m_Q(\mu)$. To obtain $m_Q(\mu)$, we used the estimation of the cell dry weight $m_{h,cDW}(\mu)$ for the *E. coli* host native cell, i.e., without expression of exogenous genes (see Section SI.7 for details), assuming that $m_Q(\mu)$ does not depend on the expression of exogenous genes. This allowed us to estimate the mass fractions with respect to the total cell dry weight.

To evaluate the productivity rate of a given protein of interest, we obtained its mass synthesis rate as the steady-state mass of protein produced per cell and generation (see Section SI.9, SI). In the case of an exogenous protein *A* and using eqs 16 and 17, we obtained

$$\begin{aligned} \Pi_A &\triangleq m_{A,ss}\mu \\ &= m_s(\mu) \frac{N_A J_A(\mu, r)}{N_r J_r(\mu, r) + N_{nr} J_{nr}(\mu, r) + N_A J_A(\mu, r)} \mu \end{aligned} \tag{18}$$

We defined the specific mass synthesis rate (spMSR) relative to the cell dry weight as

$$\begin{aligned} \pi_A &\triangleq \frac{\Pi_A}{m_{cDW}(\mu)} \\ &= \frac{m_s(\mu)}{m_{cDW}(\mu)} \frac{N_A J_A(\mu, r)}{N_r J_r(\mu, r) + N_{nr} J_{nr}(\mu, r) + N_A J_A(\mu, r)} \mu \end{aligned} \tag{19}$$

For a given protein *A*, both the protein mass synthesis rate ($g \cdot min^{-1}$) and the specific one ($g \cdot min^{-1} \cdot gCDW^{-1}$) are directly related to its relative RRS fraction.

From the results above, we obtained the cell specific growth rate at steady-state exponential balanced growth as

$$\mu_{ss}(s_i) = \frac{m_{aa}}{m_{rib}} \nu_t(s_i) \Phi_m \Phi_t^r \tag{20}$$

where m_{rib} is the average ribosome mass (see Section SI.9 for details). That is, the cell growth rate at steady state depends linearly on the fraction $\Phi_t \Phi_r^t$ of bound ribosomes being actively used to build up ribosomes themselves (i.e., ribosomes actively elongating along and translating ribosomal transcripts) relative to the total number of ribosomes.

2.2. Ribosomal and Nonribosomal Genes Differ in Their Average Resources Recruitment Strength. Using the experimental data in ref 19, we evaluated the maximum expected magnitude of the resources recruitment strength for each gene using eq 7 with $r = 1$, i.e., the theoretical maximum RRS for a given availability of intracellular substrate, $J_{k,max} = J_k|_{r=1,\nu_t(s_i)}$. The data in ref 19 correspond to *E. coli* cells under fast-growing conditions (doubling time $t_d = 21.5$ min). Therefore, we could assume saturation of substrate, allowing us to consider the maximum substrate-dependent effective translation elongation rate $\nu_t(s_i) = \nu$ to evaluate eq 7. Note that this is equivalent to estimating the maximum RRS for the maximum specific growth rate. *E. coli* has around 4225 protein-coding genes.[23,24] From ref 19, we obtained data for a representative enough set of genes, comprising 3551 nonribosomal and 68 ribosomal ones, accounting for around 86% of all *E. coli* genes.

First, the results allowed us to estimate the order of magnitude of the resources recruitment strength for ribosomal and nonribosomal genes in *E. coli* and their maximum average value. Then, we obtain how many genes of each class are active at any one time.

As expected, the values obtained spanned several orders of magnitude. For the ribosomal genes, the average value $J_{max,r}^{avg} = 124.5$ and a coefficient of variation $CV_{J_{max,r}} \approx 1$, while for the nonribosomal ones, the values were $J_{max,nr}^{avg} = 3.78$ and $CV_{J_{max,nr}} \approx 6$. The average maximum RRS for the ribosomal genes was 2 orders of magnitude higher than for nonribosomal ones. Yet, the coefficient of variation was much smaller for the ribosomal resources recruitment strengths than for the nonribosomal ones. Figure 2 shows the values of $J_{k,max}$ we obtained for each gene sorted by the log-magnitude of the ratio between the maximum RRS and the length of the associated protein. The results did not essentially change from the non-normalized ones (see Figure SI.6, SI). That is, the resources recruitment strength of *E. coli* genes is not fundamentally determined by the lengths of the proteins they code. This suggests that factors such as the effective transcription and translation rates are more relevant.

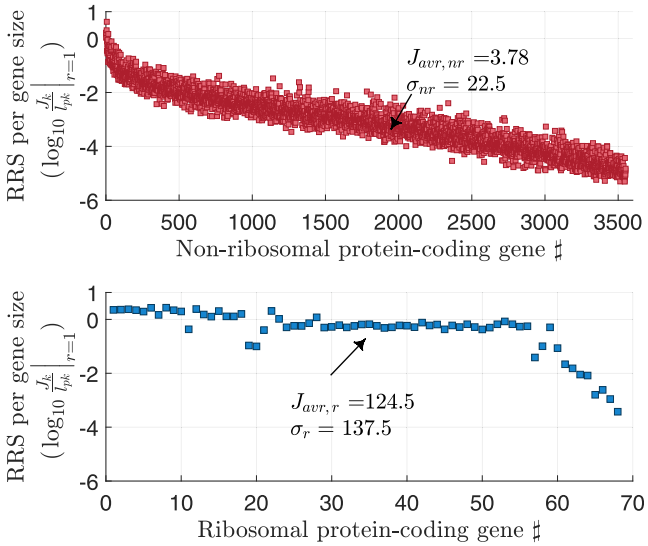


Figure 2. Log-magnitude of the ratio between the maximum resources recruitment strength and the length (aa) of the associated protein for the set of nonribosomal (top) and ribosomal (bottom) protein-coding genes in ref 19. The genes were ordered by decreasing value of the ratio.

But not all genes are expressed all of the time. As a proxy to estimate how many genes are active at any given time, we calculated the cumulative sum of the maximum RRS and obtained how many genes being expressed are required to explain both 95 and 99% of the total cumulative sum (see Figure SI.8). We did this independently for both ribosomal and

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

Table 1. Average Best-Fit Estimated Values for *E. coli* of the RBS-Strength-Related Parameters K_b^k, K_u^k and Transcription Rates ω_k for Ribosomal (k = r) and Nonribosomal (k = nr) Proteins and the Fraction Φ_t of Mature Ribosomes with respect to the Total Number of Ribosomes

parameter (units)	K_u^r (min^{-1})	K_u^{nr} (min^{-1})	K_b^r (molecule^{-1}·min^{-1})	K_b^{nr} (molecule^{-1}·min^{-1})	ω_r (mRNA·min^{-1})	ω_{nr} (mRNA·min^{-1})	Φ_m (adim.)
mean	129.9	3.09	5.57	12.86	5.65	0.028	0.90
std.	4.07	0.14	0.78	1.50	0.29	0.25×10^{-3}	0.5×10^{-2}

