

A bottom-up characterization of transfer functions for synthetic biology designs: lessons from enzymology

Max Carbonell-Ballester^{1,2,†}, Salva Duran-Nebreda^{1,2,†}, Raúl Montañez^{1,2}, Ricard Solé^{1,2,3}, Javier Macía^{1,2,*} and Carlos Rodríguez-Caso^{1,2,*}

¹ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra, 08003 Barcelona, Spain, ²Institut de Biologia Evolutiva, CSIC-UPF, Psg. de la Barceloneta 37, 08003 Barcelona, Spain and ³Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Received July 28, 2014; Revised September 29, 2014; Accepted September 30, 2014

ABSTRACT

Within the field of synthetic biology, a rational design of genetic parts should include a causal understanding of their input-output responses—the so-called transfer function—and how to tune them. However, a commonly adopted strategy is to fit data to Hill-shaped curves without considering the underlying molecular mechanisms. Here we provide a novel mathematical formalization that allows prediction of the global behavior of a synthetic device by considering the actual information from the involved biological parts. This is achieved by adopting an enzymology-like framework, where transfer functions are described in terms of their input affinity constant and maximal response. As a proof of concept, we characterize a set of Lux homoserine-lactone-inducible genetic devices with different levels of Lux receptor and signal molecule. Our model fits the experimental results and predicts the impact of the receptor's ribosome-binding site strength, as a tunable parameter that affects gene expression. The evolutionary implications are outlined.

INTRODUCTION

The advance of genetic engineering has made it possible to modify genetic programs inside cells by re-designing them in predefined ways (1). Synthetic biology has emerged as a discipline in which modular biological parts are used for the construction of genetic devices. As in any engineering discipline, mathematical and computational models provide the workbench to infer system-level behavior from the properties of the biological parts (2). Standard engineering predicts output responses of a device given a set of input signals and a specified internal set of pieces. Within synthetic biology, the proper characterization of simple blocks in a

reliable way constitutes a major challenge for the building of complex genetic devices (3–5).

The transfer function, a term borrowed from electronics, is the representation of the relationship between the input and the output of a system (6,7). This concept has been translated within synthetic biology as the response of a regulable genetic device in the presence of a signal that acts as the control variable of the system. In most relevant scenarios, nonlinear responses are often desirable in order to implement the digital logic abstraction found in man-made circuits. This can be achieved using mechanisms such as saturation of biochemical systems (8), ultrasensitivity (9), multistability (10) and transcription factor cascades (11) among others. Hill functions have been commonly used for the fitting of experimental datasets in biochemistry (12), computational biology (13), pharmacology (14), systems and synthetic biology (10,15–18). The success of this approach comes from the fact that fitting data require little *a priori* knowledge of the underlying biological mechanisms, and provide quantitative information about affinity and cooperativity of the system (8).

In genetics, Hill-like functions come from the assumption of cooperative effects due to transcription factor multimerization (19) and can be derived from equilibrium calculations on ligand-receptor binding. However, in most cases, its representation results from the correction of the hyperbolic Michaelis–Menten approach by adding an empirical exponent n (14), written as

$$\frac{v}{V_m} = \frac{[S]^n}{S_{0.5}^n + [S]^n}. \quad (1)$$

As a consequence of its empirical nature, neither the original Michaelis–Menten premises nor biological information remains in the model, losing the link between the kinetic parameters and biological mechanisms. Accordingly, models constructed by fitting have very limited predictive value

*To whom correspondence should be addressed. Tel: +34933160532; Fax: +34933160550; Email: carlos.rodriquez@upf.edu
Correspondence may also be addressed to Javier Macía. Tel: +34933160532; Fax: +34933160550; Email: javier.macia@upf.edu

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

beyond the exact conditions in which data were acquired. Thus, the approximation taken is largely a heuristic one.

Design often requires iterative optimization steps. However, any device modification may lead to some type of unpredictable behavior, forcing further empirical characterization. Unfortunately, such a circumstance is not rare in the process of construction and testing of a genetic device (4). Hence, there is a need for a more suitable framework that allows predictions and avoids time-consuming data collection.

In this regard, the Michaelis–Menten approach (20) may offer an inspiring alternative to the broadly accepted Hill fitting. Interestingly, the transfer function concept fairly matches the substrate-velocity plot for enzymatic catalysis. This classical plot constitutes a clever characterization of enzyme kinetics, connecting a simple experimental setup with a biochemically grounded model, based on very precise premises. In that way, an analogous perspective for genetic devices would confer to transfer functions a desirable predictive value. The aim of this work is to establish a quantitative relation between input's affinity, signal amplitude and the variation of the control variable (i.e. induction molecules). In order to provide an experimental validation, we shall compare our model predictions with the characterization of an engineered device: the Lux system.

The quorum sensing Lux system has been extensively used in synthetic biology (15,16,21). With a sophisticated regulation in nature (22), its engineered versions have been restricted to the transcriptional level, to which a Hill-like behavior with a wide range of cooperativities has been reported (7,13,16,23).

When we look at the biochemical characterization, the interaction of LuxR dimer with 3-oxo- C_6 -homoserine lactone (3OC6HSL) induces the binding to promoter (24). This process is mediated in a noncooperative manner by 3OC6HSL, as suggested by studies in Lux and its Car homolog systems (25,26). Interestingly, receptor without lactone cognate is able to bind the DNA promoter (27), suggesting that some expression mediated by free receptor may occur. This scenario, schematically represented in Figure 1A, provides a starting point for a more biologically meaningful model of this system. However, one issue remains: the ability to control and manipulate the elements of the device.

From an engineering perspective, modularity and orthogonal function of genetic parts is the key for the construction of tailored devices. At this point, the ribosome-binding sites (RBSs) are useful elements to control the efficiency of the translation of the mRNA pool. Efforts on the characterization of RBSs variants for different organisms have provided valuable information for the choice of one or another RBS in a genetic system (28). A comparison of the effect of these parts in the expression of the final output is given by its relative strength, which is calculated using a standard value of expression as a reference for normalization (29). The use of different RBSs constitutes a common way to modulate the expression of a particular gene. But what is the impact of RBS changes on the behavior of a device?

