Dose-Response Profiling in Heterologous Gene Expression: Insights from Proteome Fraction Analysis

Nicolás A. Vaccari,[†] Dahlin Zevallos-Aliaga,[†] Tom Peeters,[‡] and Daniel G. Guerra[*,†]

†*Unidad de Moléculas Individuales, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, Perú*

‡*Open BioLab Brussels, Erasmushogeschool Brussel, Brussels, Belgium*

E-mail: daniel.guerra@upch.pe

Phone: +51 319-0000 ext.233217

Abstract

To achieve precise control over the expression of heterologous genes, crucial to applications like biosensors and protein manufacturing, it is essential to accurately measure the system's response to varying levels of the inducer. This study introduces a formal approach, inspired by proteome fraction concepts, to analyze titration experiments with a molecular inducer. By deriving the proteome fraction definition with respect to time, we introduce the concept of expression fraction, representing the portion of the anabolic rate dedicated to the production of a specific set of proteins. Formal deduction reveals that the expression fraction equals the proteome fraction when it is at its maximum point, enabling its reliable measurement through direct quantitation of protein amounts. Additionally, experimental data from microcultures and fluorescent signals in three biosensors and six cellular contexts show a linear

correlation between protein production and growth rate during exponential growth, indicating a constant expression fraction through this time window. The values of expression fraction obtained from the slopes of such intervals or those obtained from maximum points of protein amount can be fitted independently to a Hill function and produce remarkably similar dose-response parameters. Thus, this conceptual framework provides two consistent methods for determining regulatory parameters. The model is further applied to explore possible mechanisms for how the heterologous expression negatively impacts growth rates in different strains, with *Escherichia coli* B strains generally exhibiting greater tolerance than K12.

1. Introduction

Synthetic biology involves the design and construction of novel genetic circuits, metabolic pathways, and engineered proteins[1–4]. In this field, the development of biosensors arises as a very important tool for the detection of analytes and for improving the yield of new self-regulated biosynthetic pathways. This technological advancement has captured the attention of numerous industries due to its potential to produce high-value biochemical compounds at an industrial scale[5,6]. We see that the literature accumulates transcription factors that are characterized as tools for the regulated expression of heterologous genes, typically assessing their dose-response in terms of threshold, limit of detection, and dynamic range[7–10]. We note, however, the absence of standardized or formalized procedures to address inherent variability, especially when selecting culture time points that enable reliable averaging of measurements across replicates or facilitate comparisons of outputs under different conditions.

In this paper, we make use of a conceptual framework developed in recent years for studying bacterial metabolism and growth regulation[11–17]. This framework is built upon the fundamental assumption that proteins with similar functions can be grouped together as a single entity, regulated by similar (though not necessarily identical) mech-

anisms. This simplification allows to view the entire proteome as comprised by a few proteome fractions[18].

In this work by expanding this concept with a focus on the fractionation of the cell's biosynthesis rate, we obtained a simple theorem for the evolution of proteome fractions in time. Corollary to this theorem, we formally demonstrate that the fraction of the cell's biosynthesis rate dedicated to a set of proteins can be determined directly by analysing a single point, in a peak of its accumulation, or by analysing an interval during which such expression is linear to the growth rate.

We test these two methods analysing inducer-titration experiments for sensors regulating the expression of fluorescent proteins. In this way, we analyse the production of a reporter protein in relation to an inducer and the culture's growth rate, linking easily observable data to the underlying molecular mechanisms that govern the sensitivity of a heterologous circuit.

Finally, we apply our model to explore what are the factors contributing to variations in productivity and metabolic burden that we observe between the B and K12 strains of *E. coli*.

2. Methods

2.1 *E. coli* strains

MG1655 ΔendA ΔrecA (DE3) was a gift from Kristala Prather (Addgene code # 37854)[19]. Agar stabs of BLR(DE3) and BL21(DE3) were kindly provided by Carlos Bustamante at UC-Berkeley and Mauricio Baez at Universidad de Chile, respectively. Cells were made competent by $CaCl_2$ incubation protocol and transformed by 42°C heat shock with pUC-T7GFP or sequentially with pUC-T7GFP and pACYCDUET. Selection Luria Broth (LB) agar plates were supplemented with 1% glucose to reduce T7 basal expression, and ampicillin to select for pUC-T7GFP plasmid, or ampicillin and chloramphenicol to select for

pUC-T7GFP and pACYCDUET plasmids.

2.2 Plasmids

A gene block for GFP expression was designed using Benchling RRID:SCR_013955 [Biology Software] (2020). Retrieved from https://benchling.com. Twist Bioscience synthesized the DNA fragment. It contained the following elements: T7 promoter, lac operator, Shine-Dalgarno ribosome binding sequence, a codon-optimized GFP mut3b coding sequence[20] fused with LVA tag for rapid degradation (AANDENYLVA)[21], and T7 terminator. A high-copy GFP expression plasmid, pUC-T7GFP, was built by inserting this construct into a pUC57 vector through standard digestion and ligation reactions between NcoI and KpnI restriction sites. To introduce additional expression of LacI, we used the low-copy number plasmid, pACYCDUET, acquired from GenScript.

2.3 Culture conditions

Unless indicated otherwise, all data presented here were obtained from 200 μL microcultures in M9 medium supplemented with 0.2% casamino acids and 0.24% glucose, prepared as following: Freshly transformed colonies were hand-picked to inoculate 10 mL of M9 medium supplemented with ampicillin or ampicillin and chloramphenicol. These cultures were incubated at 37°C and 250 RPM overnight, then measured for cell density and harvested by centrifugation at 5000 RPM for 5 minutes. Immediately, the cells were resuspended in fresh medium to the appropriate volume to obtain an optical density of 0.1, which equated to approximately 2 x 10^7 cells/mL. Finally, 200 μL of these suspensions were loaded in a 96 multiwell plate. The expression of reporter proteins was induced on 195 μL of culture by manually pipetting 5 μL of the appropriate concentration of IPTG, $HgBr_2$ or cumate reaching the final concentration indicated for each case, completing 200 μL in each well. These microcultures were incubated at 30 °C with 220 RPM oscillatory shaking for 16 hours without the addition of antibiotics. The multiwell

plate was covered with its plastic lid, and the outer wells were filled with 200 μL of water so that a similar humidity surrounded all the monitored cultures. This configuration resulted in less than 2.5% volume loss due to evaporation after 16 hours.

