Article

Mechanistic Models of Inducible Synthetic Circuits for Joint Description of DNA Copy Number, Regulatory Protein Level, and Cell Load

Lorenzo Pasotti [1,2,*], Massimo Bellato [1,2], Davide De Marchi [1,2] and Paolo Magni [1,2]

1   Laboratory of Bioinformatics, Mathematical Modelling and Synthetic Biology, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, I-27100 Pavia, Italy; massimo.bellato01@universitadipavia.it (M.B.); davide.demarchi01@universitadipavia.it (D.D.M.); paolo.magni@unipv.it (P.M.)
2   Centre for Health Technologies, University of Pavia, I-27100 Pavia, Italy
*   Correspondence: lorenzo.pasotti@unipv.it; Tel.: +39-0382-98-5511

Abstract: Accurate predictive mathematical models are urgently needed in synthetic biology to support the bottom-up design of complex biological systems, minimizing trial-and-error approaches. The majority of models used so far adopt empirical Hill functions to describe activation and repression in exogenously-controlled inducible promoter systems. However, such equations may be poorly predictive in practical situations that are typical in bottom-up design, including changes in promoter copy number, regulatory protein level, and cell load. In this work, we derived novel mechanistic steady-state models of the lux inducible system, used as case study, relying on different assumptions on regulatory protein (LuxR) and cognate promoter (P_{lux}) concentrations, inducer-protein complex formation, and resource usage limitation. We demonstrated that a change in the considered model assumptions can significantly affect circuit output, and preliminary experimental data are in accordance with the simulated activation curves. We finally showed that the models are identifiable a priori (in the analytically tractable cases) and a posteriori, and we determined the specific experiments needed to parametrize them. Although a larger-scale experimental validation is required, in the future the reported models may support synthetic circuits output prediction in practical situations with unprecedented details.

Keywords: mathematical modeling; mechanistic model; synthetic biology; copy number; inducible promoter; cell load; bottom-up design

1. Introduction

Recent advances in the construction and characterization of DNA encoded synthetic circuits have enabled to boost the design-build-test engineering cycle applied to biological systems, in terms of time and economic resources. This process has led to the engineering of complex engineering-inspired information processing and control systems, as well as solutions to numerous problems in industrial biotechnology and medicine [1–3].

Given a genetic circuit, its promoters, ribosome binding sites (RBSs) and the DNA copy number of its elements are some of the crucial degrees of freedom that biological engineers have to tune to implement an effectively working function [4,5]. However, the combinatorial search space to properly tune complex regulatory networks makes a trial-and-error approach unpractical with no guarantee to reach an optimally working system [6,7].

High-impact industrially-attractive control systems for different applications have been already delivered by synthetic biology, for example to adaptively regulate biofuel production as a function of pathway intermediates availability [8] or to switch cellular metabolism towards growth or product

formation as a function of cell stress [9]. Although the mentioned circuits are characterized by an elegant design and have real-world impact, the final correctly working systems were obtained only after random selection steps—e.g., to tune the expression levels of key proteins—or the affinity of repressors exerting feedback control [10]. These and a number of other examples suggest that a bottom-up design process is desirable to increase the success rate in the construction of working biological systems that behave as intended. To this aim, accurate predictive models are needed to support this design step.

Among the many issues currently limiting model-based approaches in synthetic biology, several unpredictability sources affect the re-use of biological parts in different contexts (i.e., strains, growth media, and even circuits). Cell-to-cell variability, flanking regions-dependent behavior, and cell load are among the major features causing such variability [11–16]. Many of such effects have often been neglected in mathematical models, thereby limiting their predictive power. In addition, widely used modeling approaches describe recombinant protein regulation via empirical Michaelis–Menten or Hill equations [17].

Empirical models are popular tools and have many advantages (e.g., low number of parameters, overall good descriptive capability and no need to know the biomolecular interactions underlying the described process) [18,19]. However, they may have poor predictive power on unseen data when one or more circuit elements are changed [20], like copy number of DNA or protein regulators that have been reported to be essential for the biological systems behavior [21]. In specific, transfer functions of inducible devices can be characterized in vivo through dose–response experiments, in which a constitutively expressed regulator (activator or repressor) is activated or inhibited by an exogenously added molecule and the complex eventually affects the transcription of a cognate regulated promoter [22]. While empirical models can be easily identified from such experimental data, it is not trivial to generalize them in situations in which molecule copy number changes, also for the empirical nature of model parameters [20]. This weak aspect of empirical models might be crucial in practical bottom-up design situations, in which circuit parts are interconnected and the behavior of the system is predicted from the functioning of individual parts, by using previously estimated parameters [23].

Mechanistic models are able to overcome some of the issues mentioned above: parameters usually have biological meaning (e.g., dissociation constants, copy numbers, etc.) and predictive power is expected to be higher than empirical models, since circuit changes can be translated into the variation of specific parameters [20]. However, mechanistic models usually have a larger number of parameters to be estimated, thereby raising issues of model identifiability [18,24]. Since such models require a deep knowledge of occurring biomolecular interactions in the biological system under study, they are less popular than empirical models. Different mechanistic modeling efforts have been undertaken, describing gene regulation using thermodynamics or law of mass action [18,25–28]. However, the application of these models in bottom-up design approaches remains low due to the absence of studies on their identification and lack of broad-range advantages demonstrations over the empirical ones.

In this work, we presented different mechanistic models of the widely used lux inducible system [29], derived under different key assumptions on molecule abundances and resource limitation, thereby obtaining mathematical tools able to describe copy number changes and the burden effects caused by individual parts.

Our objective was to investigate the impact of the underlying assumptions on the transfer function of the lux inducible circuit via numerical simulations. The inclusion of the considered features in a mechanistic model represents a novel contribution of this work. We extended the work by studying the usability of such models in terms of structural and practical identifiability, given experimental measurements commonly available in synthetic biology in vivo studies.