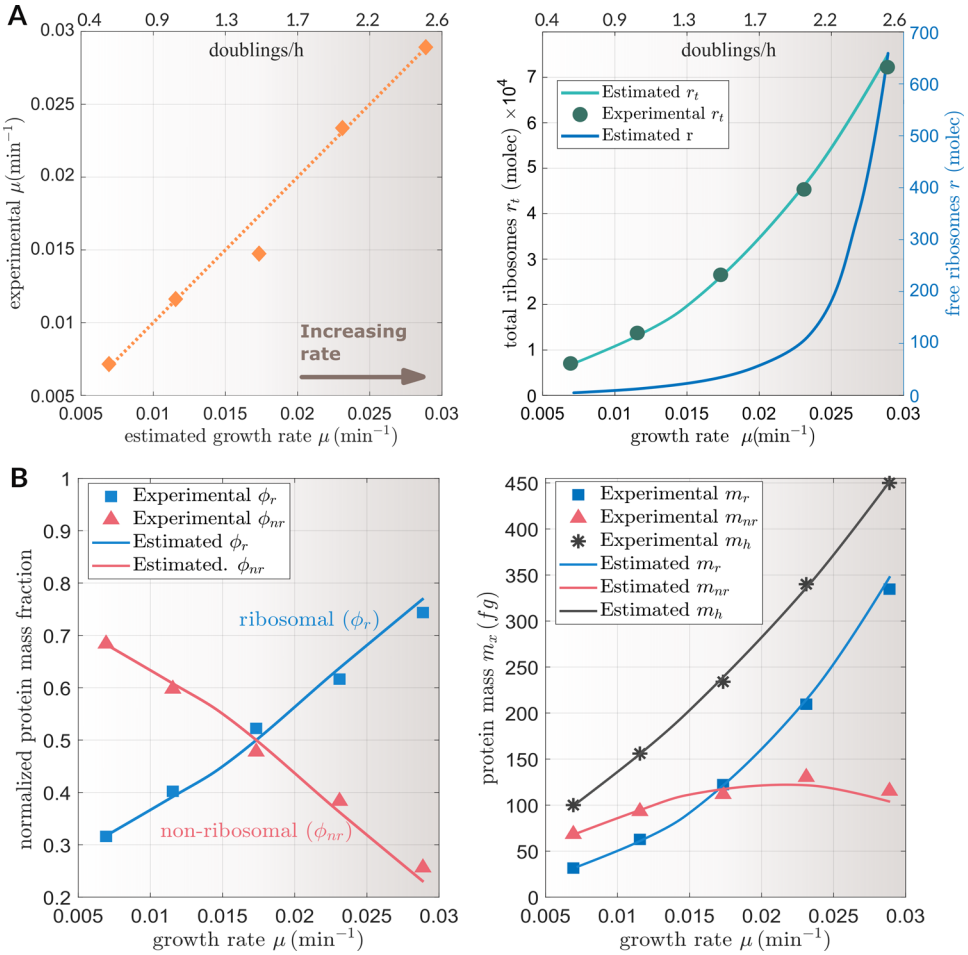


Figure 3. (A) Estimated versus experimental growth rate (left). Experimental and estimated number of total ribosomes as a function of the growth rate and estimated number of free ribosomes (right). (B) Estimated versus experimental mass fractions of ribosomal ($\phi_r = m_r/m_h$) and nonribosomal ($\phi_{nr} = m_{nr}/m_h$) proteins in *E. coli* (left). Ribosomal (m_r), nonribosomal (m_{nr}), and total host cell protein mass (right). In all plots, the x-axis corresponds to the estimated and experimental growth rates evaluated for the range of peptide chain elongation rates $\nu_t(s_i)$ extracted from ref 22.

nonribosomal proteins. Our results showed that out of the 68 ribosomal genes, 49 of them (72%) explained 95% of the cumulative sum of the maximum resources recruitment strength. To explain 99%, we needed 57 ribosomal genes (84% of them). However, for nonribosomal genes, we needed 875 out of 3551 genes (25%) to explain 95% of the cumulative sum and 1735 (49%) to explain the 99%.

2.3. Resources Recruitment Strength Explains the Distribution of Ribosomal and Nonribosomal Protein Mass Fractions. The relative mass fractions of ribosomal and nonribosomal proteins in the cell depend on the cell growth rate so that the ribosome content increases linearly with growth rate.[9,22,25,26] Existing resource allocation models explain this as a result of optimal allocation of cell resources between the ribosomal and nonribosomal fractions, balancing the demands of protein synthesis and nutrient influx under the constraint that the sum of both fractions remains constant.[9] In our model, the relative resources recruitment strength of a given protein equals its relative mass fraction in the cell at steady-state balanced growth (see eqs 14−16 and Section S.8, SI). Therefore, the relative distribution of mass between ribosomal and nonribosomal proteins must be reflected in the relative distribution of their resources recruitment strengths.

We first studied the *E. coli* host cell, i.e., without any exogenous protein-coding genes. We used the data in ref 22 to check whether our model was able to predict the linear increase of ribosomes content with growth rate and the relative distribution of endogenous ribosomal and nonribosomal protein mass fractions as a function of growth rate. We did not estimate the model parameters to try and directly fit the experimental relative distribution of resources recruitment strengths, as this would not inform on the capability of the

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

model to capture the intrinsic relationship among growth rate, use of cell resources, and distribution of protein mass fractions. Instead, we analyzed if a good fit of the specific growth rate implied our model could generalize and predict the relative mass fractions in the cell. This, in turn, implies fitting the ribosomal and nonribosomal resources recruitment strengths.

To this end, we fitted the model parameters using the experimental growth rate as output to predict. We used the values of the peptide chain elongation rates $\nu_t(s_i)$ as a function of growth rate available from ref 22 as the only input information given to the model. This is tantamount to feed the model only with the available amount of substrate s_i (see Section 5.2). Then, we minimized the sum over the experimental data points of the absolute growth rate prediction error (see Section 5.2). We considered the lumped resources recruitment strengths for both the ribosomal and nonribosomal endogenous proteins (see expressions 14 and 15) and estimated the fraction of mature ribosomes and the parameters corresponding to the RBS strength and transcription rates. This would provide our model a good fit of the specific growth rate. The best-fit estimated parameters are given in Table 1.

The estimated values of the RBS-strength-related parameters K_b^k, K_u^k implied ribosomal RBSs much weaker than the nonribosomal ones. Interestingly, the values we obtained for the transcription rates were in the same order of magnitude as the mean values obtained from the data in ref 19—ω_r = 2.4 and ω_{nr} = 0.05, respectively. Therefore, this demonstrates a much higher value for the average transcription rate of ribosomal proteins than for the nonribosomal ones. Our results also estimated an average high transcription−low translation rate expression strategy for the ribosomal endogenous genes and the opposite strategy for the nonribosomal ones.

Figure 3A shows the results of the model parameter fitting and the good agreement between the experimental and the estimated growth rate. The estimation of the number of free ribosomes for cells growing at doubling time t_d = 25 min ($\mu \approx 0.028$ min^{-1}) was consistent with the result $r \approx 350$ obtained using the experimental data in ref 19 (see Section SI.14, SI). For cells growing faster, the number of free ribosomes much increased. Note, though, that also the total number of ribosomes (both experimental and estimated) greatly increased for very fast-growing cells. Thus, the fraction of free ribosomes with respect to the total number only increased from 0.08 up to 1.37% for cell doubling times between 100 and 24 min, respectively (even though the number of free ribosomes varied by almost 200-fold). Similarly, the computed fraction Φ_m of mature ribosomes with respect to the total number of ribosomes was consistent with the estimated fraction of active bound ribosomes $\Phi_t^h\Phi_m \approx 0.78$ (see Figure SI.11) in agreement with refs 12 and 22.