Additionally, 50 mL cultures were processed as controls for total protein determination and SDS-PAGE analysis. In a similar way as performed for microcultures, 50 mL cultures were inoculated to an OD_{600} of 0.1, incubated at 30 °C with 250 RMP, and induced with 0.5 mM IPTG. For their analysis, aliquots were harvested at 0.5, 2 and 4 hours after induction and then lysed though three cycles of freeze thawing (15 minutes at –70°C and 15 minutes at 37°C).

2.4 Measurements

Culture density (OD_{600}) and fluorescence were monitored in a plate reader, Infinite 200 Pro from TECAN. The production of GFP was registered using the following settings: 475 nm Excitation and 516 nm Emission, signal gain set at 50 out of 100. RFP expression was registered using the following settings: 585 nm Excitation and 608 nm emission, signal gain set at 95. In both cases, the fluorescence was collected with 25 flashes, reading every 15 minutes.

2.5 Calculation of the heterologous fraction, growth rate and production rate

We considered fluorescence as directly proportional to the reporter protein (GFP or RFP) mass, and the optical density as a measurement of the total mass (also proportional to the total protein mass). Thus, to estimate the heterologous fraction, φ_H, we measured the ratio of Fluorescence over Optical density, $flu(t)/OD_{600}(t)$.

To calculate the growth rate at each time point, $\mu(t)$, we measured the optical density at t and at the immediate next time point $t + \Delta t$. The difference between both measure-

ments was divided by Δt and the optical density at the initial point t:

$$\mu(t) = \frac{OD_{600}(t+\Delta t) - OD_{600}(t)}{\Delta t} \cdot \frac{1}{OD_{600}(t)} \tag{1}$$

Similarly, to calculate the production rate of GFP (or RFP), ρ_H, we measured the specific fluorescence at the points t and at next time point $t+\Delta t$. This difference was divided by the time step Δt, and the optical density:

$$\rho_H(t) = \frac{flu(t+\Delta t) - flu(t)}{\Delta t} \cdot \frac{1}{OD_{600}(t)} \tag{2}$$

3. Results & Discussions

3.1 Interrelation between growth and heterologous expression.

We collected data from three *E. coli* strains: MG1655, BL21, and BLR, all containing the DE3 insert. As originally described by Studier and Moffatt in 1986, this insert codes for T7 RNA polymerase under the control of LacI repressor. Upon binding to its inducer (allolactose, or its analog IPTG), the expression of T7RNAP is allowed. These strains were transformed with pUC57-T7-GFP plasmid, which codes for GFP expression under the control of a T7 promoter followed by a LacI binding site. This setup replicates one of the most used protein expression systems at the laboratory-level[23] , where the gene of interest is controlled by LacI repressing the T7RNAP gene and the T7 promoter itself. To further evaluate the impact of LacI repressor levels, we transformed the three tested strains with a second plasmid (pACYCDUET), encoding additional LacI repressor expression. This resulted in six distinct configurations for monitoring heterologous protein production.

Titrating the expression of GFP with different IPTG concentrations, we observe -as previously described by Scott et al. - an inverse correlation between the maximum growth rates achieved in each culture and the accumulated protein production at the time of

maximum growth rate (Figure 1). Also, BL21(DE3) was not able to repress the heterologous expression without the accompanying pACYCDUET plasmid, therefore IPTG titration was evaluated in BL21(DE3) only with this extra production of LacI repressor. Similar experiments without pUC57-T7-GFP plasmid showed that IPTG has little or no effect on growth rate (Figure 1S). Therefore, the negative slopes of Figure 1 can be interpreted as caused by the burden that the production of heterologous GFP exert on each culture.

A previous model created by Scott et al. explains that such reduction of maximum growth rate takes place because the expression of heterologous proteins competes against the ribosomal proteins as fractions of the same -limited- proteome space (Figure 2C). However, those models are insufficient to explain the differences that we observe here between the B and K12 strains and between the presence and absence of additional LacI expression (by pACYCDUET plasmid). First, they were derived from data of only MG1655 (a K12 strain) provided with native transcription factors and promoters, while here we observe that the link between growth and expression is profoundly affected by the addition of extra LacI repressor. Second, with a focus on the native regulation of the biosynthetic machinery, they attempted to describe a steady-state, represented at the point of maximum growth, which is not necessarily consistent with the most common laboratory practices for heterologous systems. When analyzing the dose-response of a biosensor to an analyte, or when optimizing the production of heterologous proteins, it is common to accumulate biomass before the induction without formally testing if the heterologous expression circuit reaches a steady state. In our case, the tested strains reached their maximum growth rates at different times of incubation, making it difficult to decide when it could be valid to analyze and compare their yield and the interdependence between yield and growth (Fig 3S). Therefore, to analyse how a culture responds to the activation of heterologous expression, we must find an appropriated measuring point, and to do so we must consider the evolution of the culture in time rather than just the point of maximum growth rate.

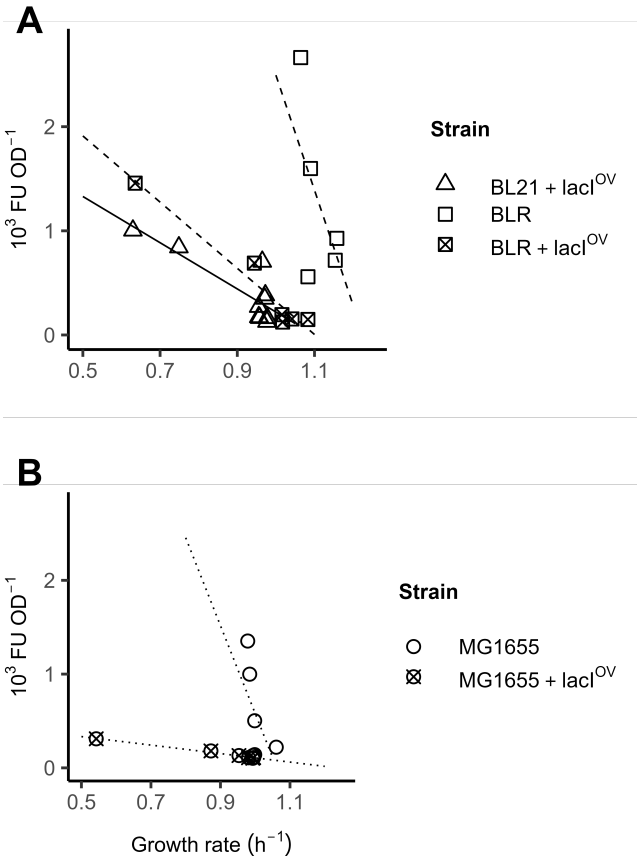


Figure 1: Titrating IPTG increases heterologous protein and decreases growth rate. Microcultures were started at a density of OD600 = 0.1 and induced with different concentrations of IPTG from this point. The maximum growth rate was obtained for each case and used to plot the corresponding fluorescence over cellular density at that point. Values correspond to triplicate cultures of three identical colonies. Lines show the linear regression for each data set. (A), B strains: BL21(DE3) and BLR(DE3) with and without pACYC for extra expression of LacI repressor. (B), MG1655(DE3) with and without pACYC for extra expression of LacI.