We showed that different assumptions can lead to relevant differences among activation curves features (i.e., all the Hill function parameters showed variations). We also showed that simulations were consistent with preliminary in vivo data, measured in recombinant *Escherichia coli* in previous and novel studies, although a larger-scale validation of the model variants will be required. Finally,

model identifiability analysis outlined the experimental designs required to enable accurate parameter estimation.

This study is expected to have high relevance for both computational biologists, since new models have been proposed and the impact of their key assumptions evaluated in silico, and for experimental biologists, since recommendations on the measurement procedures have been provided, supporting the future in vivo assessment of the models. Taken together, the results found in this work may support bottom-up design of synthetic biological systems by providing new models that, once validated, will represent accurate tools for synthetic circuits behavior prediction.

2. Materials and Methods

2.1. Inducible System Description

The inducible system studied in this work is described in Figure 1a. It includes a constitutively expressed LuxR protein, which acts as transcriptional activator of the cognate P$_{lux}$ promoter in presence of N-(3- oxohexanoyl)-L-homoserine lactone (HSL). The LuxR-HSL active complex binds to the lux box of the P$_{lux}$ promoter, triggering its activity in an HSL concentration-dependent fashion, thereby regulating the transcription of the downstream gene, which is then translated into protein. We considered a set of reactions affecting the production of the (reporter) protein-encoded gene regulated by P$_{lux}$, similar to the ones described by Carbonell-Ballestero et al. [20] (Figure 1b). Specifically, a LuxR protein dimer (R_2) and HSL (L) form a complex (Q), which binds the free promoter (P) to form a transcriptionally active bound promoter (S).

As an alternative reaction scheme, we also considered a similar set of biomolecular interactions, with complex formation occurring in two steps (Figure 1c): binding of R_2 and L to form Q, and subsequent binding of L to Q to form the hetero-tetramer Q_2. This alternative reaction set was defined to refine the mechanistic model above, in accordance with previous investigations, in which a hetero-tetrameric structure was suggested for the activated complex [30].
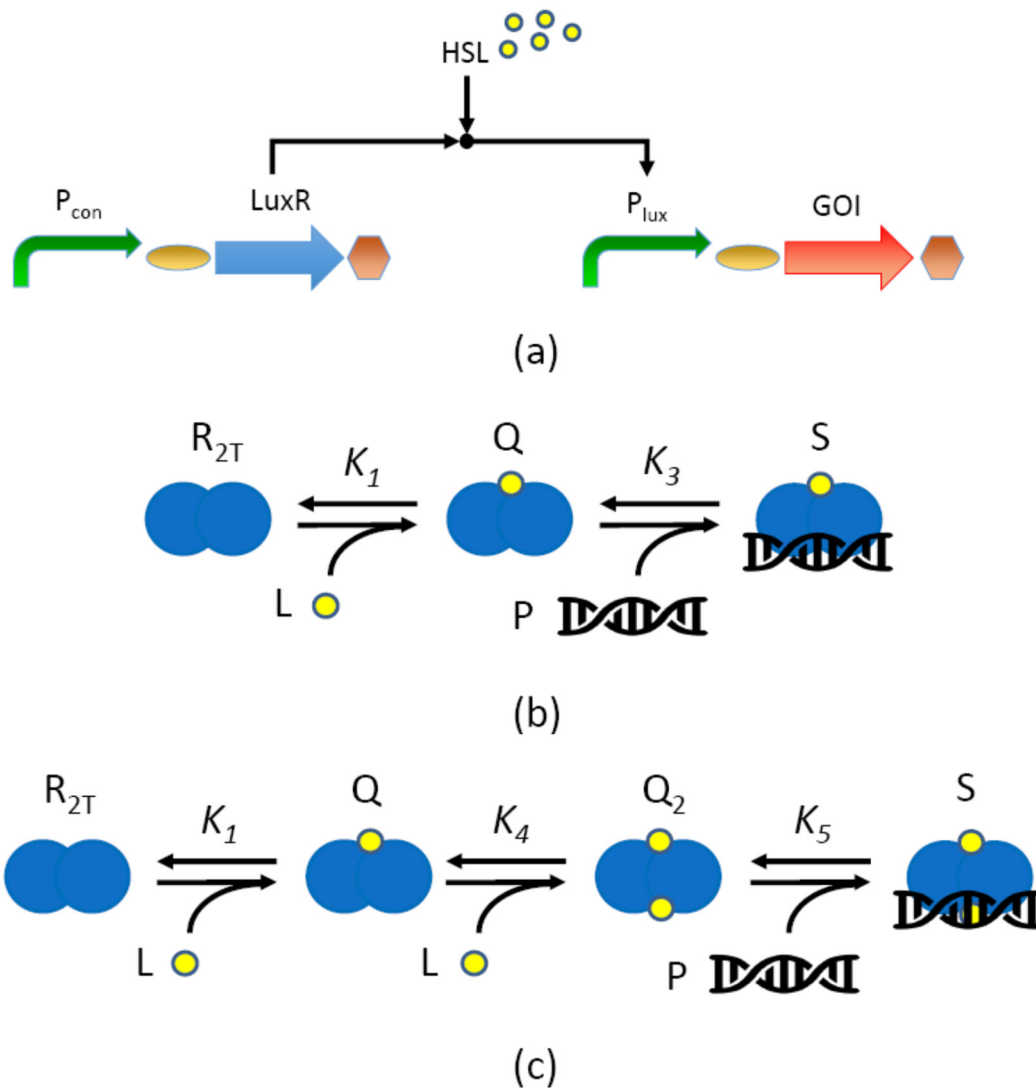
Figure 1. Description of the lux inducible system. (**a**) The gene encoding the LuxR protein transcriptional regulator is expressed by a constitutive promoter (P_{con}); the LuxR regulator becomes activated upon HSL molecule binding to form a complex which can bind the P_{lux} promoter in its single lux box sequence, thereby activating the expression of the gene of interest (GOI), placed downstream of P_{lux}. Curved arrows represent promoters; ovals represent ribosome binding sites (RBSs); straight arrows represent coding sequences; hexagons represent transcriptional terminators; circles represent HSL molecules. (**b**) and (**c**) Biomolecular reactions modeled in this work. The LuxR dimer (R_{2T}, blue overlapping circles) is assumed to be activated upon binding of one (b) or two (c) HSL molecules (L, yellow circles) to form an activated complex (Q or Q_2, respectively), which binds DNA (double helix icon) to enable transcription. In panel (c), Q is the LuxR dimer form bound to one HSL molecule and it is assumed to be unable to bind DNA. The equilibrium constants are reported for each reaction occurring in panels (b) and (c).