We evaluated the mass fractions of the endogenous ribosomal and nonribosomal proteins at steady state using the expressions 14 and 15. The model predictions were in very good agreement with the experimental values, as shown in Figure 3B. Therefore, our model reproduced the known linear increase of the ribosomal fraction with growth rate. The differential behavior between the ribosomal and nonribosomal resources recruitment strengths was behind the differential protein mass distribution as the cell growth rate increases.

The effective RBS strength used in our model is a function of the intracellular substrate because it varies with the cell growth rate according to eq 5. Figure 4 shows the estimated

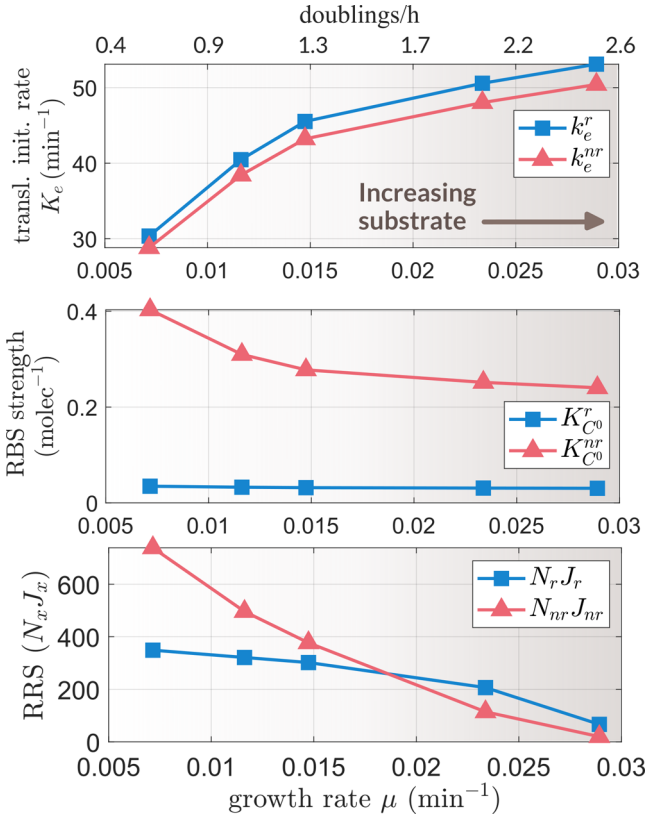


Figure 4. Estimated translation initiation rate k_e for the average ribosomal and nonribosomal endogenous genes as a function of the specific growth rate μ (top). Estimated effective RBS strengths $K_{C^0}^r$ and $K_{C^0}^{nr}$ (middle). Estimated total resources recruitment strengths N_rJ_r and $N_{nr}J_{nr}$ as a function of growth rate μ (bottom).

values as a function of the specific growth rate μ. The estimated effective RBS strength of the nonribosomal protein-coding genes ($K_{C^0}^{nr}$) was much higher than that of the ribosomal ones ($K_{C^0}^{r}$). As the growth rate increased—tantamount in our model to an increasing intracellular substrate s_i—the ribosomal effective RBS strength kept almost constant (with a slight decrease around 12%) while the nonribosomal one decreased by almost a 40%. We could explain this trend as a result of the difference in the ratio between the transcript degradation rate and the RBS strength, $d_{mk}/K_{C^0}^k$ for both ribosomal and nonribosomal genes. The ribosomal genes kept much higher values of $d_{mk}/K_{C^0}^k$ for all values of the flux of free resources μr. This, taking into account the monotonous increasing power-law relationship between the growth rate and the number of free ribosomes predicted by our model (see Section SI.14, SI) implies the observed trends in the values of the RRS in Figure 4 (bottom). The ribosomal RRS $J_r(\mu,r)$ decreases much slower than that of the nonribosomal ones as the growth rate increases.

Section 2.1 shows that in endogenous genes, steady state is reached for balanced exponential growth when their relative fraction of resources recruitment strength equals their mass relative to that of the host cell. Since the ribosomal resources recruitment strength decreases much slower than the nonribosomal one as the growth rate increases, the fraction of ribosomal RRS with respect to the total sum of ribosomal and nonribosomal RRSs will increase. As a consequence, its relative mass fraction will increase.