To tackle this question, our work presents an enzymology-like approximation that allows us to explore the role of different RBSs in a genetic pLux-LuxR-inducible circuit. This study shows how tunable parts of the device,

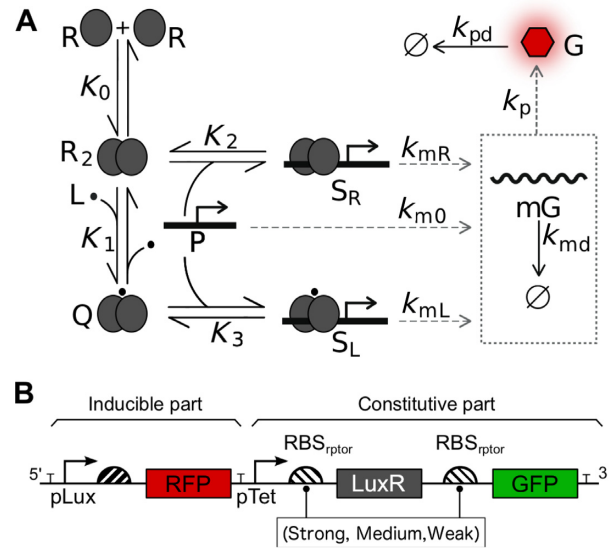


Figure 1. Schematic representation of genetic regulation for the inducible LuxR-pLux engineered devices used in this study. Notation follows the mathematical model: R LuxR receptor, L lactone, and P pLux promoter, R_2 dimerized receptor without lactone, Q dimerized receptor with lactone, S_R transcriptional complex not mediated by lactone, S_L lactone mediated transcriptional complex, mG mRNA of reporter gene and G reporter protein. The K_i notation represents equilibrium constants, while k_i refers to kinetic constants (A). Genetic architecture of the three constructs analyzed (B).

such as the expression of receptor, modulate the transfer function in a completely predictable manner. As a proof of concept, an experimental characterization and a further mathematical modeling of an inducible genetic device is presented. This picture more close to the biological mechanism suggests some limitations of the conventional Hill fitting approach.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Cloning and expression experiments were performed in *Escherichia coli* Top10 (Invitrogen, USA). Cells were grown in Lysogeny Broth (LB) at 37°C and selected with appropriate antibiotics (chloramphenicol 340 µg/ml; kanamycin 250 µg/ml; or ampicillin 100 µg/ml; Sigma, USA). Bacterial strains were preserved in LB glycerol 20% (v/v) at -80°C.

Construction of 3OC6HSL Lux genetic devices

Cloning was carried out using the Biobrick assembly method and the parts from the Spring 2010 iGEM distribution. The biobrick parts used in this study were the following: *B0014* (double terminator), *B0034* (strong RBS), *B0032* (medium RBS), *B0033* (weak RBS), *R0040* (tetracycline promoter, pTet, as a constitutive promoter), *R0062* (Lux promoter, pLux), *C0062* (LuxR coding sequence), *E1010* (red fluorescence protein, RFP), *E0040* (green fluorescence protein, GFP). Biobrick cloning was performed using an assembly kit (Ginkgo Bioworks, USA).

Three genetic devices were built, composed of a common 3OC6HSL inducible part followed by three

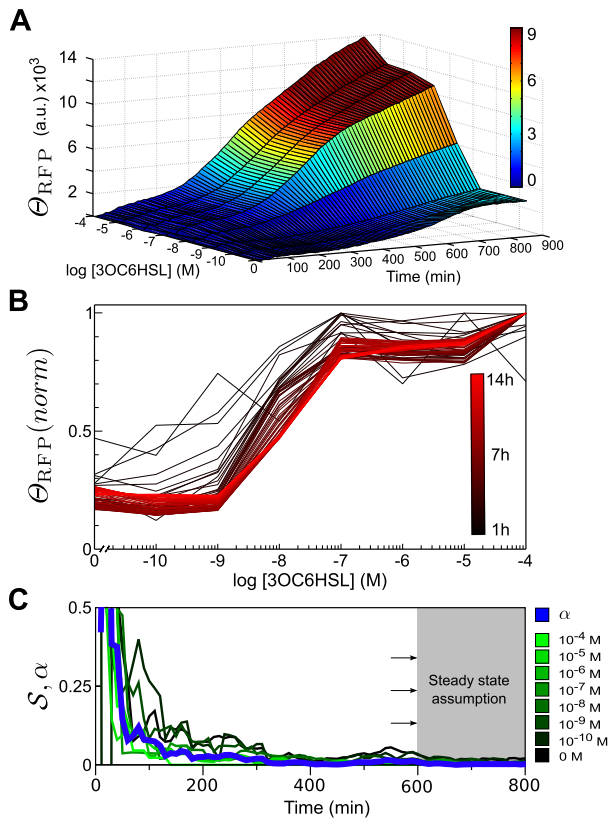


Figure 2. Transfer function acquisition RBS_{rptor} (medium) construct. A 3D chart showing response measurements of RFP expression along time at different [3OC6HSL] (A). Time series of normalized RFP expression. Colors (from black to red) indicate different times in the experiment, from 1 to 14 h (B). Estimation of the time regime for transfer function acquisition (C). In a green scale, the value of S for the eight different [3OC6HSL]. In blue, α value accounting for the overall signal variation along time. The shaded area shows the time region in which we assume the steady state has been reached.

variants of a constitutive part, as illustrated in Figure 1B. On one hand, the inducible construct had the following structure: B0014-R0062-B0032-E1010; on the other hand, there were three variants of the constitutive part, cloned as B0014-R0040-X-C0062-X-E0040-B0014, where X corresponds to the three aforementioned alternative variants: B0034, B0032 and B0033 (labeled as *strong*, *medium* and *weak*, respectively). Technical details about the relative strength of these RBSs are taken from http://parts.igem.org/Ribosome_Binding_Sites/Prokaryotic/Constitutive/Community_Collection. As detailed in the part registry, the B0034 part was used as a standard for normalization. Therefore the relative strengths used in this work were 1, 0.3 and 0.03 for B0034, B0032 and B0033, respectively. All constructs were included in the BioBricks high copy number plasmid (pSB1AK3) and transformed by chemical method. In the case of the experiments with no receptor (see Figures 3C and 5B), a construct bearing only the inducible part within the same plasmid was used. All genetic constructs were confirmed by Sanger sequencing.