2.2. Model Definition

In this section, different models describing the lux system are illustrated. In particular, the essential elements of gene expression modeling are reported in Section 2.2.1, together with the description of a basic empirical model for transcriptional activation. The description of all the models used in this work is provided in Sections 2.2.2–2.2.5. Due to the complexity of model structure under some assumptions, the theoretical analysis of the activation function parameters is only provided for the models described in Sections 2.2.1 and 2.2.2.

2.2.1. Empirical Michaelis–Menten Model (M0)

Assumptions

Assuming that no post-transcriptional or post-translational regulations are involved in the circuit, the intracellular dynamics of mRNA (M), immature protein (X) concentration regulated by P$_{lux}$ and the mature form of the protein (Y) can be described via Equations (1)–(3):

$$\frac{dM}{dt} = H(L) - \gamma_M \times M \tag{1}$$

$$\frac{dX}{dt} = \rho \times M - (\gamma_X + \sigma) \times X \tag{2}$$

$$\frac{dY}{dt} = \sigma \times X - \gamma_X \times Y \tag{3}$$

In Equation (1), γ_M is the mRNA degradation rate and $H(L)$ is an HSL-dependent activation function describing transcription rate, expressed in (RNA time^{-1}) units. In Equation (2), γ_X (time^{-1}) includes the protein degradation and dilution rates, and ρ (time^{-1}) is the translation rate. If the expressed protein is very stable, γ_X is equal to the rate of cell division. When relevant in terms of dynamics, protein maturation or folding is also included: Y represents mature protein, σ (time^{-1}) represents protein maturation rate and protein degradation rate is assumed to be the same for immature and mature forms. Assuming the steady-state of all the intracellular processes, the output commonly considered for such system, i.e., the mature protein synthesis rate per cell (y), which is equal to the synthesis term of Equation (3), is proportional to $H(L)$ (Equation (4))

$$y = \frac{\sigma \times \rho \times H(L)}{(\gamma_X + \sigma) \times \gamma_M} \tag{4}$$

Another output commonly found in literature is the per-cell mature protein (Y), which is proportional to y, thereby enabling to generalize all the modeling work presented in this study.

In many works considering a constant LuxR production, not changing throughout the experiments [22,23,31–34], $H(L)$ is modeled via a Hill equation and the circuit output, y, can be written as in Equation (5)

$$y = \delta + \frac{\alpha}{1 + \left(\frac{\kappa}{L}\right)^{\eta}} \tag{5}$$

Assuming no cooperativity, Equation (5) with $\eta = 1$ is equivalent to a Michaelis–Menten equation with three parameters: δ is the basic protein synthesis rate when P$_{lux}$ is in its off state; α is the activity range in the on state; κ is the HSL concentration corresponding to half of the maximum activation [20]. The δ and α parameters are expressed as intracellular protein concentration per time. The experimental measurements routinely performed in laboratory to characterize synthetic circuits usually exploit fluorescent reporter proteins, which are quantified via in vivo assays by means of plate reader or flow-cytometry. In these cases, per-cell arbitrary units of fluorescence (AU) can be adopted to express intracellular protein concentration, assuming their proportionality.

Derivation

When LuxR level is also needed to be described (situation of interest in the present work), another commonly found modeling approach includes the following equations (Equations (6) and (7)) [30,33,35], describing the LuxR-HSL complex formation and the subsequent activation of protein synthesis by this complex

$$C = \frac{U}{1 + \left(\frac{\kappa_R}{L}\right)^{\beta}} \tag{6}$$

$$y = \hat{\delta} + \frac{\hat{\alpha}}{1 + \left(\frac{\hat{\kappa}}{C}\right)^{\hat{\eta}}} \tag{7}$$

In Equation (6), C is the intracellular LuxR-HSL complex concentration, U is the total LuxR concentration, κ_R is the concentration of HSL required for half-activation of LuxR, and β is the Hill coefficient. In Equation (7), symbols have the same meaning as in Equation (5), with the cap denoting that this Hill function has C as input. For this reason, $\hat{\kappa}$ has the same units as C and U (protein concentration or AU, as explained above). Assuming no cooperativity, as before, the Hill coefficients (β and $\hat{\eta}$) can be fixed to 1. The expressions in Equations (6) and (7) can be lumped into a single equation describing y as a function of L (Equation (8))

$$y = \hat{\delta} + \frac{\hat{\alpha}/(1 + \hat{\kappa}/U)}{1 + \frac{\kappa_R/(1 + U/\hat{\kappa})}{L}} \tag{8}$$

which is equivalent to the Michaelis–Menten function in Equation (5) with the following parameters (Equations (9)–(11))