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

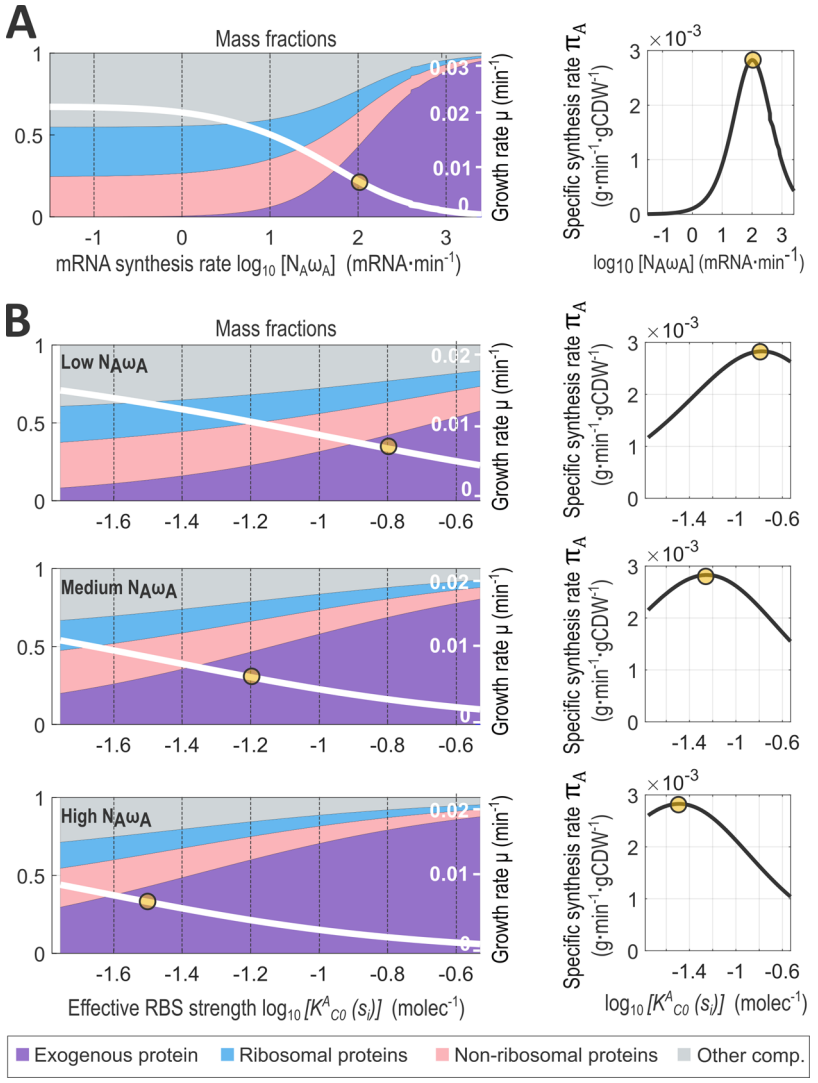


Figure 5. (A) Left: effect of increasing the mRNA synthesis rate of an exogenous protein A on the cell growth rate (right axis) and on the cell mass fractions (left axis). Right: specific protein mass synthesis rate (spMSR) for the exogenous protein A as a function of its mRNA synthesis rate. Even though the growth rate decreases for increasing mRNA synthesis rates, the spMSR increases, reaching a maximum value (yellow dot) at fast mRNA synthesis rates around 2000 mRNA·min^{-1} and eventually decreases for larger mRNA synthesis values. (B) Left: Effect of varying the RBS strength on the cell growth rate (right axis) and the protein mass fractions (left axis) for three increasing values of the mRNA synthesis rate (low, medium, high). Right: spMSR of the exogenous protein as a function of RBS strength variation.

It is important to stress again that we estimated the parameters in our model so as to fit not the experimental mass fractions but the cell growth rate. By doing that, the internal structure of the model—substantiated in the structure of the resources recruitment strength functional coefficients—captured the correct differential mass distribution between ribosomal and nonribosomal cell protein content as a function of growth rate.

2.4. Host−Circuit Interaction Shapes the Optimal Synthesis Rate of Exogenous Proteins. There are multiple ways to increase the expression of an exogenous protein of interest, including the choice of the expression vector of the synthetic gene circuit, optimizing the use of codons, co-expression of chaperones to aid protein folding, etc.[27] We focused on varying the expression space—i.e., the gene induction space defined by the values of the mature mRNA synthesis rate and the effective RBS strength—at the same values of cell growth conditions and intracellular substrate availability. We used the average host dynamics at steady-state balanced growth to evaluate the distribution of cell mass fractions and the specific protein mass synthesis rate (specific synthesis rate for short or spMSR) of a given exogenous protein of interest *A* as defined in eq 19 (see also Section SI.9) as a function of variations in the expression space.

We first considered the RBS-strength-related parameters K_b^A, K_u^A of the exogenous gene to be constant with values equal to the estimated averages for an endogenous nonribosomal protein in *E. coli* (see Table 1) and only the mRNA synthesis rate was varied. To this end, we changed the gene copy number times the transcription rate (or promoter strength) $N_A\omega_A$ in the range $[10^{-1}, 10^5]$ times the average promoter strength of endogenous nonribosomal genes given in Table 1. This gave a maximum value $N_A\omega_A \approx 3.3 \times 10^3$ mRNA·min^{-1}, which is an attainable value for *E. coli* considering an average transcription rate $\omega_A = 3$ mRNA·min^{-1} and a high copy number plasmid with $N_A = 1100$.

Figure 5A shows the variation across the mRNA synthesis space $N_A\omega_A$ of the mass fractions and the cell growth rate (left)

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

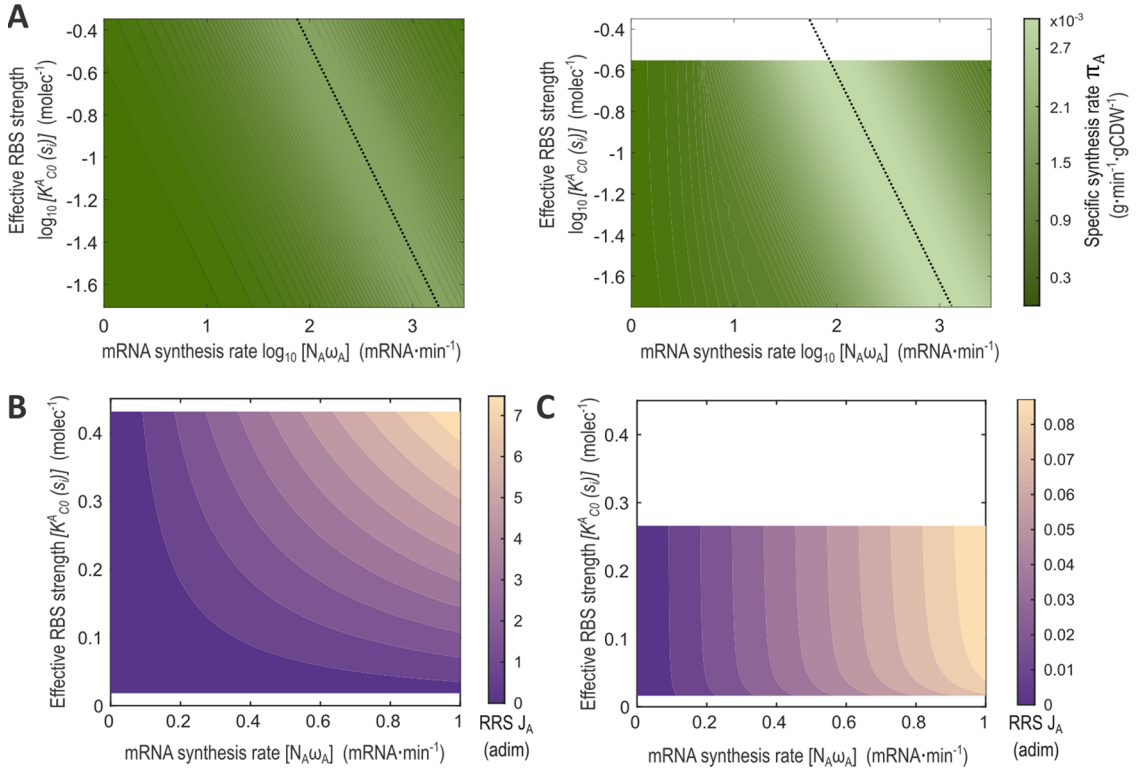


Figure 6. Effect of varying the mRNA synthesis rate and the effective RBS strength on (A) the specific synthesis rate of protein A and (B, C) the resources recruitment strength (J_A) for different substrates. (B) Low substrate scenario $\nu_t(s_i)$ = 720 min^{-1}. (C) High substrate scenario $\nu_t(s_i)$ = 1260 min^{-1}. The value of J_A was evaluated for the full range of RBS values ($K^A_{C^0}(s_i)$) and a representative range of promoter values ($N_A\omega_A$), with E_{mA} and d_{mA} equal to endogenous ribosomal values (without loss of generality).

and the spMSR, π_A, of the exogenous protein (right). The distribution of mass fractions was consistent with the behavior of the cell. As the mRNA synthesis rate of the gene A was increased (moving toward the right in the plot Figure 5A, left), the mass fraction corresponding to the protein A also increased (purple) while that of ribosomal proteins decreased (blue) with a corresponding decrease in the cell growth rate (white line). Consequently, there appeared a maximum specific protein mass synthesis rate value (Figure 5A yellow dot, $\pi_A \approx 2.8 \times 10^3$ g·min^{-1}·gCDW^{-1}) which was achieved for an mRNA synthesis rate of 100 mRNA·min^{-1}. This value represents a low copy plasmid number $N_A \approx 20$ and a constitutive promoter with a transcription rate $\omega_A \approx 5$ mRNA·min^{-1}.