3.2 Model building

A theorem for the evolution of proteome fractions in time.

We conceptualized that the general biosynthesis rate of the bacteria is divided into a manifold of rates producing different types of protein that are termed proteomic fractions (φ_i). In a similar way as it has been done in previously reported models[11–17], proteins will be grouped together as a unified proteome fraction if they share a similar biological function and are regulated by the same cues[14,18]. One fraction of the proteome will be constantly expressed fulfilling core, structural functions (φ_Q). The two other fractions are variable and perform nutrient assimilation ($\varphi_C(t)$) and biomass production ($\varphi_R(t)$), respectively. Importantly, the total flux through the latter, 'ribosomal' fraction, φ_R, equates to the biosynthesis rate times the total proteome mass m (rate ρ in figure 2D). The relative proportion of this flux allocated to each proteome fraction $\varphi_i(t)$ will be denoted as the expression fraction and is represented by $f_i(t)$ (f_R, f_Q, f_H, f_C in Figure 2D).

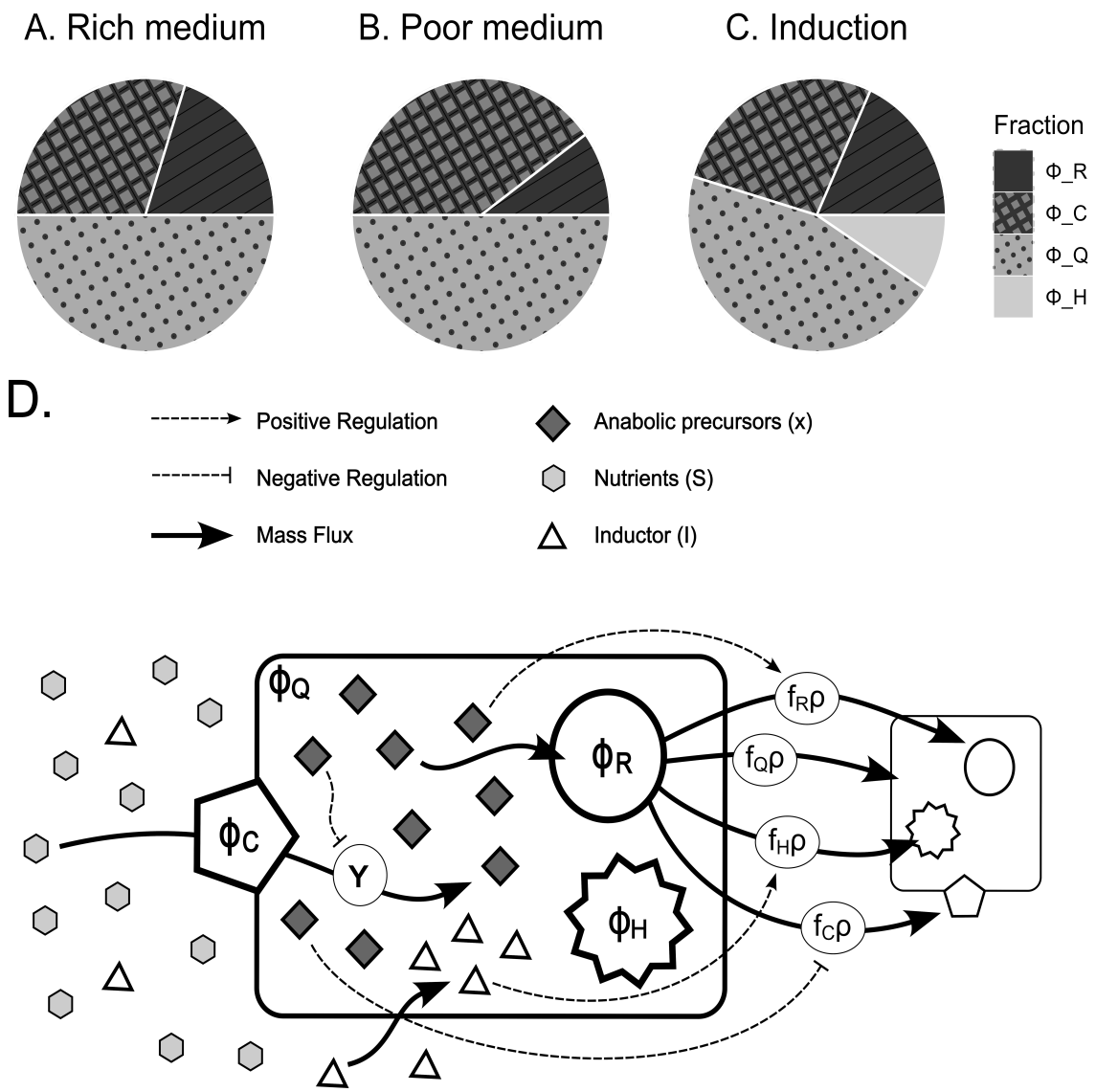
With these definitions, solving the derivation of a proteome fraction over time (full derivation in the Biosynthesis Allocation Theorem section of the supplementary material) produces an equation that links the change in a proteome fraction to its expression fraction and the total biosynthesis rate:

$$\frac{d\varphi_i(t)}{dt} = \rho(t) \cdot [f_i(t) - \varphi_i(t)] \tag{3}$$

3.3 Testing the model

Solving the regulation of a sensor at the maximum fraction point.

A remarkable consequence of the equations 3, is that for any proteome fraction, when its differential over time equals zero ($\frac{d\varphi_i(t)}{dt} = 0$), the expression fraction and the proteome fraction must be equal, $f_i(t) = \varphi_i(t)$ (assuming that the cell's anabolic rate is never zero).



E. $Y = \beta \cdot \phi_C \cdot \frac{S}{S+k_S} \cdot \frac{k_1}{x+k_1}$

F. $\rho = \gamma \cdot \phi_R \cdot \frac{x}{x+k_2}$

G. $\rho = \sum_i (f_i \cdot \rho)$

Figure 2: A model of proteome allocation and metabolic rate. Representation of the proteome fractioning of a cell: A) in a rich medium maximizing its growth rate through the activity of ribosomal fraction, φ_R; B) in a poor medium, maximizing its nutrient fixation through the activity of nutrient fixators fraction, φ_C; C) producing a heterologous fraction, φ_H, causing a redistribution of the native proteome allocation. D) Representation of rates within bacteria.