$$\delta = \hat{\delta} \tag{9}$$

$$\alpha = \frac{\hat{\alpha}}{1 + \frac{\hat{\kappa}}{U}} \tag{10}$$

$$\kappa = \frac{\kappa_R}{1 + \frac{U}{\hat{\kappa}}} \tag{11}$$

Final Expression

Although the parameters of Equation (8) have empirical nature, the copy numbers of P$_{lux}$ (n_1) and LuxR (n_2) can be included in the model. Specifically, n_1 and n_2 can act as known scale factors of $\hat{\delta}$, $\hat{\alpha}$ and U (Equations (12)–(14)), thereby enabling the description of copy number changes among different situations. Following a bottom-up approach, model parameters ($\hat{\delta}$ or d, $\hat{\alpha}$ or v, $\hat{\kappa}$ and κ_R) can be estimated from experimental data and the parametrized model can be adopted to predict unseen situations with different n_1 and n_2

$$\hat{\delta} = n_1 \times d \tag{12}$$

$$\hat{\alpha} = n_1 \times v \tag{13}$$

$$U = n_2 \times u \tag{14}$$

As before, the system output (y) can be expressed in per-time intracellular protein concentration or AU time^{-1}, like $\hat{\delta}$ and $\hat{\alpha}$. Analogously, U and $\hat{\kappa}$ are intracellular protein concentrations that can also be expressed in AU, while κ_R has the same units as HSL concentration. For this reason, the relative copy number changes, instead of the absolute ones, are sufficient to express n_1 and n_2. Specifically, n_1 can be set according to plasmid copy number and n_2 is proportional to the strength of the promoter expressing LuxR. Finally, the u value (Equation (14)) can be set to 1 without any loss of generality, since U is always present in ratio with $\hat{\kappa}$ (Equation (8)). The final empirical model, called M0, describing the output of the lux inducible system as a function of HSL and the per-cell copy numbers of P$_{lux}$ and LuxR, is reported in Equation (15)

$$y = n_1 \cdot d + \frac{n_1 \times v/(1 + \hat{\kappa}/(n_2 \times u))}{1 + \frac{\kappa_R/(1 + n_2 \times u/\hat{\kappa})}{L}} \tag{15}$$

2.2.2. Mechanistic Model with LuxR Abundance Assumption (M1)

Assumptions

A mechanistic model of the biomolecular reactions shown in Figure 1b has been previously reported [20] and is herein briefly illustrated. The main underlying assumptions are that HSL molecules bound to LuxR are negligible compared to the total HSL amount, and that $P << R$. A difference from the work by Carbonell-Ballestero et al. is that no binding is assumed between LuxR dimer and P$_{lux}$, and the low-entity promoter activation by this unspecific complex is neglected. Previous experimental work showed that this activity increase is undetectable in the commonly tested conditions [35].

Derivation

Considering the same equations for transcription, translation and maturation (Equations (1)–(3)) as in Section 2.2.1, $H(L)$ can be expressed in a more mechanistic fashion (Equation (16))

$$H(L) = k_{m0} \times P + k_{mL} \times S \tag{16}$$

where P and S represent the intracellular concentrations of free and complex-bound promoter, while k_{m0} and k_{mL} are the transcription rate constants in off and on state, respectively.

Based on the law of mass action, the biomolecular interactions in Figure 1b depend on the following equilibrium constants and conservation expressions (Equations (17)–(20))

$$K_1 = \frac{Q}{L \times R_2} \tag{17}$$

$$K_3 = \frac{S}{P \times Q} \tag{18}$$

$$P_T = P + S \tag{19}$$

$$R_{2T} = R_2 + Q + S \tag{20}$$

Symbols in Equations (17)–(20) are described in Figure 1b; briefly, R_{2T} and R_2 are the total and free LuxR dimer concentrations, respectively, where $R_{2T} \approx R_T/2$ (and R_T is the total concentration of LuxR monomer). Under the assumption that $P << R$, the bound promoter S in Equation (20) can be neglected: $R_{2T} = R_2 + Q$. Considering Equations (1)–(3) and Equations (16)–(20), the following expressions for LuxR-HSL complex and circuit output can be written (Equations (21) and (22)) [20]

$$Q = \frac{R_{2T}}{1 + \dfrac{1}{K_1 \times L}} \tag{21}$$

$$y = \hat{k}_{m0} \times P_T + \frac{\left(\hat{k}_{mL} - \hat{k}_{m0}\right) \times P_T}{1 + \dfrac{1}{K_3 \times Q}} \tag{22}$$

where $\hat{k}_{m0}$ and $\hat{k}_{mL}$ are k_{m0} and k_{mL}, respectively, scaled by $\sigma \times \rho/((\gamma_X + \sigma) \times \gamma_M)$ (see Equation (4)).

These equations show that, like in the empirical model in Equations (6) and (7), in this mechanistic model the intracellular complex concentration and circuit output can be both described by Michaelis–Menten functions. Analogously, circuit output as a function of HSL is a Michaelis–Menten function (Equation (23)) with coefficients described in Equations (24)–(26)

$$y = P_T \times \hat{k}_{m0} + \frac{P_T \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)}{1 + 1/(K_3 \times R_{2T})} \times \frac{1}{1 + 1/(K_1 \times (1 + K_3 \times R_{2T}) \times L)} \tag{23}$$

$$\delta = P_T \times \hat{k}_{m0} \tag{24}$$

$$\alpha = \frac{P_T \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)}{1 + 1/(K_3 \times R_{2T})} \tag{25}$$

$$\kappa = 1/\left(K_1 \times (1 + K_3 \times R_{2T})\right) \tag{26}$$

In summary, M0 (Equation (15)) and M1 (Equation (23)) have the same $y(L)$ function and analogous LuxR-dependent Michaelis–Menten parameters expressions (Equations (9)–(11) and Equations (24)–(26)). However, the mechanistic nature of M1 gives the advantages that molecule concentrations are explicitly described and the link between model parameters and biological mechanism is not lost. As a result, M1 offers the opportunity to parametrize the model with biologically meaningful parameters, like intracellular concentrations, equilibrium constants, and copy numbers. Nonetheless, hard-to-measure quantities can still be expressed in AUs instead of concentrations, as in M0. For instance, while promoter concentration and DNA copy number are easy to retrieve (e.g., from assumptions about cell volume and plasmid datasheets or quantitative PCR), protein concentration requires more resource consuming experiments (e.g., Western blot) and depends not only on DNA copy number, but also on transcription, translation and degradation rates [20]. Nevertheless, the relative protein level can still be approximated via the strength of the upstream promoter.