The model predicted an increasing mass fraction of the protein A as we continue increasing the value of the mRNA synthesis rate. However, this situation happens at the cost of reducing the fraction of ribosomal proteins, resulting in a very small growth rate. The relationship between the fraction of exogenous protein and growth rate in our model is a decreasing exponential (something consistent given its mathematical smooth differential continuous-time nature). Therefore, even if the zero growth rate is achieved in the limit for 100% of exogenous protein, note that this is a theoretical point only achieved in the limit, i.e., at infinite cell doubling time. In practice, the cell viability will be lost before.

Figure 5B (left) shows the results obtained when we analyzed three representative values of the mRNA synthesis rate $N_A\omega_A$ = {150, 400, 800} corresponding to an average transcription rate in *E. coli* ($\omega_A \approx 3$ mRNA·min^{-1}) combined with a low, medium, and high plasmid copy number, respectively. Then, we varied K^A_b, K^A_u in the ranges considered in Section 5.2 to obtain a range of values for the effective RBS strength $K^A_{C^0}(s_i)$. The mass fraction corresponding to protein A increased and the cell growth rate decreased for high levels of the RBS strength. Figure 5B (right) shows that the main factor affecting the spMSR is the mRNA synthesis rate $N_A\omega_A$. Thus, for low values of the mRNA synthesis rate, the spMSR increased for strong RBSs until a maximum appeared for one of the stronger ones (e.g., $K^A_{C^0} = 10^{-0.8} = 0.15$ molecule^{-1}). For medium values of $N_A\omega_A$, there soon appeared a maximum spMSR for the exogenous protein as a function of the RBS strength. Finally, for high values of the mRNA synthesis rate, increasing the RBS strength rapidly produced a decrease in the specific protein mass synthesis rate. Our model correctly predicted that there is a critical (optimal) protein synthesis rate that is achieved for lower RBS strength as the mRNA synthesis rate increases.

The location of the optimal spMSR as a function of variations in the full range of the expression space can be seen in Figure S.12C in Section SI.15, which shows the variation of the specific synthesis rate of the exogenous protein across the expression space $N_A\omega_A$, $K^A_{C^0}(s_i)$ in log−log scale. The optimal subspace corresponded to a line in the log−log promoter−RBS space, showing the existence of a trade-off between the mRNA synthesis rate (tantamount to the gene induction) and the RBS strength. The pronounced slope of the optimal subspace explains the different sensitivity of the specific synthesis rate to the variations of either the promoter or the RBS strengths that were obtained in Figure 5A,B. Our model predicted that the specific synthesis rate is more sensitive to variations of the mRNA synthesis rate than to variations of the RBS strength.

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

Thus, for intermediate values of $N_A\omega_A$, there is a wide range of RBS strengths that keep the specific synthesis rate close to its optimal value. This is also reflected in the smoother transition between the mass fractions resulting when the RBS strength is modified compared to changing the mRNA synthesis rate. Note that, as predicted by eq 19, for a given substrate availability, different expression strategies resulting in the same specific synthesis rate will correspond to the same distribution of mass fractions.

Differently from modifying the mRNA synthesis rate for a fixed RBS strength value, or vice versa, the maximum spMSR of the exogenous protein significantly changed when the substrate availability does not remain constant. The effect of the differential role of RBS and promoter combinations for scenarios with varying substrate is analyzed in the next section.

2.5. Substrate Level Emphasizes 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[28] and the trade-off between desired expression levels and noise[19] and between expression of endogenous and synthetic genes and growth.[11,13] In the previous section, we showed that for a constant availability of substrate rich in nutrients, there are different promoter and RBS combinations that can achieve the same expression level (tantamount the same specific protein mass synthesis rate) of an exogenous protein A. This leads to a multimodal design problem. One can choose between design strategies ranging from using a combination of a weak promoter strength and a strong RBS ($N_A\omega_A \downarrow K^A_{C^0} \uparrow$) to using a strong promoter and a weak RBS ($N_A\omega_A \uparrow K^A_{C^0} \downarrow$). The results depicted in Figure 5 show that for the case of constant substrate, there is no difference between using one promoter−RBS combination or another as long as the desired spMSR of the protein A remains the same.

However, changes in the substrate have a different impact on the protein expression depending on which one is the promoter−RBS combination selected. Figure 1B,D illustrates how the mass fractions of ribosomal, nonribosomal, and exogenous proteins change as a function of the growth rate μ, which is indirectly dependent on the availability and quality of the substrate. For a given gene following the weak-RBS strong-promoter strategy (the one followed by the endogenous ribosomal genes), the mass fraction corresponding to the exogenous protein increased as the availability of substrate increased. On the contrary, the strong-RBS weak-promoter strategy (as followed by the endogenous nonribosomal genes) caused the exogenous protein mass fraction to decrease with increasing availability of substrate.

To understand the differential role of RBS and promoter strengths, we first evaluated the dependence of the specific protein mass synthesis rate of an exogenous protein A on the mRNA synthesis rate and the effective RBS strength as a function of the substrate. Figure 6A shows the results for two representative substrate levels: low substrate $\nu_t(s_i) = 720\ \text{min}^{-1}$ (left) and high substrate $\nu_t(s_i) = 1260\ \text{min}^{-1}$ (right). The maximum protein synthesis rates (black dashed lines) are located at different places in the design space. Increasing the substrate had the effect of increasing the spMSR (the right plot is whiter than the left one). In addition, the optimal synthesis rate moved to the right, i.e., for the same mRNA synthesis rate, a higher effective RBS was required to reach the optimum. This implies that a cell configured to obtain the optimum protein synthesis rate for some substrate level will become suboptimal when changing the substrate level.

The resources recruitment strength (RRS) explains this differential effect of RBS and promoter strength on protein expression. For a given protein, its RRS eq 3 is directly proportional to the mRNA synthesis rate. Figure 6B,C shows that the mRNA synthesis rate effectively modifies the RRS value regardless of the substrate or the growth rate. As the mRNA synthesis rate increases (displacement to the right in the x-axis) the value of the RRS increases. Therefore, tuning the promoter strength implies tuning the RRS level without affecting the RRS sensitivity to changes in the substrate, the growth rate or the changing availability of free ribosomes.

Different from the promoter strength, the RBS strength determines the sensitivity of the resources recruitment strength to changes in the substrate. It has two different effects on the value of the RRS. The first effect is related to the definition of the RBS in eq 5. It depends on the association−dissociation rate constants K^k_b and K^k_u and indirectly on the substrate through $K_e(s_i)$. For a given substrate s_i, there is a set of infinite combinations of K^k_b and K^k_u that might provide the same RBS strength level. This causes the strength of the RBS to vary with changes in the substrate so that it decreases as the substrate increases. However, the RBS strength (and therefore the RRS value) with a high dissociation constant rate $K^k_u \gg K_e(s_i)$ will be less sensitive to changes in the substrate.