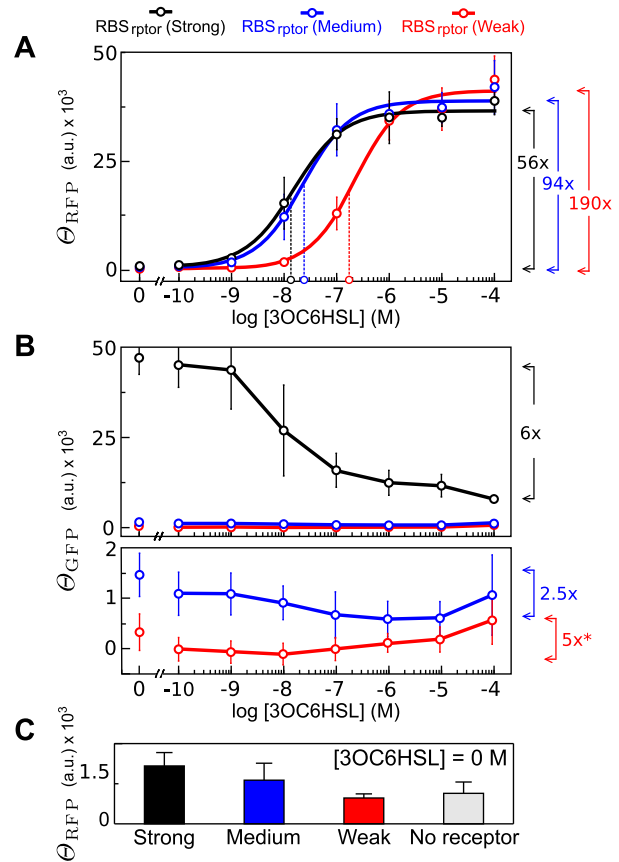


Figure 3. Transfer function varying RBS at receptor. Open circles in the X-axis indicate the $K_{0.5}^n$ according to the fitting. Arrows in the right side of the chart show the induction with respect to basality according to fitted equations (A). Effect of [3OC6HSL] on Θ_{GFP} , as an estimator of receptor expression. The bottom chart is a zoom of medium and weak constructs. Arrows in the right side of the charts indicate the variation between minimum and maximum value of Θ_{GFP} . Asterisk indicates that Θ_{GFP} was very close to zero giving rise to negative values. In this case, variation was calculated using the lowest positive value (B). Leakiness of constructs at [3OC6HSL] = 0 M (C).

Fluorescence assays for gene expression determination

Strains containing the plasmid of interest were grown overnight in LB ampicillin at 37°C and continuous shaking. A 1000-fold dilution from overnight culture was grown until exponential phase, $OD_{660} \approx 0.4$. Cultures were centrifuged at 4000 g, during 5 min and resuspended in fresh LB ampicillin up to an $OD_{660} = 0.3$. Incubation for *in vivo* measures was carried out by transferring 100 μ l of the diluted cultures and 100 μ l of LB ampicillin with the appropriate 3OC6HSL (N-[β -ketocaproyl]-L-homoserine lactone; Cayman Chemical Company, USA) concentrations into a flat bottom 96-well microplate (Nunc, Thermo Fisher Scientific, USA). LB without cells was included in the incubation as a background control for both fluorescence and absorbance.

Gene expression was monitored in time for a battery of 3OC6HSL concentrations by quantification of the RFP. LuxR was indirectly reported by measuring the concomitant expression of GFP placed in tandem with LuxR (Figure 1B). Incubation and measures of bacterial cultures dur-

Table 1. Data fitting for the parameters using Hill function approach and the enzymology-like model assuming leakiness

RBS _{rptor}	a	$b (\times 10^4)$	$K_{0.5} (\times 10^{-8})$ M	n	r^2	$a_1 (\times 10^4)$	$a_2 (\times 10^{-10})$ M	$a_3 (\times 10^{-8})$ M	r^2
Strong	638 ± 1990	3.6 ± 0.3	1.5 ± 8.9	0.93 ± 0.16	0.96	3.6 ± 0.1	3.2	1.5 ± 0.7	0.95
Medium	404 ± 2103	3.8 ± 0.3	2.3 ± 4.8	0.97 ± 0.13	0.95	4.0 ± 0.2	4.0	2.6 ± 1.1	0.95
Weak	215 ± 1605	4.1 ± 0.3	22 ± 59	0.99 ± 0.23	0.96	4.1 ± 0.2	12.7	22.0 ± 6.8	0.96

ing characterization were performed on a Synergy MX microplate reader (BioTek Instruments, USA) every 10 min for 14 h. Fluorescence measures for RFP (ex: 578 ± 9 nm, em: 616 ± 9 nm) and GFP (ex: 478 ± 9 nm, em: 516 ± 9 nm) with gain 70 were carried out, as well as optical density (OD at 660 nm) measures. Incubation was done at 37°C with continuous orbital shaking (medium intensity). 3OC6HSL concentration conditions were prepared from an initial stock at 10⁻² M (3:1, phosphate buffered saline:ethanol). Serial dilutions in LB ampicillin ranging from 10⁻⁴ to 10⁻¹⁰ M were prepared the day of the experiment.

Data transformation and Hill function fitting

Sample absorbance and fluorescence readings (OD₆₆₀(S), $f(S)$) were corrected using signal background control (OD₆₆₀(B), $f(B)$). Averaged data were obtained from six independent experiments. As described in (18), output signal Θ_i was calculated according to the formula:

$$\Theta_i = \frac{f_i(S) - f_i(B)}{OD_{660}(S) - OD_{660}(B)} \quad (2)$$

where i refers to GFP or RFP. The value Θ corresponds, with a factor of proportionality, to the concentration of the fluorescent protein i per cell. Matlab R2013a software was used for fitting according to the following formula, assuming a cooperative behavior and signal basality:

$$\Theta_{RFP} = a + \frac{b[3OC6HSL]^n}{K_{0.5}^n + [3OC6HSL]^n} \quad (3)$$

Nonlinear least squares were computed using the trust region algorithm with default settings.

Time series and signal variation computation

A normalized value of Θ_{RFP} , $\Theta_{RFP}(norm)$, was calculated for every time step as following: given a time, Θ_{RFP} for each [3OC6HSL] was divided by the maximal value. We defined signal variation, \mathcal{S} , as the coefficient of variation of $\Theta_{RFP}(norm)$ in a time interval. Evolution of \mathcal{S} over time was calculated using a moving window, consisting of five points of $\Theta_{RFP}(norm)$, therefore capturing the information of 50 min in total.

The relation between the values of \mathcal{S} for the different [3OC6HSL] was evaluated using the estimator α . This value, defined as the standard deviation of the different \mathcal{S} curves, was used to establish the region of the steady-state condition for transfer function acquisition. According to α analysis, the time for the transfer function acquisition was set at 14 h, applying a gain 75 for fluorescence measures.

RESULTS

Transfer function acquisition

The exact time for transfer function acquisition is often arbitrary and still constitutes an open issue (7,30). As an illustrative case, Figure 2A shows the evolution of Θ_{RFP} along time for different [3OC6HSL] for the construct with the RBS_{rptor} (medium). Similar qualitative behavior was observed in the other two constructs (data not shown). Representation of $\Theta_{RFP}(norm)$ of the different [3OC6HSL] curves allowed a qualitative comparison of the transfer function along time. Figure 2B shows how the transfer function converged (from black to red lines) to a more and more signal overlap.