For the reasons above, in this work M0 will not be further analyzed, and only serves as a reference for M1 parameters expressions.

Final Expression

The final expression of M1, explicitly including n_1 and n_2, is reported in Equation (27)

$$y = n_1 \times P_U \times \hat{k}_{m0} + \frac{n_1 \times P_U \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)/(1 + 1/(K_3 \times n_2 \times r_T))}{1 + 1/(K_1 \times (1 + K_3 \times n_2 \times r_T) \times L)} \tag{27}$$

where P_U and r_T are the intracellular concentrations of one DNA copy of promoter and LuxR protein, respectively. The n_1 and n_2 parameters are, respectively, the P$_{lux}$ copy number, and the scale factor between one protein monomer and the actual dimer concentration.

2.2.3. Mechanistic Model without LuxR Abundance Assumption (M2)

Assumptions and Final Expression

Following the same biomolecular reactions and calculation steps as in M1, it is possible to calculate circuit output without the $P \ll R$ assumption. In this situation, S cannot be neglected in Equation (20). Circuit output can be computed from Equation (28), in which R_2 and P are the roots of a second order two-equation system (Equations (29)–(30)). In this equation system, only one root has a biologically acceptable meaning for any value of L (not shown). Analytical formulas expressing R_2 and P are not reported due to their complexity and, as a result, Equations (28)–(30) represent the final reported expression for M2

$$y = P \times \left(\hat{k}_{m0} + \hat{k}_{mL} \times K_1 \times K_3 \times R_2 \times L\right) \tag{28}$$

$$\begin{aligned}(K_1 \times K_3 \times L) \times P^2 + (1 + K_1 \times L + K_1 \times K_3 \times n_2 \times r_T \times L \\ - K_1 \times K_3 \times n_1 \times P_U \times L) \times P - (P_T + K_1 \times n_1 \times P_U \times L) = 0\end{aligned} \tag{29}$$

$$\begin{aligned}(K_1 \times K_3 \times L + K_1^2 \times K_3 \times L^2) \times R_2^2 \\ + (1 + K_1 \times L + K_1 \times K_3 \times n_1 \times P_U \times L - K_1 \times K_3 \times n_2 \times r_T \times L) \\ \times R_2 - n_2 \times r_T = 0\end{aligned} \tag{30}$$

2.2.4. Modeling LuxR-HSL Hetero-Tetramerization (M1T, M2T)

Assumptions

The biomolecular interactions illustrated in Figure 1c can be used to derive mathematical models in which, differently from Figure 1b, the activated complex is a hetero-tetramer, formed by two LuxR and two HSL molecules. In particular, under the $P << R$ assumption, we also assumed that: i) the LuxR dimer has two binding sites for HSL; ii) the probability of HSL binding to a free site of R_2 and Q is the same (i.e., there is no cooperative behavior); iii) the probability of HSL unbinding from an occupied site of R_2 and Q is the same.

Derivation

The output can be expressed as in Equation (16), the K_1 equilibrium constant as in Equation (17) and the conservation of total, free and bound promoter as in Equation (19). The following expressions describe equilibrium constants K_4 and K_5, as well as the LuxR dimer (free, bound with L, bound with $2L$ and bound with $2L$ and the promoter) conservation (Equations (31)–(33))

$$K_4 = \frac{Q_2}{Q \times L} \tag{31}$$

$$K_5 = \frac{S}{P \times Q_2} \tag{32}$$

$$R_{2T} = R_2 + Q + Q_2 + S \tag{33}$$

Thanks to the $P << R$ assumption, as in M1 the S concentration becomes negligible in Equation (33), thereby having: $R_{2T} = R_2 + Q + Q_2$.

The forward and reverse rate constants in Equation (34) (k_+ and k_-) and Equation (35) (k'_+ and k'_-), describing the two HSL binding steps, have the following relations: $k_+ = 2 \times k'_+$ and $2 \times k_- = k'_-$:

$$R_2 + L \overset{k_+}{\underset{k_-}{\rightleftarrows}} Q \tag{34}$$

$$Q + L \overset{k'_+}{\underset{k'_-}{\rightleftarrows}} Q_2 \tag{35}$$

For the described reasons, a relation between K_1 and K_4 can be written (Equation (36))

$$K_4 = \frac{k'_+}{k'_-} = \frac{k_+}{4 \cdot k_-} = \frac{K_1}{4} \tag{36}$$

Final Expression

Based on the relations above and following the same mathematical steps as in M1, the final output for this model (M1T) is reported in Equation (37); the M1T expression in Equation (37) does not resemble a Michaelis–Menten function (differently from M1).

$$
\begin{aligned}
y &= n_1 \times P_U \\
&\times \frac{\hat{k}_{m0} + (\hat{k}_{m0} \times K_1) \times L + (\hat{k}_{m0} \times K_1^2/4 + \hat{k}_{mL} \times K_1^2/4 \times K_5 \times n_2 \times r_T) \times L^2}{1 + K_1 \times L + (K_1^2/4 + K_1^2/4 \times K_5 \times n_2 \times r_T) \times L^2}
\end{aligned}
\tag{37}
$$