On the other hand, note, from eq 5, that the RBS strength defines the sensitivity of the RRS to the flux of free resources μr. Decreasing the RBS strength will always reduce the RRS value. However, increasing the RBS strength will increase the RRS value until it eventually saturates. In particular, when $d_m/K^k_{C^0} \ll \mu r$, the RRS eq 5 becomes

$$J_k(\mu, r) = E_{mk}(l_{pk}, l_e)\frac{\omega_k(T_f)}{\mu r} \tag{21}$$

in this case, the RRS value becomes independent of the RBS strength. Thus, there is a maximum RRS value that can be obtained by increasing the RBS strength. Figure 6B shows that for low substrate availability, the RBS can increase and yet the RRS value decrease, and Figure 6C shows the saturating effect of increasing the RBS strength for a high substrate.

For exogenous protein-coding genes, the situation is different depending on whether they do add or not a relevant burden on the cell. In case the exogenous genes do not overload the cell, the expression patterns will be the same as those for the endogenous genes analyzed above. In case the exogenous genes impose an important burden on the cell, the effects of RBS and the promoter change. In this case, μr will be very small and the differential effect of the promoter and RBS strengths is partly lost. In this overloaded scenario, the RRS can be approximated as

$$J_k(\mu, r) = E_{mk}(l_{pk}, l_e)\frac{\omega_k(T_f)K^k_{C^0}(s_i)}{d_{mk}} \tag{22}$$

Therefore, the RRS only depends on the substrate through the substrate-dependent value of the RBS strength. This causes the RBS−promoter strength strategies to become less differentiated. Yet, it is still possible to apply the analysis above for the distribution of mass fractions as a function of the RRS. Thus, the strong-promoter weak-RBS strategy will allow us to have RRS whose value is less sensitive to changes in the

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

substrate compared to the weak-promoter strong-RBS one, as observed in Figure 1D.

3. DISCUSSION

Our model defined the gene resources recruitment strength as the key functional coefficient that explains the distribution of resources among the host−circuit and the relationship between the use of these resources and cell growth. The RRS generalizes similar proposals in the literature, allowing us to analyze not only scenarios with high cell burden but also scenarios where the competition for cell resources does not overload the cell extremely. Conversely from the resource demand coefficient defined in ref 4, where the resource limitation effect is local, we considered that the cell resources (ribosomes) are accessible to all genes in the cell, so exogenous and endogenous host genes compete to recruit cellular resources. The assumption of constant growth rate, constant total number of ribosomes, and highly overloaded cell in ref 4 implies a static resource demand coefficient that is independent of the availability of free resources or the growth rate. This assumption is equivalent to the overloaded scenario in our model with RRS given in expression 22. However, our RBS strength does depend on the substrate. Therefore, our model can be used in scenarios where the demand on resources changes over time since the RRS explicitly captures the mass distribution dependence on cellular growth and substrate availability.

The RRS of a gene plays an important role in the value of the specific protein mass synthesis rate. Note, from eq 19, that RRS and the spMSR are related. The specific mass synthesis rate is essentially a function of the ratio between the RRS of the gene of interest and the total sum of RRSs of the cell. Therefore, it provides information about the resources that the gene of interest is capturing and sharing with other cell components to get expressed. In this sense, spMSR is a context-dependent magnitude that requires knowledge of the spMSRs of the remaining genes. The resources recruitment strength is somewhat a more fundamental characterization of a protein-coding gene than the specific protein mass synthesis rate. It is kind of a context-dependent intrinsic magnitude. Its shape only depends on the gene characteristics. Its actual value is only defined by the generic flux of free resources μr and the effect of the substrate availability (which may integrate the nutrient quality) on the effective RBS strength. Therefore, the RRS measures the intrinsic avidity of a given protein-coding gene for cell resources.

Interestingly, the spMSR, i.e., the mass synthesis rate of a given protein per cell mass can be related to the definition of capacity as proposed in refs 29 and 30. There, a cell capacity monitor is implemented by including the constitutive expression of a GFP gene and determining capacity as the GFP production rate per cell of their capacity monitor. Both concepts, capacity and spMSR, are not the same but are related. In ref 30, the authors show the existence of a critical capacity. Our results, as seen in Figure 5A,B, also showed an upper bound or "critical" spMSR as a function of the mRNA synthesis rate ($mRNA{\cdot}min^{-1}$). The existence of this critical spMSR is not directly related to energy limitation, but it is the result of the peptide optimal allocation for building blocks (aas) to synthesize either a given (possibly exogenous) protein or more ribosomes. Indeed, energy limitations will indirectly affect the critical capacity value insofar as they interfere with the flow of building blocks to build up the peptide chains. From the perspective of energy as a resource, our model implicitly incorporates this concept as a fundamental part of the substrate. That is, all of the resources needed by the cell eventually come from the substrate. Consequently, our model captures this substrate−energy interaction and it is quantified by the resources recruitment strength and the substrate-dependent effective RBS strength. This approach differs from others such as ref 13, where energy is modeled explicitly after defining additional gene expression thresholds and a sigmoidal transcription/translation dependence on the energy levels.

The results obtained with our model were relevant both for the analysis of the native host cell, i.e., without exogenous protein-coding genes, and for the case of having a strain hosting exogenous protein-coding genes.

In the first case, we showed that endogenous ribosomal and nonribosomal genes clearly differ in their average resources recruitment strength and, therefore, in their average requirement for cell resources. The ribosomal proteins, essential for the cell and continuously being expressed, have higher RRS values than the nonribosomal ones. Moreover, its range of variation over the ribosomal proteins was much lower than for nonribosomal ones. This result was not fundamentally determined by the lengths of the coded proteins and is consistent with the fact that to great extent all ribosomal proteins are equally important for the cell. Transcription and translation are energetically expensive processes. It is usually accepted that around 60% of genes are expressed in standard laboratory conditions at any one time in *E. coli*, with only a small fraction making up a large percentage of the total protein. The cumulative sum of the maximum resources recruitment strength gave a good estimation of the percentage of genes expressed at any time. This is consistent with the fact that ribosomal genes are continuously needed for the cell so they are continuously expressed. On the contrary, nonribosomal genes are regulated to be expressed only when they are required. This also explains the very low RRS values obtained for them and reflects these genes are down-regulated most of the time.

It is known that weakly expressed endogenous genes exhibit low RNA polymerase (RNAP)/ribosome ratios, while strongly expressed genes have higher RNAP/ribosome ratios, as this is metabolically efficient.[11] Our model predicted that it is not possible to achieve high expression and high robustness with respect to the resources recruitment strength by only adjusting the RBS strength. There is a trade-off among protein expression, RBS strength, robustness, and flux of free resources. The RBS strength sets the sensitivity of the resources recruitment strength with respect to the flux of free resources. Thus, strong RBSs were predicted to be associated with resources recruitment strengths more sensitive to variations in the flux of free resources (i.e., at different growth rates) while weak RBSs provide robustness with respect to the growth rate. This defines how much of a given protein (e.g., ribosomal or nonribosomal) will be expressed at different growth rates.