After the introduction of 3OC6HSL, cells require time for protein production and maturation. At the steady state, i.e. $\Delta\Theta/\Delta t \sim 0$, production is compensated by degradation processes, giving rise to a constant value of Θ along time. However, given any arbitrary interval time Δt , $\Delta\Theta$ consists of the genetic behavior due to the actual change associated to protein production and the noise associated to the measure in that interval. This imposes a limitation for the application of the steady state definition.

To overcome this limitation, we used \mathcal{S} as a way to establish a practical definition for the steady-state condition from experimental data (see Materials and Methods section for definition). As at the steady state biological signal does not contribute to the variation of the output, this one must be given only by noise. In Figure 2C, where the time evolution of \mathcal{S} for every [3OC6HSL] (labeled from black to green) is shown, one can see how dispersion values depended on time but also on [3OC6HSL] (inversely proportional). The differences among \mathcal{S} curves were captured in the time evolution of α value (see Materials and Methods section), represented by a blue thick line in Figure 2C. Looking at α , we could arbitrarily select a threshold to define the time when the steady steady is assumed for the transfer function acquisition. This mathematical transformation allowed us to usefully collapse the information about the level of fluctuations in a single curve. In our study, we chose a final point at 14 h to perform the transfer function characterization and further modeling fitting.

Characterization of the pLux device varying RBS_{rptor}

In order to modulate the strength of gene expression, we characterized a set of constructions using different RBS_{rptor} as illustrated in Figure 1B. The characterization of the respective transfer functions is summarized in Figure 3. The amount of output in response to [3OC6HSL] suggested a noncooperative effect, i.e. $n \approx 1$ (see Figure 3A and Table 1 for numerical details). Furthermore, the results showed a decrease of the turning point when stronger RBS_{rptor} were used.

The expression of receptor monitored by fluorescence measurements of the GFP tandem construction (Figure 1B) showed how the receptor expression was practically unaffected in RBS_{rptor} (medium) and RBS_{rptor} (weak) constructs. However, a significant decrease of Θ_{GFP} was observed in the case of the RBS_{rptor} (strong) construct.

Measures with zero ligand concentration produced a basal signal. Such a response was also dependent on the strength of the receptor, as shown in Figure 3C. The analysis of constructs without receptor also produced a basal expression, independently of lactone concentration. This suggests that the engineered pLux alone was able to promote the expression without receptor.

Enzymology-based model premises

The mathematical model for the inducible device was established according to the following premises:

- Transfer function is evaluated at the steady state where a single stable fixed point is expected.
- Concentrations of molecular species regarding the genetic device, i.e. receptor, promoters and transcriptional and translational machinery, remain invariable along time.
- Total promoter concentration $[P_T]$ is much smaller than the total transcriptional activator receptor $[R_T]$.
- Ligand concentration $[L]$ (the input) is assumed invariant in time and, according to the previous premise, large enough to consider that the bound part is negligible with respect to the total one, i.e. $[L] \approx [L_T]$.
- Genetic devices work under the condition in which its load neither affects the behavior of the cell nor causes a limitation of resources. Metabolic burden is therefore not included in the formalization.
- The parts of the genetics device have an orthogonal functionality, i.e. the components perform one predefined function and they have not got undesirable interactions with the other cellular components. Therefore, they perform the expected behavior.

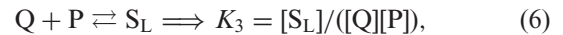
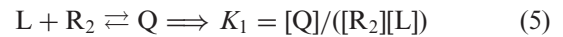
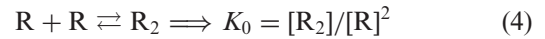
Besides these general assumptions, for our specific LuxR device and according to the results shown in the previous section, we shall consider the biological mechanism shown in Figure 1A. In agreement to this, RFP expression is modulated by lactone concentration and both luxR-independent and luxR-dependent leakiness are also considered. In order to provide a more understandable view of this characterization, we formulate a mathematical model using the former set of enzymology-inspired premises.

A simple model predicts the role of the RBS_{rptor} strength in output amplitude and turning point

Departing from the genetic circuit illustrated in Figure 1A, our model considers a process of sensing a molecule L by joining with a dimeric receptor R_2 . It is worth mentioning that the requirement of more than one L molecule would introduce a power over $[L]$. For the sake of simplicity, we shall assume that one ligand molecule is enough for receptor's activation and further addition has no impact on the behavior

of active receptor complex. This assumption is supported by the noncooperative process observed in our experimental characterization and by the reported evidence previously mentioned in the Introduction section.

In this simplified version, we assume that leakiness due to receptor binding and naked promoter does not occur, i.e. $k_{m0} = 0$ and $K_2 = 0$. The reactions considered in this simple scenario are:



giving rise to the following differential equations:

$$\frac{d[mG]}{dt} = k_{mL}[S_L] - k_{md}[mG] \quad (7)$$

$$\frac{d[G]}{dt} = k_p[mG] - k_{pd}[G]. \quad (8)$$

At the steady state, $d[mG]/dt = d[G]/dt = 0$, $[G^*]$ can be written as a function of the active complex $[S_L]$:

$$[G^*] = \frac{k_p k_{mL}}{k_{pd} k_{md}} [S_L]. \quad (9)$$

Now we need to obtain a mathematical expression of $[G^*]$ in terms of the control variables of this model, i.e. $[L]$, $[R_T]$ and $[P_T]$. These correspond to [3OC6HSL] concentration, externally fixed in the experimental setup; amount of LuxR, kept invariant by its constitutive expression ($[R_T]$); and the number of Lux promoters ($[P_T]$), defined by a constant population of plasmids.