Without the $P << R$ assumption, the expressions of circuit output become more complex. By following the same steps previously done in 2.2.3 for M2, the final model (M2T) is described by the following expressions (Equations (38)–(40)).

$$y = P \times \left(\hat{k}_{m0} + \hat{k}_{mL} \times \frac{K_1^2}{4} \times K_5 \times R_2 \times L \right) \tag{38}$$

$$\left(\frac{K_1^2}{4} \times K_5 \times L^2 + \frac{K_1^3}{4} \times K_5 \times L^3 + \frac{K_1^4}{16} \times K_5 \times L^4 \right) \times R_2^2 + (1 + K_1 \times L + \frac{K_1^2}{4} \times L^2 + \dots$$

$$\dots \frac{K_1^2}{4} \times K_5 \times n_1 \times P_U \times L^2 - \frac{K_1^2}{4} \times K_5 \times n_2 \times r_T \times L^2) \times R_2 - n_2 \times r_T = 0 \tag{39}$$

$$\left(\frac{K_1^2}{4} \times K_5 \times L^2 \right) \times P^2 + \left(1 + K_1 \times L + \frac{K_1^2}{4} \times L^2 + \frac{K_1^2}{4} \times K_5 \times n_2 \times r_T \times L^2 - \frac{K_1^2}{4} \times K_5 \times n_1 \times P_U \times \right.$$

$$\left. L^2 \right)$$

$$\dots n_1 \times P_U \times \left(1 + K_1 \times L + \frac{K_1^2}{4} \times K_5 \times L^2 \right) = 0 \tag{40}$$

2.2.5. Modeling Cell Load (M1L, M2L)

Assumptions

Unless differently indicated, in the two-gene circuit considered in this work (Figure 1a) we assume that only the protein with expression regulated by P$_{lux}$ is characterized by a relevant resource usage, while LuxR does not cause relevant burden. Since LuxR is constitutively expressed, thereby giving constant load, this assumption will not affect any HSL-dependent function, as previously discussed for genes with constant resource usage acting as background [23].

Derivation

A mechanistic model has been recently proposed to describe the effects of cell load caused by the expression of proteins with high resource demand [12,36]. In such context of transcriptional/translational resource limitation, the synthesis rates of all the proteins of a synthetic circuits are globally scaled by a factor D (Equation (41)).

$$D = 1 + \sum_{i=1}^{c} J_i \times s_i \tag{41}$$

where c indicates the number of expressed proteins in the synthetic circuit, s_i represents the synthesis rate of the i-th protein and J_i its resource usage parameter. Model derivation is extensively discussed in the original publications [12,36], while in this study only the expression with lumped parameters is reported and used (Equation (41)).

Final Expression

The final expression of model output in presence of cell load is reported in Equations (42)–(44).

$$D(L, n_1, n_{2,b}) = 1 + E + J \times \left(\frac{\sigma + \gamma_X}{\sigma} \right) \times y(L, n_1, n_{2,b}) \tag{42}$$

$$n_{2,b}(L, n_1, n_{2,b}) = \frac{n_2}{D(L, n_1, n_{2,b})} \tag{43}$$

$$y_b(L, n_1, n_{2,b}) = \frac{y(L, n_1, n_{2,b})}{D(n_1, n_{2,b})} \tag{44}$$

where y_b and $n_{2,b}$ indicate the model output and the LuxR scale factor affected by cell load, y is the output of one of the models described in Section 2.2.1–2.2.4, n_1 and n_2 have the same meaning as before, and J is the resource usage parameter (in min/AU) associated to the output protein. Finally, the contribution of an external constant load (E) was studied analogously, with the exception that the value of E was added to Equation (42).

2.3. Model Analysis

2.3.1. Model Parametrization

Models were parametrized using plausible values, according to available biological knowledge (e.g., DNA/protein concentrations) and previously published experimental data (e.g., P$_{lux}$ activation curve, resource usage [23,31,37]). A summary of parameters is provided in Table 1 for each model used in this work. When indicated, structural parameters were fixed to the reported values in simulations and they were assumed to be unknown during parameter estimation tasks (e.g., when studying a posteriori identifiability). *E. coli* cell volume was assumed to be 1 μm^3, corresponding to 10^{-15} L. Under this assumption, the concentration of one molecule of promoter DNA or protein corresponds to 1.66 nM [38]. A variation of the average *E. coli* cell volume has been previously reported in the range 0.5–2 μm^3 for different growth rates and conditions [38]. This variation is expected to affect the absolute values but not the trends of the numerical simulations shown of this work, thereby not affecting the drawn conclusions. Nonetheless, the variation of cell volume will be highly relevant in the analysis of real data, in which model parameters have to be estimated based on the knowledge of this value.

As anticipated in Section 2.2.2, to enable the application of the models in popular situations occurring in synthetic biology (i.e., model identification with the data routinely measured in fluorescent reporter protein-based assays) promoter copy number (n_1) was assumed to be available; on the other hand, the actual LuxR protein concentration (dependent from luxR DNA copy number, transcription rate, translation rate, dimerization, and mRNA/protein degradation) is harder to measure. While in simulation the R_{2T} quantity was spanned to explore the effects of having wide ranges of LuxR values, its value was assumed to be unavailable during model identification. Nonetheless, the relative strength of the promoter expressing the luxR gene is commonly known, thereby enabling to approximate the relative level of LuxR. For model amenability reasons, in estimation steps we parametrized n_2 as the (known) relative strength of this promoter (in AU) and r_T as the (unknown) scale factor between protein concentration (in nM) and AU, in which all the biologically occurring processes, described above, are lumped without any loss of generality and maintaining the mechanistic nature of the model. In M1, the r_T quantity always appears multiplied by K_3 (Equation (27)) and, for this reason, only their product is identifiable. By defining $\widehat{K}_3 = K_3 \times r_T$ (in AU^{-1}), the M1 model can be re-parametrized as Equation (45)

$$y = n_1 \times P_U \times \hat{k}_{m0} + \frac{n_1 \times P_U \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)/\left(1 + 1/\left(\widehat{K}_3 \times n_2\right)\right)}{1 + 1/\left(K_1 \times \left(1 + \widehat{K}_3 \times n_2\right) \times L\right)} \tag{45}$$