This trade-off was consistent with the estimated values of the average transcription rates and RBS strengths we obtained for the cell endogenous ribosomal and nonribosomal genes. We found that the low RBS strength and high transcription rate of ribosomal genes make their resources recruitment strength robust with respect to changes in the flux of free resources with growth rate. On the contrary, for nonribosomal genes, our model predicted an average high RBS strength and a low transcription rate expression strategy. This differential strategy

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

allowed us to explain the relative mass fractions distributions of endogenous ribosomal and nonribosomal proteins as a function of growth rate. Thus, the differential expression strategies in *E. coli* encode the mass distribution of ribosomal and nonribosomal proteins for varying growth rates. Our model suggested that the cell achieves a fairly constant absolute expression of nonribosomal proteins using a high RBS strength to express them. On the other hand, the cell uses much weaker RBSs to express the ribosomal proteins. This way, the value of the total ribosomal RRS remains mostly constant with respect to the nonribosomal one. As a consequence, the absolute expression of ribosomal proteins increases with growth rate.

The results were applicable to the expression of exogenous protein-coding genes. For a given ratio of RRSs, increasing the expression of exogenous genes decreases the growth rate thereby reducing the absolute mass of endogenous proteins. However, the mass of exogenous proteins accumulates in the cell, which allows the total mass of the cell to increase even if the growth rate decreases. Two extreme cases can be differentiated: either the exogenous genes imposing negligible loading on the cell or strongly overloading it. In the first case, the exogenous proteins behave in an equivalent way to the endogenous ones. Therefore, all of the results obtained for the last are applicable. This situation is of interest in situations like, e.g., when designing gene synthetic circuits for feedback regulation of enzymes expression in metabolic pathways. In this case, one of the goals is that the exogenous circuit does not load the cell in excess, as this will affect the overall performance of the regulated pathway. In the highly overloaded scenario, the RRS no longer depends on the flux of free resources. This causes a diminished differential effect of the RBS and promoter strength. Yet, the different sensitivity of the RBS to the available substrate as a function of its strength still has consequences in scenarios with variable substrate. In between, the definition of the RRS allows us to consider a wide range of scenarios with varying cell burden.

4. 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 of our model and experimental data highlights the relevance of the cellular resources recruitment strength defined in our model as a key functional coefficient. Our resources recruitment strength coefficient allows us to explain the distribution of resources between the host and the genes of interest. Additionally, it shapes the relationship between the use of resources, cell growth, and protein productivity. This functional coefficient explicitly considers the interplay between the flux of available free resources and lab-accessible gene expression characteristics. In particular, the promoter and RBS strengths.

Though we only considered *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 microorganisms to encode the mass distribution between ribosomal and nonribosomal proteins as a function of cell growth rate. Weak transcription and strong translation and the complementary strong transcription and weak translation emerge as two potentially equally optimal strategies in the expression space but with different characteristics from the point of view of the sensitivity of the specific synthesis rate of the expressed protein to variations in the cell growth. The capacity of the defined resources recruitment strength functional coefficients to capture the interaction between growth, cell resources, and gene expression characteristics is reflected in the fact that the model was able to infer good predictions of the experimental distribution of the cell endogenous ribosomal and nonribosomal protein mass fractions when fitted to estimate the cell specific growth rate.

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. 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.[13] In this context, our model may also be useful for design purposes in synthetic biology where it can be used to design the proper promoter−RBS strategy depending on the desired behavior of the genes expression as a function of growth rate. In this sense, the resources recruitment strength can be used as a context-dependent intrinsic magnitude for the standard characterization of protein-coding transcription units.

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. As a consequence, it can be integrated within a multiscale framework that considers the macroscopic extracellular dynamics of the substrate and population of cells in a bioreactor. The model only requires as input a measure of the fraction of available substrate with respect to the saturated case and predicts both the resulting cell specific cell growth rate and the mass and mass rates of the expressed proteins. This makes its integration with constraint-based models of metabolism rather straightforward. The possibility to consider expression systems using orthogonal ribosomes can also be implemented without much difficulties. All this makes the model useful in the context of model-based design of gene synthetic circuits and protein expression systems.

5. METHODS

5.1. Model Parameters. Table SI.2 in Section SI.11 shows the set of parameters used in the model.

5.2. Estimation of the Parameters for Ribosomal and Nonribosomal Endogenous Proteins. We considered the model expressions at steady state in Section 2.1 and estimated the RBS-strength-related parameters K_b^k, K_u^k with the transcription rates ω_k with $k = \{r, nr\}$ and the fraction Φ_m so that our model provided a good fit of the specific growth rate at steady state.

The only input information given to the model was the value of the peptide chain elongation rate values $\nu_t(s_i)$ as a function of growth rate obtained from ref 22. This is equivalent to feeding the model only with the available amount of substrate s_i. To this end, we expressed the effective maximum translation rate as $\nu_t(s_i) = \nu f(s_i)$, where ν is the maximum attainable peptide synthesis rate (see eq 2) and $f(s_i) = s_i/(K_{sc} + s_i)$. Note that $f(s_i)$ is monotonous with the amount of intracellular substrate s_i. From the experimental values of $\nu_t(s_i)$ as a function

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303

of growth rate and knowing the maximum attainable peptide synthesis rate ν (see Table SI.2), we obtained the experimental values of $f(s_i)$ for each growth rate. We used these to feed our model. This is tantamount to feeding the model with the substrate s_i, but the value of the substrate- and host-dependent Michaelis constant K_{sc} needs not to be known.

Then, we fitted the model parameters using the experimental growth rate as the output to predict. So as not to penalize large errors in excess, which in our case are more prone to happen for larger values of the growth rate, we minimized the sum over the experimental data points of the absolute prediction error of the growth rate

$$I = \sum_{k=1}^{n_p} |\mu_{exp}(s_{i,k}) - \hat{\mu}(s_{i,k})|$$

We considered $N_r = 57$; and $N_{nr} = 1735$, 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 2.2). We also considered the average mRNA degradation rates $d_{m,r} = 0.16$ min^{-1} and $d_{m,nr} = 0.2$ min^{-1} (see Table SI.2). Using the value of ν in Table SI.2 and the range of l_e obtained in Section SI.14, we estimated $\frac{\nu}{l_e} \subset [40, 70]$ $molecule^{-1}\cdot min^{-1}$.

Moreover, 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 $K_b^k \subset [3, 15]$ $molecule^{-1}\cdot min^{-1}$ are found in the literature (see Table SI.2). We used a conservative upper bound $K_b^{max} = 10$ $molecule^{-1}$ for the search space, considering binding is diffusion controlled. From the literature, we also considered a search range for the dissociation rate $K_u^k \subset [3, 135]$ min^{-1}. Overall, these estimates gave us a range $K_{C^0}^k \subset [0.02, 0.2]$ $molecule^{-1}$ for the effective RBS strength under the assumption of intracellular substrate saturation. We ran 200 instances of the parameter fitting algorithm using the global optimization software MEIGO[31] (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. The resulting average best-fit estimated parameters are given in Table 1.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.1c00131.