In this simple model, $[R_T]$ is a function of $[R_{2T}]$, as detailed in the mathematical appendix. For the sake of simplicity and according to the biology of the system, we assume that the process of dimerization is prior to the binding with the ligand. Furthermore, we mathematically impose that the ligand interaction has a negligible effect on the equilibrium between dimer and monomer. According to this, we define two conservation equations:

$$[R_{2T}] = [R_2] + [Q] + [S_L] \quad (10)$$

$$[P_T] = [P] + [Q] + [S_L], \quad (11)$$

where by the assumption of $[R_T] \gg [P_T]$, $[S_L]$ is not considered in (10). Applying the equilibrium constant definitions (5) and (6), we rewrite the $[S_L]$ expression and the conservation Equations (10) and (11) as

$$[S_L] = K_1 K_3 [P][R_2][L] \quad (12)$$

$$[R_{2T}] = [R_2](1 + K_1[L]) \quad (13)$$

$$[P_T] = [P](1 + K_1 K_3 R_2 [L]). \quad (14)$$

According to these equations, we can rewrite $[S_L]$ in function $[R_{2T}]$ and $[P_T]$ in Equation (9), giving rise to the expres-

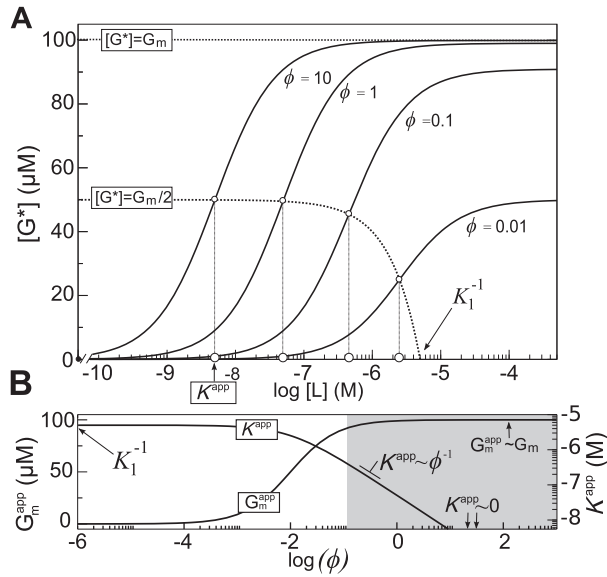


Figure 4. Transfer function for four different ϕ according to the simplified model without leakiness. Upper dotted line corresponds to the maximum achievable expression. Lower dotted line shows the $G_m^{\text{app}}/2$ as function of ϕ . The intersections with the transfer functions (little open circles) highlight the $G_m^{\text{app}}/2$ for the four transfer functions. Their projections over the $[L]$ axis show their respective K^{app} values (bigger open circles) (A). Dependence of K^{app} and G_m^{app} with respect to the value of ϕ . Note that when $\phi \sim 0$, $K^{\text{app}} \sim K_1^{-1}$ and $G_m^{\text{app}} \sim 0$. The gray area corresponds to $\phi \gg (K_3 r')^{-1}$. In this region G_m^{app} acquires the maximum value and K^{app} decays as the inverse of ϕ (B). Parameters for the simulation were: $K_1 = 0.2 \mu\text{M}^{-1}$, $K_3 = 1 \mu\text{M}^{-1}$, $r' = 100 \mu\text{M}$ and $\gamma g k_{\text{mL}}[\text{P}_T] = 100 \mu\text{M}$.

sion

$$[G^*] = \frac{\gamma g k_{\text{mL}}[\text{P}_T][\text{R}_{2\text{T}}]}{K_3^{-1} + [\text{R}_{2\text{T}}]} \left(\frac{[\text{L}]}{\left(\frac{K_1^{-1}}{1 + K_3[\text{R}_{2\text{T}}]} \right) + [\text{L}]} \right), \quad (15)$$

where γ and g are groups of kinetics constants as detailed in Equations (35) and (36) of the mathematical appendix. $[\text{R}_{2\text{T}}]$ can also be written as a function of the normalized RBS strength (ϕ), with respect to the canonical RBS part (see the mathematical appendix for further details). Now, by defining

$$G_m^{\text{app}} = \gamma g k_{\text{mL}}[\text{P}_T] \frac{\phi}{(K_3 r')^{-1} + \phi} \quad (16)$$

$$K^{\text{app}} = K_1^{-1} \frac{(K_3 r')^{-1}}{(K_3 r')^{-1} + \phi}, \quad (17)$$

we obtain a more compact expression in a familiar Michaelis–Menten form:

$$[G^*] = G_m^{\text{app}} \frac{[\text{L}]}{K^{\text{app}} + [\text{L}]}. \quad (18)$$

Interestingly, the model allows us to define the maximum expression of the reporter (G_m^{app}) and the affinity constant of the device (K^{app}) as functions of the normalized strength of the RBS receptor (ϕ). According to the expressions (16), (17) and (18), Figure 4 illustrates the Michaelian effect of $[L]$ and ϕ in the device. Notice that G_m^{app} reaches its maximum

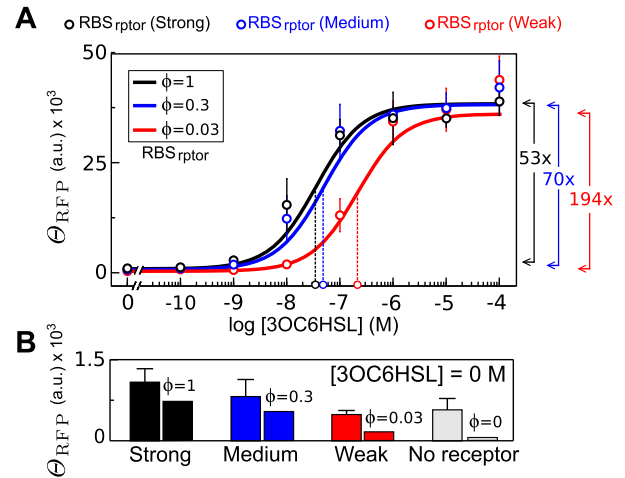


Figure 5. Transfer function (A) and basal expression (B) obtained from fitting our mathematical model including leakiness with real data. It explains the effect of receptor in amplitude signal and ligand affinity. Arrows show the times of variation from the ratio of G_m/G_0 from fitted model equation. Model was manually fitted to data using the following parameter values: $\gamma k_{\text{m0}}[\text{P}_T] = 85 \text{ M}$, $K_2 r' = 5$, $K_3 r' = 500$, $\varepsilon_R = 10$, $\varepsilon_L = 10$ and $K_1^{-1} = 3 \times 10^{-6} \text{ M}$. Pearson test was applied to the different constructs giving the following correlation coefficients: 0.996, 0.987 and 0.987 for weak, medium and strong, respectively. Comparison of model and real data leakiness in the panel (B) gave a correlation coefficient of 0.984.

value when $\phi \gg (K_3 r')^{-1}$ is satisfied (gray region in the Figure 4B). Under such a condition, the affinity of the device increases, reducing the value of K^{app} as ϕ^{-1} (see Figure 4B).