The light gray hexagons represent the nutrients, $S(t)$, which are converted into anabolic precursors, $x(t)$, (dark gray diamonds) by the nutrient fixator's fraction, φ_C (pentagon) at a rate Y. The anabolic precursors are then used by the ribosomal fraction, φ_R (circle) at a rate ρ, to produce new mass (represented by the small cell on the right side). The biosynthesis is divided into a manifold of different protein production rates (black arrows) with their respective expression fractions, $f_R(t)$, $f_C(t)$, $f_Q(t)$, $f_H(t)$; the concentration of anabolic precursors stimulates the rate of φ_R production and represses the rate of φ_C production; the rate of φ_Q production is constant and the rate of φ_H production is activated by its specific inducer (triangles) which diffuses from the extracellular space. E) Rate equation for nutrient fixation, $Y(t)$. F) Rate equation for total proteome biosynthesis, $\rho(t)$. The total biosynthetic activity is modeled as the specific rate through one enzyme (Detailed in supplementary material). G) The total biosynthesis rate is the sum of various anabolic sub-rates, each one with a specific expression fraction, $f_i(t)$.

Therefore, regardless of the growth behavior of a culture, registering the heterologous fraction when it has reached a maximum should correctly represent the regulation of the circuit. This analysis leads us to experimentally characterize the expression fraction for a heterologous reporter protein, f_H, by plotting the heterologous proteome fraction at its maximum point in multiple inducer's titration assays. Then, using the Hill function for representing the regulation of the heterologous fraction[24], we obtain that:

$$\varphi_H^{max}(I) = f_H(I) = H \cdot \frac{I^n}{(H+1) \cdot I^n + {k_I}^n} \tag{4}$$

Here, H represents the maximum possible expression for the circuit, k_I is related to the affinity for the inducer, I, and n is the Hill's coefficient expressing the cooperativity of the response to the inducer. Because the experimental curves represent the behavior of the whole system, k_I and n will not necessarily correspond to the dissociation constant and the molecular cooperativity of the binding reaction between the transcription factor and the inducer. We verified by SDS PAGE analysis that even for the highest GFP overexpression, H represented at most about 10% of total biomass (Figure 9S); therefore, we infer that in most cases, the $(H+1)$ coefficient can be approximated to 1, thus resulting a typical Hill function .

We confirmed that the analysis of experimental data based on φ_H^{max} yields a Hill func-

tion for the T7 expression system as we tested it in three different *E. coli* strains with or without additional LacI repressor (Figure 3A). To test the generality of this method, we analysed two additional systems in the same way: a cumate biosensor based on CymR repressor (Fig. 5S, and Table 1) and a mercury biosensor based on MerR transcription factor (Figure 6S, and Table 1). The coefficients of determination indicate excellent fitting with the exception of the mercury system, where only a transient, local φ_H^{max} was identified (Table 2[1]).

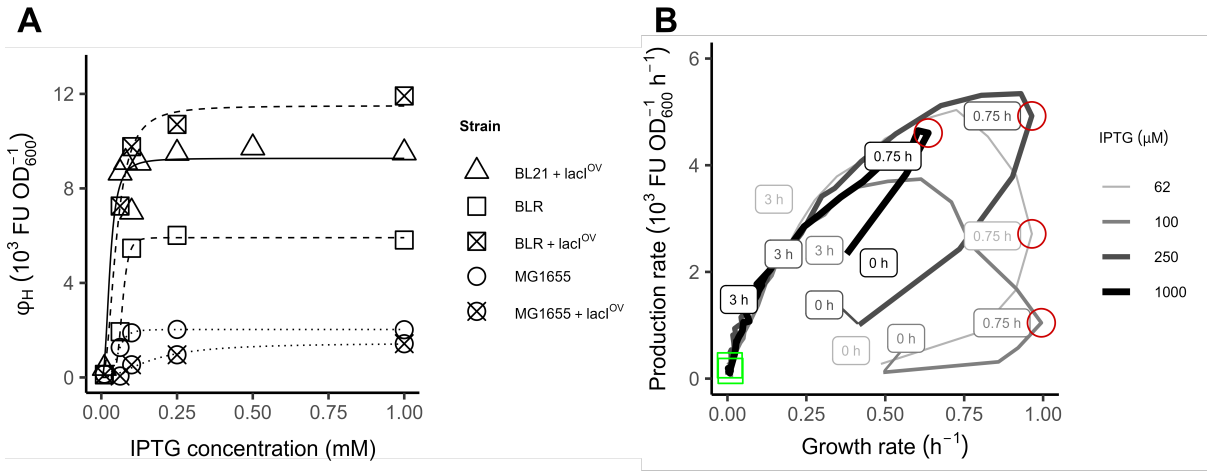


Figure 3: (A) **Heterologous fraction at its maximum points as a function of IPTG concentration.** Markers correspond to maximum values of fluorescence over cellular density obtained from triplicate cultures of three identical colonies (n = 9) with the indicated IPTG concentration concentration in 5 different cellular contexts. Standard errors are smaller than the data symbols and are not indicated. Lines show the Hill function fitting for each data set. (B) **Production rate Vs Growth rate in BL21(DE3) + LacI^{OV}**. Each line represents one culture from the time of induction (0 h) until the onset of the stationary phase, when growth rate approaches zero (typically around 5 hours). Circles signal the point of maximum growth rate. Squares signal the points selected following the criteria of heterologous fraction maxima, φ_H^{max}. The cultures draw the shape of a counterclockwise 'hook' by increasing and then lowering the growth and production rates. Almost all the cultures reached an approximately constant linear relation between production and growth after 0.75 h. Higher concentrations of IPTG led to shorter nonlinear behaviours.