Analogously, the $K_5 \times r_T$ product in M1T (Equation (37)) can be re-parametrized as $\widehat{K}_5 = K_5 \times r_T$ (in AU^{-1}), thereby yielding the model in Equation (46)

$$y = n_1 \times P_U \times \frac{\hat{k}_{m0} + \left(\hat{k}_{m0} \times K_1\right) \times L + \left(\hat{k}_{m0} \times K_1^2/4 + \hat{k}_{mL} \times K_1^2/4 \times \widehat{K}_5 \times n_2\right) \times L^2}{1 + K_1 \times L + \left(K_1^2/4 + K_1^2/4 \times \widehat{K}_5 \times n_2\right) \times L^2} \tag{46}$$

2.3.2. A Priori Identifiability

Given a circuit output expression with known form (e.g., a Michaelis–Menten or a rational function) with specific coefficients that could be estimated from data, an expression for each parameter of the model was found as a function of the theoretically known coefficients. If the system can be solved uniquely, the model was considered as structurally (or a priori) identifiable.

Table 1. Model parameters. The indicated values were used for simulation and, unless differently indicated, for the study of a posteriori identifiability.

Parameter	Units	Models	Value
P_U	nM	All	1.66
r_T	nM [a]	All	1.66
n_1	-	All	5[b]
n_2	- [a]	All	301[b]
γ_X	min^{-1}	All	0.01[c]
$\boldsymbol{\sigma}$	min^{-1}	All	0.0167
$\hat{k}_{m0}$	$AU\ nM^{-1}\ min^{-1}$	All	0.0192
$\hat{k}_{mL}$	$AU\ nM^{-1}\ min^{-1}$	All	1.907
K_1	nM^{-1}	M1, M1L, M2, M2L	0.001
		M1T, M1TL, M2T, M2TL	0.0055
K_3	nM^{-1}	M1, M1L, M2, M2L	0.2
K_4	nM^{-1}	M1T, M2T, M1TL, M2TL	0.001375
K_5	nM^{-1}	M1T, M2T, M1TL, M2TL	0.2
J_{RFP}	$min\ AU^{-1}$	M1L, M2L, M1TL, M2TL	0.04
E	-	M1L, M2L, M1TL, M2TL	0[d]

[a] During the study of a posteriori identifiability, n_2 was expressed as AU with known value, while r_T was expressed as nM/AU with unknown value; [b] When indicated, a range of copy number values was spanned; [c] Typical value of *E. coli* growth rate, which is used as the rate of intracellular protein dilution; [d] A value of 2 was used to simulate the presence of external load where indicated.

2.3.3. A Posteriori Identifiability

Given a model, simulated data were generated and parameters were estimated. Different experiments were simulated varying n_2 as required, while the other structural parameters were kept constant to the values in Table 1, unless differently indicated. Proportional Gaussian random noise was added to the simulated data with a coefficient of variation (CV) of 5% as default value; other values were also evaluated (0, 1, 2.5, and 10%). A limited number of data points (12) was assumed to be available, resembling a realistic dose–response experimental setup. Parameter estimation was performed via least squares method with the MATLAB R2017b (MathWorks, Natick, MA, USA) *lsqnonlin* routine. If the estimated parameters are consistent with the ones used to generate the data, the model is considered as practically (or a posteriori) identifiable. For each proportional error entity, 200 simulation and estimation steps were carried out, thereby identifying the model using different synthetic data with random noise. Relative estimation error ($REE = 100 \cdot |p_{est} - p_{true}|/p_{true}$, where p_{est} and p_{true} are the estimated and true parameter values, respectively) was used to express parameter consistency. Uncertainty of parameter estimates, in terms of CV, was also computed as reported previously [39]. For each run, the maximum REE and CV among all the estimated parameters was considered.

2.3.4. Simulations

The MATLAB roots function was used to find P and R_2 as polynomial roots to solve the M2, M2T, and M2L model equations. Implicit equations, commonly occurring in models including cell load terms, were solved using the fixed point method, as previously described [23].

2.3.5. Analysis of Activation Curves

Hill function parameters, described in Equation (5), were calculated for each activation curve: δ and α were computed as the y values at lowest and highest L value, respectively; κ was computed as the value of L corresponding to half-maximum activation; η was computed according to Equation (47).

$$\eta = log(81)/log(L90/L10) \tag{47}$$

where $L90$ and $L10$ are the L values corresponding to 90% and 10% of the maximum value of y.

On the other hand, in vivo measured activation curves were fitted with Equation (5) to estimate its parameters, as described in Section 2.2.3.

2.3.6. In Vivo Experiments

Circuit output, i.e., RFP synthesis rate per cell (S$_{cell}$) at steady-state, was measured for recombinant MG1655-Z1 strain [31] bearing the low-copy plasmid pSB4C5 with X$_3$r as insert [23]. In this construct, previously described in [23] and with sequence available as BBa_J107032 code in the Registry of Standard Biological Parts (http://parts.igem.org), luxR is under the control of an anhydrotetracycline (ATc) inducible promoter, which works as gene expression knob in *Escherichia coli* strains overexpressing TetR, like the one used in this study.