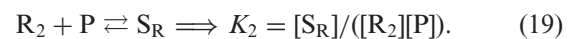
In this model, G_m^{app} ranges from zero at $\phi = 0$ to a maximal value, determined by the efficiency of the machinery, formally $\gamma g k_{\text{mL}}[\text{P}_T]$. Conversely, K^{app} at $\phi = 0$ acquires its maximum value K_1^{-1} , corresponding with $G_m^{\text{app}} = 0$.

It is worth mentioning that extreme cases address an ideal trend and they would not match real behavior. This is due to considerations of the simple model, such as the availability of ligand and resources, as well as the extreme variation of the ϕ values, which are far to be satisfied under physiological conditions. Unlike Hill approximation, this simple equation allows us to see in a qualitative way the role of this device according to the variation of its components.

It is worth to note the similarity of equation (18) with the so-called reversible acompetitive inhibition in enzymology (8). Input affinity and signal amplitude is affected in a saturable way by the RBS strength as it occurs with the inhibitor in the enzymological model. However, in our case such a dependence occurs in an activatory fashion.

Mathematical model including leakiness predicts the effect of receptor on output basality

The model presented in this section is a generalization of the previous one. In contrast to the simplified version, this model incorporates two lactone-independent expression pathways: one mediated by the free receptor and another by the naked promoter. Departing from the chemical reactions previously presented in Equations (4), (5) and (6), we now add the following binding equilibrium:



The concentration of the protein reporter [G] is determined by the equation based on the intermediates species—including the free promoter activity mediated by k_{m0} —as described in the following ODEs:

$$\frac{d[mG]}{dt} = k_{m0}[P] + k_{mR}[S_R] + k_{mL}[S_L] - k_{md}[mG] \quad (20)$$

$$\frac{d[G]}{dt} = k_p[mG] - k_{pd}[G]. \quad (21)$$

Analogously to the simple model, we shall write an expression of $[G^*]$ based on equilibrium constants and total concentration of the chemical species. As detailed in the mathematical appendix, the resulting expression in its compact version is:

$$[G^*] = a_1 \left(\frac{a_2 + L}{a_3 + L} \right), \quad (22)$$

where a_1 , a_2 and a_3 are functions of ϕ (detailed in Equations (53), (54) and (55) of the mathematical appendix), accounting for amplitude signal, basalness and ligand affinity, respectively.

By considering the two limit cases $[L] = 0$ and $[L] = \infty$ of the (22) and defining the turning point as that $[L] = L_{0.5}$ for $G_m/2$, we connect mathematical parameters with measurable values of the transfer function as follows:

$$G_0 = \frac{a_1 a_2}{a_3} = \gamma g P_T \frac{k_{m0} + k_{mR} K_2 r' \phi}{1 + K_2 r' \phi} \quad (23)$$

$$G_m = a_1 = \gamma g P_T \frac{k_{m0} + k_{mL} K_3 r' \phi}{1 + K_3 r' \phi} \quad (24)$$

$$L_{0.5} = a_3 - 2a_2. \quad (25)$$

The parameters G_0 , G_m and $L_{0.5}$ correspond to the basal expression, maximal expression and turning point, respectively. Continuous lines in Figure 5A show that the effect of ϕ is similar to the observed one from the simple model behavior for signal amplitude (G_m) and turning point ($L_{0.5}$) (see Figure 4A for comparison). Leakiness by free receptor gives rise to a basal signal at $[L] = 0$ that also depends on ϕ in a saturable way, according to Equation (23).

This expression, although less treatable than the simpler version, offers a predictable behavior useful for data fitting. Figure 5 shows the fitting of experimental data using values of normalized strength for the strong ($\phi = 1$), medium ($\phi = 0.3$) and weak ($\phi = 0.03$) RBS parts, as are described in the parts registry collection (see Materials and Methods). The fitting using the Equation (22) gave a basal output and turning point that fairly matched the experimental values (see Table 1). In the same way, amplitudes from the model and the experiments followed a similar trend as suggested by Pearson correlation between data and the model, as detailed in Figure 5.

DISCUSSION

Despite the potential applicability of Synthetic Biology in biomedicine and environmental issues, a proper characterization of global behavior of devices in a reliable and predictable way is missing. In this regard, the use of biological

meaningful models allows us to determine the behavior of tunable genetic devices.

Inspired by well established formal approaches from enzymology, our results show that the impact of tunable parts such as the RBS receptor, the plasmid copy number (which would modulate $[P_T]$) and even the RBS directly affecting RPF, can be predictably studied without the use of the Hill function approach. From an engineering perspective, the model reveals that the sensation by an inducible device can be adjusted by varying the receptor levels. A strategy based on the changes of the receptor expression is worth considering rather than complicated protein engineering.

Interestingly, the model offers a suitable explanation about the lack of cooperativity observed in the experimental results: only one molecule of lactone is required for the activation complex. This is in agreement with evidence found in literature for Lux and homologous quorum sensing receptors. The free ligand receptor formation allows us to explain the leakiness mediated by free receptor described in this work. Such evidence provides some concerns about the behavior of the engineered Lux device, specially when positive cooperativity is desired (13,15).

Noteworthy, for the RBS_{rptor} (*strong*) we observed a reduction in the receptor levels, despite its expression took place under a constitutive promoter. This reduction may affect the device response when scaling up the system. These experimental results may suggest the existence of an indirect negative interaction between inducible gene expression, in our case RFP, and other nonregulated genes, such as LuxR. We speculate that the metabolic load associated to device induction could be responsible for this behavior. Looking at Figure 5B, the higher levels of basal RFP expression in constructs with and without LuxR in absence of lactone can be interpreted in similar terms. This latter example illustrates the possibility that experimental setup may not adjust well to our model premises, being the nonorthogonality of genetic devices another possible source of discrepancies. The predictive value of the model precisely relies on the assumption that its premises are fulfilled, and therefore deviations from the expected behavior could be a hint that some premises are broken. Further research incorporating factors such as metabolic load and crosstalks might improve the predictive capabilities of the model.

Bearing all this in mind, principles of organization described in this work may offer an evolutionary insight, in particular, in processes of adaptation or even the emergence of some type of diseases. According to our results, mutations at the level of receptor expression may offer a finer tuning process than those occurring at the polypeptidic chain of the receptor. In this context, the work provides a proof of concept for an interesting evolutionary perspective of the principles of biological design extracted from a synthetic biology approximation.

CONCLUSION

The use of an enzymology-based approach provides a framework for the study and reliable characterization of synthetic devices uncovering interesting connections of the principles of organization of natural systems. Further work on the extension of an enzymological approach to the study

of more complex genetic behaviors would be of great interest for the controllability and development of new synthetic genetic devices.