[1]All cultures in the mercury titration experiment showed an early peak at 30 mins, presumably too early for the slow equilibration of extracellular Hg^{2+}. The reporter expression continued to accumulate beyond the stationary phase, preventing the identification of a true φ_H^{max}

Sensor	Strain	$H(flu/OD_{600})$		$k_I(\mu M)^a$		n	
		Maxima	Slopes	Maxima	Slopes	Maxima	Slopes
IPTG	**BL21(DE3) + LacI^{OV}**	9757 ± 420	8215 ± 513	37.2 ± 8.1	12.5 ± 9.1	2.2 ± 0.7	1.3 ± 0.6
	BLR	5915 ± 78	6198 ± 225	69 ± 1.1	67 ± 4.2	6.75 ± 0.7	10 ± 7.8
	BLR + LacI^{OV}	11499 ± 337	10240 ± 88	49.2 ± 4.4	50 ± 1.4	2.3 ± 0.5	3.1 ± 0.3
	MG1655(DE3)	2029 ± 59	2928 ± 85	54.9 ± 3.4	55 ± 23	4.3 ± 1.5	10 ± 38
	MG1655(DE3) + LacI^{OV}	1458 ± 148	3498 ± 792	164 ± 33	236 ± 117	1.9 ± 0.5	1.4 ± 0.6
Cumate	**MG1655(DE3)**	1546 ± 72	4731 ± 597	256 ± 20	340 ± 62	1.7 ± 0.2	2.1±0.6
Hg^{2+}	**NEB® Stable**	7749 ± 2170	8847 ± 184	0.93 ± 0.44	0.16 ± 0.01	1 (fixed)	1.3 ± 0.05

Table 1: **Hill function parameters describing the sensors responses.** The Hill function was fitted to the values of the fraction of expression plotted as a function of inducer concentration (IPTG, cumate, or $HgBr_2$). The fraction of expression was determined through two methods: Maxima, the fraction of expression is measured as equal to the heterologous protein fraction at its maximum value, $f_H = \varphi_H^{max}$. *Slopes*, the fraction of expression is determined by measuring the slope within the linear interval of the production rate Vs growth rate plot, $f_H = \frac{\Delta\rho_H}{\Delta\mu}$.

[a] We didn't use a true ϕ_H^{max}, in its place we used a local one to do the calculations.

3.4 Solving the regulation of a sensor during exponential growth.

In the cases analyzed here, we observed that the maximum point for the heterologous fraction is reached close to or during the stationary phase. It will be important for biosensors developers, especially in the field of metabolic engineering, to test if the Hill parameters thus obtained keep their validity when the heterologous expression is quantified closer to the steady state, during exponential growth. During exponential growth, the protein to total cellular mass ratio remains relatively constant[25], then we can approximate $\rho(t)$ to the growth rate $\mu(t)$. This approximation yields a convenient equation since measuring the growth rate directly is simpler than determining the protein production rate.

$$\frac{d\varphi_i(t)}{dt} = \mu(t) \cdot [f_i(t) - \varphi_i(t)] \tag{5}$$

Then, to test the validity of the Hill parameters determined using the heterologous maximum point (φ_H^{max}, where $f_H = \varphi_H^{max}$), we graphed the fluorescence production rate (ρ_H) against the growth rate (μ) at various time points. As shown in figure 3B (and 4S), data points from each culture draw a non-linear behaviour (hook shape) followed by a linear portion. Importantly, we verify that φ_H^{max} occurs within the linear interval, and therefore the Hill parameters ($H,\ k_I, n$) obtained from data of the equilibrium points (where $f_H = \varphi_H$) should be valid through the interval where the linear trend is maintained, which includes a significant portion of the exponential growth. Moreover, because in this phase of growth the general anabolic rate is represented by the growth rate ($\rho \approx \mu$), then it follows that the linear segment of the graph indicates an interval where the fraction of expression of the heterologous protein, f_H, is constant and is equal to the slope.

In Figure 4 (and Figure 5S) we present a visual comparison of the f_H determined as equal to the slope of production Vs growth rate, or as equal to the maximum of heterologous fraction, φ_H^{max}. In Table 1 we present the Hill parameters obtained from fitting these f_H values as a function of the inducer concentration in different expression systems. The

Table 2: Coefficients of determination for the fitted Hill functions displayed in Table 1, using the φ_H^{max} and slope data.

Sensor	**Strain**	R^2	
		Maxima	Slopes
IPTG	**BL21 + LacI^{OV}**	0.997	0.903
	BLR	0.999	0.997
	BLR + LacI^{OV}	0.996	0.985
	MG1655(DE3)	0.996	0.997
	MG1655(DE3) + LacI^{OV}	0.979	0.992
Cumate	**MG1655(DE3)**	0.999	0.989
Hg^{2+}	**NEB® Stable**	0.599*	0.794

similarity of the results obtained through these two different methods further proves that the expression fractions are constant within these intervals. Also, it validates the linear regression as a method for determining f_H. This alternative method of analysis will be of special value for sensor circuits that don't reach a conspicuous maximum of heterologous protein φ_H^{max} such as in the case of the mercury sensor (Figure 6S).

3.5 Using the model

Exploring possible mechanisms for the strain-specific growth rate sensitivity to heterologous protein production.

We contemplated if our model could account for the divergent behaviors that were noted in K12 and B strains and between the presence or absence of additional LacI repressor (i.e., the different slopes that we observe in Figure 1). Specifically, we inquired what modifications in our model's parameters could simulate strains that are more tolerant or more vulnerable to have an impact in growth rate because of heterologous protein production. Manipulating all the values of the constants and coefficients failed to yield varying slopes in a graph of maximum growth rate Vs heterologous proteome fraction, like Figure 1; consequently, we pursued more substantial adjustments by suggesting two hypotheses.