Biochemical pseudoreactions of the model; assumptions and detailed derivation of the burden-aware dynamic model; estimation of model parameters from experimental data and additional results relating the spMSR with the substrate availability and the cell growth rate as a function of the expression space; and software code for the model computational simulations and parameters optimization (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jesús Picó − *Synthetic Biology and Biosystems Control Lab, Institut d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, 46022 Valencia, Spain;* orcid.org/0000-0003-4144-3521; Email: jpico@upv.es

Authors

Fernando N. Santos-Navarro − *Synthetic Biology and Biosystems Control Lab, Institut d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, 46022 Valencia, Spain;* orcid.org/0000-0003-2710-8612

Alejandro Vignoni − *Synthetic Biology and Biosystems Control Lab, Institut d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, 46022 Valencia, Spain*

Yadira Boada − *Synthetic Biology and Biosystems Control Lab, Institut d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, 46022 Valencia, Spain;* orcid.org/0000-0001-5677-2702

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.1c00131

Author Contributions

J.P. and F.N.S.-N. conceived the study and designed the experiments. F.N.S.-N. implemented the method. F.N.S.-N., A.V., and Y.B. performed the experiments. F.N.S.-N., A.V., and Y.B. performed data analysis and visualization. J.P. and F.N.S.-N. wrote the manuscript. All authors participated in the writing of the revised manuscript. All of the authors edited and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially supported by grants MINECO/AEI, EU DPI2017-82896-C2-1-R, and MCIN/AEI/10.13039/501100011033 grant number PID2020-117271RB-C21. F.N.S.-N. is grateful to grant PAID-01-2017 (Universitat Politècnica de València). The authors are very grateful to the anonymous reviewers for their comprehensive and in-depth reviews.

■ REFERENCES

(1) Gorochowski, T. E.; Chelysheva, I.; Eriksen, M.; Nair, P.; Pedersen, S.; Ignatova, Z. Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol.* **2019**, *15*, No. e8719.

(2) Cardinale, S.; Arkin, A. Contextualizing context for synthetic biology- identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **2012**, *7*, 856−866.

(3) Towbin, B. D.; Korem, Y.; Bren, A.; Doron, S.; Sorek, R.; Alon, U. Optimality and sub-optimality in a bacterial growth law. *Nat. Commun.* **2017**, *8*, No. 14123.

(4) Qian, Y.; Huang, H.-H.; Jiménez, J. I.; Vecchio, D. D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **2017**, *6*, 1263−1272.

(5) de Lorenzo, V. Evolutionary tinkering vs. rational engineering in the times of synthetic biology. *Life Sci. Soc. Policy* **2018**, *14*, 1263−1272.

(6) Sabi, R.; Tuller, T. Modeling and measuring intracellular competition for finite resources during gene expression. *J. R. Soc. Interface* **2019**, *16*, No. 20180887.

(7) Jayanthi, S.; Nilgiriwala, K.; Del Vecchio, D. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synth. Biol.* **2013**, *2*, 431−441.

(8) Carbonell, M.; et al. Dealing with the genetic load in bacterial synthetic biology circuits: covergences with the Ohm's law. *Nucleic Acids Res.* **2016**, *44*, 496−507.

(9) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **2010**, *330*, 1099−1102.

(10) Bienick, M. S.; Young, K. W.; Klesmith, J. R.; Detwiler, E. E.; Tomek, K. J.; Whitehead, T. A. The interrelationship between promoter strength, gene expression, and growth rate. *PLoS One* **2014**, *9*, No. e109105.
(11) Gorochowski, T. E.; Avcilar-Kucukgoze, I.; Bovenberg, R. A. L.; Roubos, J. A.; Ignatova, Z. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth. Biol.* **2016**, *5*, 710−720.
(12) Bosdriesz, E.; Molenaar, D.; Teusink, B.; Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J.* **2015**, *282*, 2029−2044.
(13) Weiße, A. Y.; Oyarzún, D. A.; Danos, V.; Swain, P. S. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E1038−E1047.
(14) Guillaume, J.; Goelzer, A.; Tebbani, S.; Dumur, D.; Fromion, V. Dynamical resource allocation models for bioreactor optimization. *IFAC PapersOnLine* **2018**, *51*, 20−23.
(15) Macklin, D. N.; et al. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* **2020**, *369*, No. eaav3751.
(16) Blanchard, A. E.; Liao, C.; Lu, T. Circuit-Host coupling induces multifaceted behavioral modulations of a gene switch. *Biophys. J.* **2018**, *114*, 737−746.
(17) Boada, Y.; Vignoni, A.; Oyarzún, D.; Picó, J. Host-circuit interactions explain unexpected behavior of a gene circuit. *IFAC-PapersOnLine* **2018**, *51*, 86−89.
(18) Frei, T.; Cella, F.; Tedeschi, F.; Gutierrez, J.; Stan, G.; Khammash, M.; Siciliano, V. Characterization, modelling and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* **2020**, *11*, 4641.
(19) Hausser, J.; Mayo, A.; Keren, L.; Alon, U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nat. Commun.* **2019**, *10*, No. 68.
(20) Zheng, H.; Bai, Y.; Jiang, M.; Tokuyasu, T. A.; Huang, X.; Zhong, F.; Wu, Y.; Fu, X.; Kleckner, N.; Hwa, T.; Liu, C. General quantitative relations linking cell growth and the cell cycle in *Escherichia coli*. *Nat. Microbiol.* **2020**, 995.
(21) Si, F.; Li, D.; Cox, S. E.; Sauls, J. T.; Azizi, O.; Sou, C.; Schwartz, A. B.; Erickstad, M. J.; Jun, Y.; Li, X.; Jun, S. Invariance of Initiation Mass and Predictability of Cell Size in *Escherichia coli*. *Curr. Biol.* **2017**, *27*, 1278−1287.
(22) Bremer, H.; Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. *EcoSal Plus* **2008**, DOI: 10.1128/ecosal.5.2.3.
(23) Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **2010**, *38*, D750−D753.
(24) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedi of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27−30.
(25) Metzl-Raz, E.; Kafri, M.; Yaakov, G.; Soifer, I.; Gurvich, Y.; Barkai, N. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife* **2017**, *6*, No. e28034.
(26) Dai, X.; Zhu, M. Coupling of Ribosome Synthesis and Translational Capacity with Cell Growth. *Trends Biochem. Sci.* **2020**, *45*, 681−692.
(27) Rosano, G. L.; Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **2014**, *5*, No. 172.
(28) Oyarzún, D. A.; Stan, G.-B. V. Synthetic gene circuits for metabolic control: Design trade-offs and constraints. *J. R. Soc., Interface* **2013**, *10*, No. 20120671.
(29) Ceroni, F.; Algar, R.; Stan, G. B.; Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **2015**, *12*, 415−418.
(30) Nikolados, E. M.; Weiße, A. Y.; Ceroni, F.; Oyarzún, D. A. Growth Defects and Loss-of-Function in Synthetic Gene Circuits. *ACS Synth. Biol.* **2019**, *8*, 1231−1240.
(31) Egea, J. A.; Henriques, D.; Cokelaer, T.; Villaverde, A. F.; MacNamara, A.; Danciu, D.-P.; Banga, J. R.; Saez-Rodriguez, J. MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinf.* **2014**, *15*, No. 136.

https://doi.org/10.1021/acssynbio.1c00131
ACS Synth. Biol. 2021, 10, 3290−3303