ACKNOWLEDGMENTS

We thank Eva G. Ramallo for technical support and Complex Systems lab members for fruitful conversations.

FUNDING

Fundación Botín, Banco de Santander through its Santander Universities Global Division [BES-2010-038940 to R.M., C.R.C.]; ERC SYNCOM [291294 to M.C.B.]; FPI MINECO fellowship [to S.D.N.]. Funding for open access charge: ERC SYNCOM [291294].

Conflict of interest statement. None declared.

REFERENCES

- Endy, D. (2005) Foundations for engineering biology. *Nature*, **438**, 449–453.
- Tyson, J.J., Chen, K.C. and Novak, B. (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.*, **15**, 221–231.
- Baldwin, G., Bayer, T., Dickinson, R., Ellis, T.P.S., Freemont, R.I.K., Polizzi, K. and Stan, G.B. (2012) Synthetic Biology: A Primer. 1st edn. Imperial College Press, London.
- Mutalik, V.K., Guimaraes, J.C., Cambray, G., Mai, Q.A., Christoffersen, M.J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G. *et al.* (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods*, **10**, 347–353.
- Kwok, R. (2010) Five hard truths for synthetic biology. *Nature*, **463**, 288–290.
- Mukherji, S. and van Oudenaarden, A. (2009) Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.*, **10**, 859–871.
- Wang, B., Kitney, R.I., Joly, N. and Buck, M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.*, **2**, 508.
- Cornish-Bowden, A. (2004) Fundamentals of Enzyme Kinetics. 3rd edn. Portland Press, London.
- Trunnell, N., Poon, A., Kim, S. and Ferrell, J.J. (2011) Ultrasensitivity in the regulation of Cdc25C by Cdk1. *Mol. Cell*, **41**, 263–274.
- Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339–342.
- Hoochangi, S., Thiberge, S. and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. U.S.A.*, **102**, 3581–3586.
- Huang, C.Y. and Ferrell, J. Jr (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. U.S.A.*, **93**, 10078–10083.
- Weber, M. and Buceta, J. (2013) Dynamics of the quorum sensing switch: stochastic and non-stationary effects. *BMC Syst. Biol.*, **7**, 6.
- Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M. and Maire, P. (2008) The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam. Clin. Pharmacol.*, **22**, 633–648.
- Garcia-Ojalvo, J., Elowitz, M.B. and Strogatz, S.H. (2004) Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 10955–10960.
- Balagadde, F.K., Song, H., Ozaki, J., Collins, C.H., Barnett, M., Arnold, F.H., Quake, S.R. and You, L. (2008) A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.*, **4**, 187.
- Swinburne, I.A., Miguez, D.G., Landgraf, D. and Silver, P.A. (2008) Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev.*, **22**, 2342–2346.
- Chappell, J., Jensen, K. and Freemont, P.S. (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.*, **41**, 3471–3481.
- Aron, U. (2006) An Introduction to Systems Biology: Design Principles of Biological Circuits. 1st edn. Chapman and Hall/CRC, Boca Raton, FL.
- Michaelis, L. and Menten, M.L. (1913) Die Kinetik der Invertinwirkung. *Biochem. Z.*, **49**, 333–369.
- Williams, J.W., Cui, X., Levchenko, A. and Stevens, A.M. (2008) Robust and sensitive control of a quorum-sensing circuit by two interlocked feedback loops. *Mol. Syst. Biol.*, **4**, 234.
- Tu, K.C., Long, T., Svenningsen, S.L., Wingreen, N.S. and Bassler, B.L. (2010) Negative feedback loops involving small regulatory RNAs precisely control the *Vibrio harveyi* quorum-sensing response. *Mol. Cell*, **37**, 567–579.
- Zucca, S., Pasotti, L., Mazzini, G., De Angelis, M.G.C. and Magni, P. (2012) Characterization of an inducible promoter in different DNA copy number conditions. *BMC Bioinformatics*, **13**(Suppl. 4), S11.
- Fuqua, C., Winans, S.C. and Greenberg, E.P. (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu. Rev. Microbiol.*, **50**, 727–751.
- Welch, M., Todd, D.E., Whitehead, N.A., McGowan, S.J., Bycroft, B.W. and Salmond, G.P. (2000) N-acyl homoserine lactone binding to the CarR receptor determines quorum-sensing specificity in *Erwinia*. *EMBO J.*, **19**, 631–641.
- Urbanowski, M.L., Lostroh, C.P. and Greenberg, E.P. (2004) Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein. *J. Bacteriol.*, **186**, 631–637.
- van Kessel, J.C., Ulrich, L.E., Zhulin, I.B. and Bassler, B.L. (2013) Analysis of activator and repressor functions reveals the requirements for transcriptional control by LuxR, the master regulator of quorum sensing in *Vibrio harveyi*. *MBio*, **4**, doi:10.1128/mBio.00378-13.
- Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.
- Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glielberman, A.L., Monie, D.D. and Endy, D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, **3**, 4.
- Saeidi, N., Wong, C.K., Lo, T.M., Nguyen, H.X., Ling, H., Leong, S.S., Poh, C.L. and Chang, M.W. (2011) Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol.*, **7**, 521.

MATHEMATICAL APPENDIX

According to the chemical reactions described in Equations (4), (5), (19) and (6), the expression of the protein reporter [G] (which can be tuned by the modification of the RBS parts of the genetic device) is determined by two parts: one constitutive, related to the receptor expression [R_T] and another one inducible, affected by the presence of ligand [L].

DERIVATION OF THE CONSTITUTIVE EXPRESSION OF THE RECEPTOR IN RELATION TO THE RELATIVE RBS STRENGTH

We assume that the receptor expression is controlled by a constitutive promoter with a linear degradation of the gene product. Prefixes m and p are used for transcription and translation of related species and kinetic constants, respectively. The resulting equations for the receptor are split in its transcription and translation processes as follows:

$$\frac{d[mR_T]}{dt} = k'_{m0} - k'_{md}[mR_T] \quad (26)$$

where [mR_T] is the mRNA species that, at the steady state, is defined as [mR_T]^{*} = k'_{m0}/k'_{md}. The translation process is described by the following ODE and the respective steady

state for the value of total receptor concentration:

$$\frac{d[R_T]}{dt} = k'_p[mR_T] - k'_{pd}[R_T] \quad (27)$$

$$[R_T^*] = k'_p k'_{m0} / (k'_{md} k'_{pd}). \quad (28)$$

Normalizing k'_p by the kinetic constant of the canonical RBS sequence B0034 (cita web) denoted by k_p^o we get that $k'_p = \phi k_p^o$. In this way we define ϕ as the relative strength of a particular RBS part, respecting the canonical sequence. Grouping all the constants except ϕ we obtain