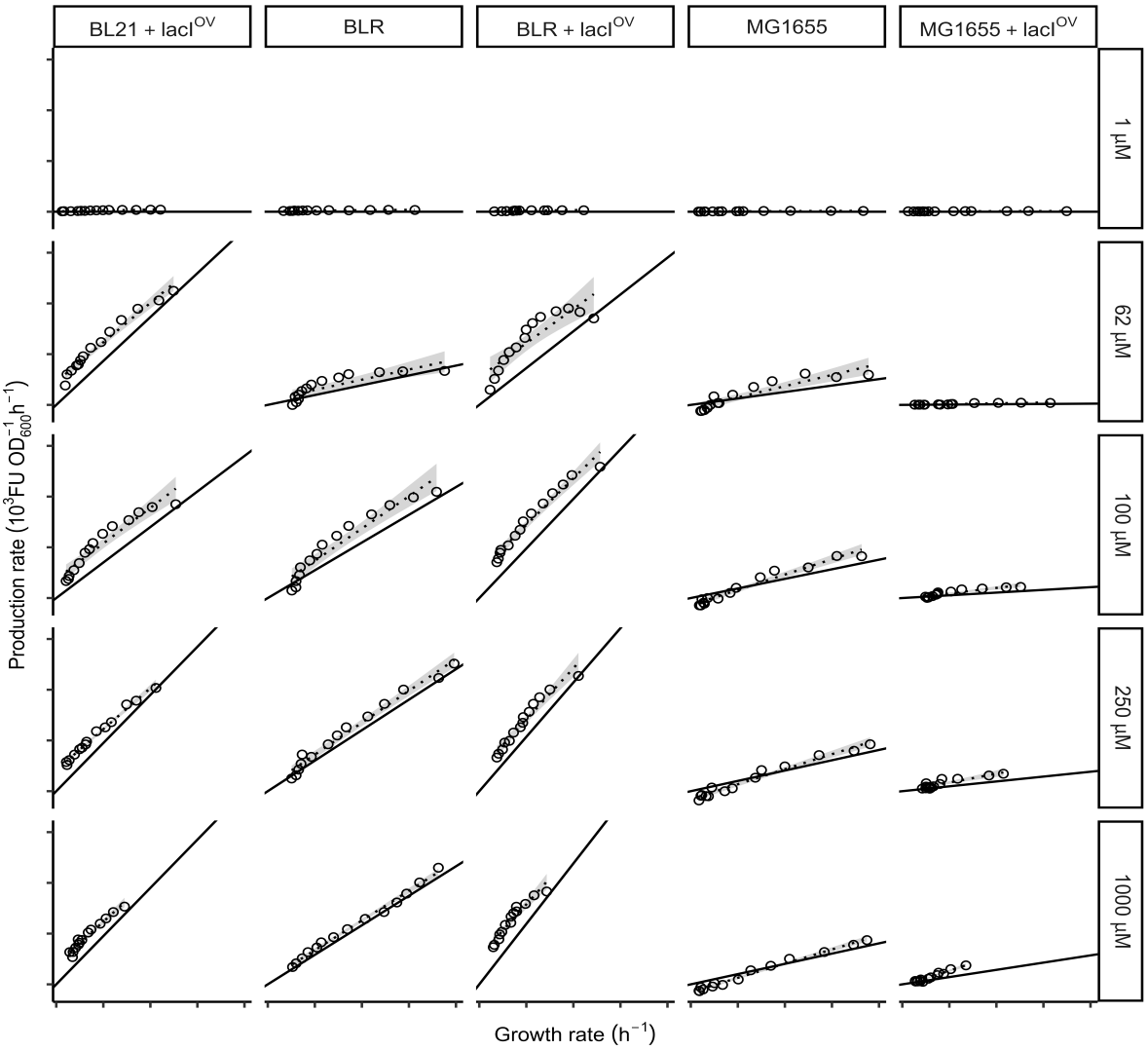


Figure 4: **Expression fractions at different IPTG concentrations as obtained from two analysis methods**. Dots are directly plotted from experimental data and then fitted to a linear regression with its confidence interval (0.95) shaded in gray. Each solid line represents a line function where the slope is set to be equal to the value of φ_H^{max} found at the indicated condition and strain. Similar analysis with CymR-Cumate, and MerR-Hg systems are shown in Figures 5S and 6S in SUPP MATERIAL.

The first hypothesis proposes that B strains, due to their knocked-out genes, might have an excess of anabolic capacity. We would represent such excess capacity as a pool of inactive ribosomes that are present but have nothing to translate. We thought that this pool could activate when the heterologous RNAs are produced. For this "slacker ribosome" hypothesis to be true, the protein-per-cell ratio should be higher for induced bacteria because a group of ribosomes would suddenly start producing extra protein. We tested this hypothesis experimentally by measuring the total protein of the cultures and normalizing it by dividing it by the OD600 of the sample at three different culture times: 30 minutes, 2 hours, and 4 hours. We could not determine a significant difference in this ratio for induced and uninduced bacteria (Fig. 8S), thus refuting this hypothesis. Nevertheless, we simulated the slacker ribosomes hypothesis to see if the model could emulate the observed differences among *E. coli* strains. Altering the value of inactive ribosomes $\varphi_{R,0}$ (see Fig. 5 and supplementary material for details), the model predicts differences in the basal growth rate, without heterologous induction (Figure 5B). This prediction clearly differs from experimental observation, because in the absence of inducer, B and K12 strains with and without pACYCDUET plasmid show similar growth rates (1 – 1.1 h^{-1}, Figure 1).

Our second hypothesis proposed that the bacteria could be producing stress-related proteins in response to GFP, as it has been reported that strong heterologous protein expression lead to general stress[26,27]. To test this idea in our model, we added a new proteome fraction induced as a response to the presence of GFP. Such a response could be, for example, that stress sigma factors respond to heterologous misfolded proteins. By varying the sensitivity of the simulated stress response, we obtained predictions that had the same behavior as our experimental results (Figure 5C).

According to this interpretation, the BLR(DE3) should have the lowest production of stress-related proteins induced by the presence of GFP. The mutated, partially functional version of sigma factor S, σ_S, present in BLR at difference with BL21 might explain this

behavior[28]. In contrast, MG1655(DE3) shows the highest sensitivity against GFP expression, implying a greater stress-related response, thus displacing φ_R and φ_C and leading to a more severe reduction in the growth rate.