$$r = k_p^o k'_{m0} / (k'_{md} k'_{pd}). \quad (29)$$

Then, we can express $[R_T^*]$ as a function of ϕ as follows

$$[R_T^*] = \phi r. \quad (30)$$

DERIVATION OF THE INDUCIBLE PART

According to the chemical reactions described in Equations (4), (5), (6) and (19), the expression of the protein reporter [G], is determined by the following equation:

$$\frac{d[G]}{dt} = k_0[P] + k_R[S_R] + k_L[S_L] - k_d[G]. \quad (31)$$

In this equation, we capture by order of the summations: (i) the basal expression due to the naked promoter [P], (ii) the expression by the promoter bound to receptor without ligand [S_R], (iii) the promoter bound ligand-receptor complex, denoted as [S_L] and (iv) the natural degradation of the protein [G]. Processes of transcription and translation are collapsed in the following kinetic constants for the alternative types of G synthesis: k_0 , k_R and k_L . This is possible by the steady-state assumption also for the RNA species. Transcription and translation processes are defined by these two ODEs:

$$\frac{d[mG]}{dt} = k_{m0}[P] + k_{mR}[S_R] + k_{mL}[S_L] - k_{md}[mG] \quad (32)$$

$$\frac{d[G]}{dt} = k_p[mG] - k_{pd}[G]. \quad (33)$$

Giving rise to the value of the protein G at the steady state:

$$[G^*] = k_p / (k_{md} k_{pd} (k_{m0}[P] + k_{mR}[S_R] + k_{mL}[S_L])) \quad (34)$$

We now normalize k_p by k_p^o , obtaining a relative strength for the RBS of the inducible part, defined as

$$\gamma = k_p / k_p^o. \quad (35)$$

By grouping some kinetic constants we get

$$g = k_p^o / (k_{md} k_{pd}) \quad (36)$$

obtaining the following expression:

$$G^* = \gamma g (k_{m0}[P] + k_{mR}[S_R] + k_{mL}[S_L]) \quad (37)$$

where γ represents the normalized RBS strength used in a particular construct. Now we shall proceed to get an expression in terms of parameters of the system. For the sake of simplicity, we will start with obtaining the expression as a function of R_2 and then to give the final expression in terms of the RBS strengths. This provides a more tractable and intuitive expression for the behavior of the gene expression.

According to model premises, $[R_T]$ and $[P_T]$ satisfy the following conservation expressions:

$$[R_T] = [R] + [R_2] + [Q] + [S_R] + [S_L] \quad (38)$$

$$[P_T] = [P] + [S_R] + [S_L]. \quad (39)$$

$[R_T] \gg [P_T]$, $[S_R]$ and $[S_L]$ are not considered in Equation (38) giving rise to a simpler equation:

$$[R_T] = [R] + [R_2] + [Q]. \quad (40)$$

Taking the equilibrium constants of Equations (4), (5), (19) and (6), we express $[S_R]$, $[S_L]$ and $[Q]$ in function of $[R_{2T}]$ and $[L]$ and $[P]$:

$$[Q] = K_1 R_2 [L] \quad (41)$$

$$[S_R] = K_2 [P][R_2] \quad (42)$$

$$[S_L] = K_3 K_1 [P][R_2][L]. \quad (43)$$

Now, from the equation conservation defined in Equation (40), we replace the [Q] expression of the Equation (41) and we get

$$[R_2] = \frac{[R_{2T}]}{(1 + K_1[L])}. \quad (44)$$

Now, from the conservation equation of $[P_T]$ (see Equation (11)) we derive the expression of [P] by replacing $[S_R]$ and $[S_L]$ from Equations (42) and (43). Accordingly, we obtain

$$[P] = \frac{[P_T]}{(1 + (K_2 + K_1 K_3 [L])[R_2])} \quad (45)$$

Replacing [Q], $[S_R]$ and $[S_L]$ from Equations (41), (42) and (43), and [P] from (45) in Equation (37), we obtain after re-ordering

$$[G^*] = \gamma g P_T \frac{k_{m0} + (k_{mR} K_2 + k_{mL} K_1 K_3 [L])[R_2]}{1 + (K_2 + K_1 K_3 [L])[R_2]}. \quad (46)$$

Now, replacing $[R_2]$ by the eq. (44) we obtain

$$G^* = \frac{\gamma g [P_T] (\sigma_1 + \sigma_2 L)}{\sigma_3 + \sigma_4 L} \quad (47)$$

where

$$\sigma_1 = \frac{k_{m0}}{R_{2T}} + k_{mR} K_2, \quad \sigma_2 = \frac{k_{m0} K_1}{R_{2T}} + k_{mL} K_1 K_3 \quad (48)$$

$$\sigma_3 = \frac{1}{R_{2T}} + K_2, \quad \sigma_4 = \frac{K_1}{R_{2T}} + K_1 K_3. \quad (49)$$

We can rewrite Equation (47) as follows

$$G^* = \gamma g P_T \frac{\sigma_2}{\sigma_4} \left(\frac{\sigma_1 / \sigma_2 + L}{\sigma_3 / \sigma_4 + L} \right) \quad (50)$$

and we finally obtain the expression (18) by defining

$$a_1 = \gamma g P_T \sigma_2 / \sigma_4 \quad a_2 = \sigma_1 / \sigma_2 \quad a_3 = \sigma_3 / \sigma_4. \quad (51)$$

RELATION BETWEEN [R_{2T}] AND [R_T]

For the [R_{2T}] value, the simplification allows to establish a constant relation with [R_T] just solving the second degree equation: $4K_0[R_{2T}]^2 + (1 - 4K_0[R_T])[R_{2T}] + K_0[R_T]^2 = 0$. Taking the positive solution of the equation and assuming that $K_0 \gg 1$, we obtain

$$[R_{2T}] \approx \frac{[R_T]}{2} = \frac{\phi r}{2} = \phi r'. \quad (52)$$

For the general case, rewriting (51) applying (52) we obtain the expressions

$$a_1 = \gamma g[P_T](k_{m0} + k_{mL}K_3r'\phi)/(1 + K_3r'\phi) \quad (53)$$

$$a_2 = K_1^{-1} \left(\frac{k_{m0} + k_{mR}K_2r'\phi}{k_{m0} + k_{mL}K_3r'\phi} \right) \quad (54)$$

$$a_3 = K_1^{-1}(1 + K_2r'\phi)/(1 + K_3r'\phi). \quad (55)$$

Simplified model (15) can be similarly treated applying (52) to obtain (18).