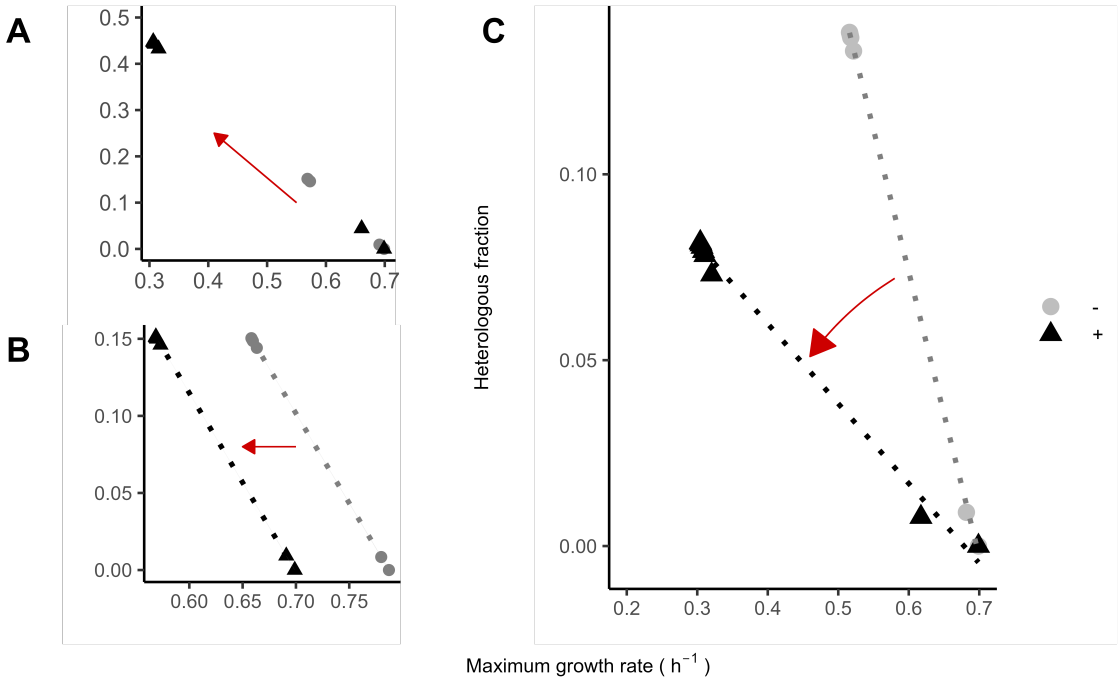


Figure 5: **Heterologous proteome fraction vs. maximum rate in different simulations.** Each point represents a simulation run with a different concentration of inducer. Each graph represents two sets of simulations run with our original settings (grey) and modified settings (black). A) Changing the maximum capacity of the circuit (H, in equation (4)). Arrow indicates higher values for H. B) Adding initially passive, "slacker" ribosomes that only activate upon induction. C) Introducing an additional protein fraction designed to respond to the presence of heterologous protein. The new protein fraction follows a Hill function where the heterologous protein is the inducer.

4. Conclusions

By utilizing the proteome fractionation framework initially proposed by Scott et al. and further developed by the same research group and others[11–13,15–17], we built a generalized kinetic model based on the fractionation of anabolic rate. The model defines a fraction of expression that theoretically accommodates any genetic regulatory circuit, thereby

providing a formal foundation for the analysis of biosensors and their response to inducers.

For the characterization of new regulatory circuits and the development of biosensors, it is nowadays common to rely on the expression of fluorescent proteins, which can be monitored in real time. Employing microcultures in plate readers greatly facilitates data collection from different circuits, strains, and culture conditions. However, the analysis of such data remains a complex challenge. Bacterial cultures inherently exhibit variability, necessitating the evaluation of replicates from different colonies. This makes it difficult to identify time points that can be accurately compared or averaged between them. Introducing the concept of the expression fraction, it is deduced that the peaks of accumulated protein, where $f_H = \varphi_H^{max}$, serve as reliable experimental time points for assessing the activity of a genetic sensor. Additionally, evaluating the reporter protein expression as a function of culture growth rate may reveal a linear range ($\rho_H \propto \mu$) in which the fraction of anabolic rate dedicated to the protein of interest (the expression fraction) remains constant. The slopes of these graphs also offer an alternative means of determining f_H, although its use for determining Hill's equation parameters may result in a broader error margin, being the product of two consecutive curve fittings.

Despite a common practice of selecting the time of maximum growth for data analysis, in our case -and probably in other experimental setups- the maximum growth rate falls within the interval of non-linear relation between growth and reporter protein production (figure 4 and 4S). Consequently, these time points are unsuitable for analysis. The production of the heterologous protein in early time points after induction may deviate from linearity because the extracellular inducers need more time to permeate cells and equilibrate with the regulatory machinery. Consistently with this interpretation, the presence of additional LacI repressor delays equilibrium, likely by reducing the expression of LacY transporter, a contributor to the uptake of IPTG[29]. Conversely, higher IPTG concentrations expedite equilibrium reducing the time required for the production rate

ρ_H to exhibit linear correlation with the growth rate μ.

Using our theoretical framework, we explored hypotheses explaining how heterologous protein expression exerts different metabolic burdens depending on the cellular context. Such differences are evidenced by the observed effects of GFP production on growth rate in B and K12 strains in the presence or absence of the pACYCDUET plasmid.

Other models take into account the specific regulations of the nutrient fixators and biosynthetic machinery[11,30], where differences among strains might explain particular growth rate phenotypes. In contrast, our model is a simplified approach that assumes that φ_R and φ_C possess directly opposite regulations without specifying their individual allocation cues such as cAMP and ppGpp. Therefore, our simulations could not explore how different sensitivities at those levels could represent different strains more accurately. Nevertheless, our simulations suggest that at least a portion of the induced burden is a result of the expression of additional proteins in response to heterologous protein production. This idea is partly supported by existing literature reporting expression profiles following heterologous protein induction[31]. Further experimentation may confirm whether a stress response sensitive to heterologous protein expression is a major contributor to the observed metabolic burden, and whether a strain-specific sensitivity of such response correlates to the distinct load that each strain bears.

Author information

Corresponding author

Daniel G. Guerra - Unidad de Moléculas Individuales, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima 15102, Perú. https://orcid.org/0000-0002-2410-722X, daniel.guerra@upch.pe, +51 319-0000 ext.233217.

Author contribution

Nicolás A. Vaccari -Conceptualized and developed the model, designed and cloned the IPTG biosensor system, conducted titration experiments, and performed data analysis. **Dahlin Zevallos-Aliaga** - Designed and cloned the biosensors for mercury and cumate, and performed their titration assays. **Tom Peeters** - Offered critical feedback on the design and testing of the mercury biosensor. **Daniel G. Guerra** - Served as the project supervisor, providing guidance and critical feedback for the experimental design, modeling, and data analysis.

Notes

The authors declare no competing financial interest.

Acknowledgement

This work was supported by Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica – FONDECYT / PROCIENCIA, through the program EF-041 - Proyectos de investigación básica y aplicada . grant number CONV-000114-2015-FONDECYT-DE.

The development of the mercury biosensor was funded by VLIR-UOS through the South Initiative grant code PE2020SIN292B122.

Supporting Information Available

Supporting Experimental Information: Additional experimental results showing the effects of IPTG induction with and without GFP production, the change of growth rates in time and fluorescence accumulation, growth vs. production graphs for BLR(DE3) and MG1655(DE3) strains, growth vs. production regressions for MerR-Hg^{2+}, and CymR-

cumate systems, total protein analysis by SDS-PAGE for all strains containing T7-GFP system, total protein quantitation for induced and non-induced bacteria. Supporting Modelling Information: Detailed definitions, step by step derivations and algebraic arrangements, and references for assigned constant values.

References

(1) Abatemarco, J.; Hill, A.; Alper, H. S. Expanding the metabolic engineering toolbox with directed evolution. *Biotechnology Journal* **2013**, *8*, 1397–1410, Number: 12.

(2) Hadadi, N.; Hatzimanikatis, V. Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Current Opinion in Chemical Biology* **2015**, *28*, 99–104.

(3) Hassanpour, N.; Ullah, E.; Yousofshahi, M.; Nair, N. U.; Hassoun, S. Selection Finder (SelFi): A computational metabolic engineering tool to enable directed evolution of enzymes. *Metabolic Engineering Communications* **2017**, *4*, 37–47.

(4) Küken, A.; Nikoloski, Z. Computational Approaches to Design and Test Plant Synthetic Metabolic Pathways. *Plant Physiology* **2019**, *179*, 894–906, Number: 3.

(5) Gong, Z.; Nielsen, J.; Zhou, Y. J. Engineering Robustness of Microbial Cell Factories. *Biotechnology Journal* **2017**, *12*, 1700014, Number: 10.

(6) Lee, G. N.; Na, J. The Impact of Synthetic Biology. *ACS Synthetic Biology* **2013**, *2*, 210–212, Number: 5.

(7) Jeon, Y.; Lee, Y.; Kim, K.; Jang, G.; Yoon, Y. Transcription Factor-Based Biosensors for Detecting Pathogens. *Biosensors* **2022**, *12*, 470.

(8) Mitchler, M. M.; Garcia, J. M.; Montero, N. E.; Williams, G. J. Transcription factor-

based biosensors: a molecular-guided approach for natural product engineering. *Current Opinion in Biotechnology* **2021**, *69*, 172–181.

(9) Ding, N.; Zhou, S.; Deng, Y. Transcription-Factor-based Biosensor Engineering for Applications in Synthetic Biology. *ACS Synthetic Biology* **2021**, *10*, 911–922, Publisher: American Chemical Society.

(10) Mahr, R.; Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Applied Microbiology and Biotechnology* **2016**, *100*, 79–90.

(11) Erickson, D. W.; Schink, S. J.; Patsalo, V.; Williamson, J. R.; Gerland, U.; Hwa, T. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature* **2017**, *551*, 119–123, Number: 7678.

(12) Mairet, F.; Gouzé, J.-L.; De Jong, H. Optimal proteome allocation and the temperature dependence of microbial growth laws. *npj Systems Biology and Applications* **2021**, *7*, 14, Number: 1.

(13) Mori, M.; Cheng, C.; Taylor, B. R.; Okano, H.; Hwa, T. Functional decomposition of metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions. *Nature Communications* **2023**, *14*, 4161, Number: 1.

(14) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science (New York, N.Y.)* **2010**, *330*, 1099–1102.

(15) Scott, M.; Hwa, T. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nature Reviews Microbiology* **2023**, *21*, 327–342, Number: 5.

(16) Sharma, P.; Pandey, P. P.; Jain, S. Modeling the cost and benefit of proteome regulation in a growing bacterial cell. *Physical Biology* **2018**, *15*, 046005, Number: 4.

(17) Wu, C.; Mori, M.; Abele, M.; Banaei-Esfahani, A.; Zhang, Z.; Okano, H.; Aebersold, R.; Ludwig, C.; Hwa, T. Enzyme expression kinetics by Escherichia coli during transition from rich to minimal media depends on proteome reserves. *Nature Microbiology* **2023**, *8*, 347–359, Number: 2.

(18) Towbin, B. D.; Korem, Y.; Bren, A.; Doron, S.; Sorek, R.; Alon, U. Optimality and sub-optimality in a bacterial growth law. *Nature Communications* **2017**, *8*, 14123.

(19) Tseng, H.-C.; Harwell, C. L.; Martin, C. H.; Prather, K. L. J. Biosynthesis of chiral 3-hydroxyvalerate from single propionate-unrelated carbon sources in metabolically engineered E. coli. *Microbial Cell Factories* **2010**, *9*, 96.

(20) Balleza, E.; Kim, J. M.; Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods* **2018**, *15*, 47–51.

(21) Andersen, J. B.; Sternberg, C.; Poulsen, L. K.; Bjørn, S. P.; Givskov, M.; Molin, S. New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *Applied and Environmental Microbiology* **1998**, *64*, 2240–2246.

(22) Studier, F. W.; Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **1986**, *189*, 113–130.

(23) Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **2006**, *72*, 211–222.

(24) Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; Chapman and Hall/CRC: New York, 2006.

(25) Dennis, P. P.; Bremer, H. Macromolecular composition during steady-state growth of Escherichia coli B-r. *Journal of Bacteriology* **1974**, *119*, 270–281.

(26) Carneiro, S.; Villas-Bôas, S. G.; Ferreira, E. C.; Rocha, I. Metabolic footprint analysis of recombinant Escherichia coli strains during fed-batch fermentations. *Mol. BioSyst.* **2011**, *7*, 899–910, Number: 3.

(27) Hoffmann, F.; Rinas, U. *Physiological Stress Responses in Bioprocesses*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; Vol. 89; pp 73–92, Series Title: Advances in Biochemical Engineering/Biotechnology.

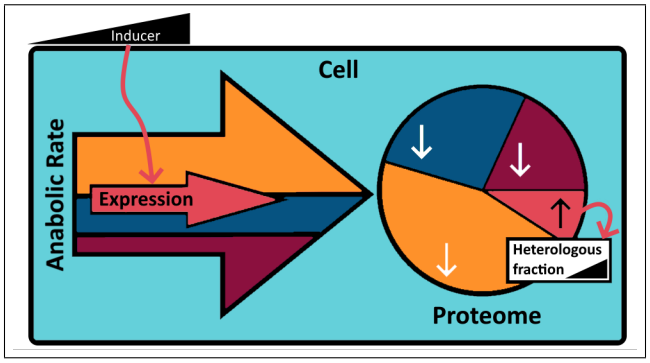
(28) Goffin, P.; Dehottay, P. Complete Genome Sequence of Escherichia coli BLR(DE3), a recA-Deficient Derivative of E. coli BL21(DE3). *Genome Announcements* **2017**, *5*, e00441–17.

(29) Hansen, L. H.; Knudsen, S.; Sørensen, S. J. The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens. *Current Microbiology* **1998**, *36*, 341–347.

(30) You, C.; Okano, H.; Hui, S.; Zhang, Z.; Kim, M.; Gunderson, C. W.; Wang, Y.-P.; Lenz, P.; Yan, D.; Hwa, T. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *500*, 301–306.

(31) Ceroni, F.; Boo, A.; Furini, S.; Gorochowski, T. E.; Borkowski, O.; Ladak, Y. N.; Awan, A. R.; Gilbert, C.; Stan, G.-B.; Ellis, T. Burden-driven feedback control of gene expression. *Nature Methods* **2018**, *15*, 387–393.

TOC Graphic


Inducer
Cell
Anabolic Rate
Expression
Heterologous fraction
Proteome