The detailed experimental protocol for S$_{cell}$ measurement was previously described [23]. Briefly, 0.5 ml of M9 medium (11.28 g/L M9 salts, 1 mM thiamine hydrochloride, 2 mM MgSO$_4$, 0.1 mM CaCl$_2$, 0.2% casamino acids and 0.4% glycerol) were inoculated with a colony from a freshly streaked LB agar plate in a 2-ml tube and the culture was incubated at 37 °C, 220 rpm for at least 16 h. The grown culture was 100-fold diluted in 200 μL of M9 in a 96-well microplate. ATc and HSL (2 μl) were added to the microplate wells to reach the desired concentrations. The microplate was incubated at 37 °C in the Infinite F200 (Tecan) microplate reader and an automatic measurement procedure was programmed via the i-control software v.2.0.10 (Tecan, Switzerland): shaking (15 s, 3-mm amplitude), wait (5 s), optical density (600 nm) acquisition, red fluorescence (excitation: 535 nm, emission: 620 nm, gain = 50) acquisition, sampling time = 5 min. Raw data time series were background-subtracted using sterile medium (absorbance) and a non-fluorescent culture (fluorescence), incubated in the same experiment. The resulting data were used to compute S$_{cell}$ as the numeric time derivative of fluorescence, divided by absorbance over time. S$_{cell}$ was averaged in the exponential growth phase, typically occurring at absorbance values between 0.05 and 0.18 [32]. Finally, the average S$_{cell}$ values of the X$_3$r strains at the desired inductions were normalized by the S$_{cell}$ value of a reference culture (MG1655-Z1 bearing the BBa_J107029 constitutive RFP expression cassette in the pSB4C5 plasmid) as internal control to obtain S$_{cell}$ in standardized relative units.

The resulting dose–response curves were fitted using the MATLAB *lsqnonlin* routine to estimate Hill function parameters. For each HSL and ATc condition, at least three independent experiments were carried out.

3. Results

3.1. Simulated Output for Different Model Assumptions

The effect of different model assumptions on the transfer functions shape was studied via numerical simulations. The considered assumptions were on LuxR abundance (Section 3.1.1), cell load (Section 3.1.2) and LuxR-HSL binding mechanism (Section 3.1.3). We evaluated if such assumptions exert a relevant contribution to output variation and, in some cases, if their inclusion contributes to the superior descriptive power of preliminary experimental data.

3.1.1. Effect of LuxR Abundance Assumption: M1 vs. M2

We compared the simulated outputs of M1 and M2, for different DNA/protein copy number situations, to evaluate the effect of the assumption of LuxR concentration abundance over P$_{lux}$ (Figure 2).

As expected from Equations (25) and (26) and previous works [20], for different values of R_{2T} (equal to $n_2 \times r_T$) the output curves generated by M1 showed diverse maximum (α) and switch point (κ) values, increasing and decreasing respectively as a function of R_{2T} (Figure 2a,c,e). On the other hand, for increasing values of plasmid copy number (n_1), the values of α showed a linear increase and κ remained constant (Figure 2b,d,f) as expected from Equation (24). In both cases, the Hill

coefficient was always equal to 1 as expected (Figure 2g,h). While the M1 model was able to capture the effects of LuxR level variation in some experimentally tested model systems, as previously demonstrated [20], the effect of changes in plasmid copy number are intuitively non-realistic since M1 assumes that the concentration of LuxR is much higher than the one of the promoter, thereby resulting in an unlimited increase of protein synthesis rate for high DNA copy numbers, with unchanged shape of the curve in terms of κ and η. Since LuxR may be expressed over a wide range of values to tune the sensitivity of the inducible device [17], the removal of its abundance assumption can be of interest and led us to develop M2.

The M2 model was investigated in the same situations considered for M1 (Figure 2). Upon changes of LuxR values, the relationship between R_{2T} and α or κ were qualitatively analogous to M1. However, when R2T decreases below the promoter concentration value (about 8 nM) the output of M2 showed both lower α and κ levels compared to M1, with the decrease of κ showing the highest-entity effect (>2-fold with the used parameters, Figure 2e). When plasmid copy number is varied in a range below a constant LuxR concentration value (500 nM), α still showed a linear increase as observed in M1, but, differently from M1, κ showed a low-entity increase (less than 2-fold). For concentrations of promoter higher than LuxR, the α value showed saturation, intuitively because all the P$_{lux}$ promoters in the cell cannot be occupied by the limiting amount of LuxR, and as a result RFP synthesis cannot increase anymore for higher concentrations of promoter. Although in this latter condition RFP maximum expression is constant, an increase in plasmid copy number results in a decrease of κ (about 2-fold). Also, in the M2 model, the Hill coefficient was equal to 1 upon R_{2T} variation. However, interestingly, it decreased to values slightly lower than 1 (up to 0.8) upon variations of plasmid copy number (Figure 2g,h).

The illustrated results demonstrate that the removal of the LuxR abundance assumption can affect all the parameters of a Hill function, even when R_{2T} and P_T are tuned over ranges of values not violating this assumption.

3.1.2. Effects of Cell Load

The M1L model was simulated to investigate the effects of cell load, which was assumed to derive from RFP expression alone, or from both RFP and a constant load outside the inducible circuit, caused by the expression of another heterologous protein (Figure 3).

If the load was caused by RFP, its expression affected both LuxR and RFP itself when induction levels became high upon HSL addition. As expected, the maximum level of RFP expression reached by M1L was lower than the respective levels reached by M1 (Figure 3c,d) upon changes of both R_{2T} and plasmid copy number. In particular, the increase of DNA copy number resulted in a saturated maximum RFP expression, which was much lower than in the no-burden model (Figure 3d). The burden effect on κ resulted in a slight decrease compared with M1 upon R_{2T} variation (Figure 3e), while the decreasing effect on κ was more relevant upon plasmid copy number increase (Figure 3f).
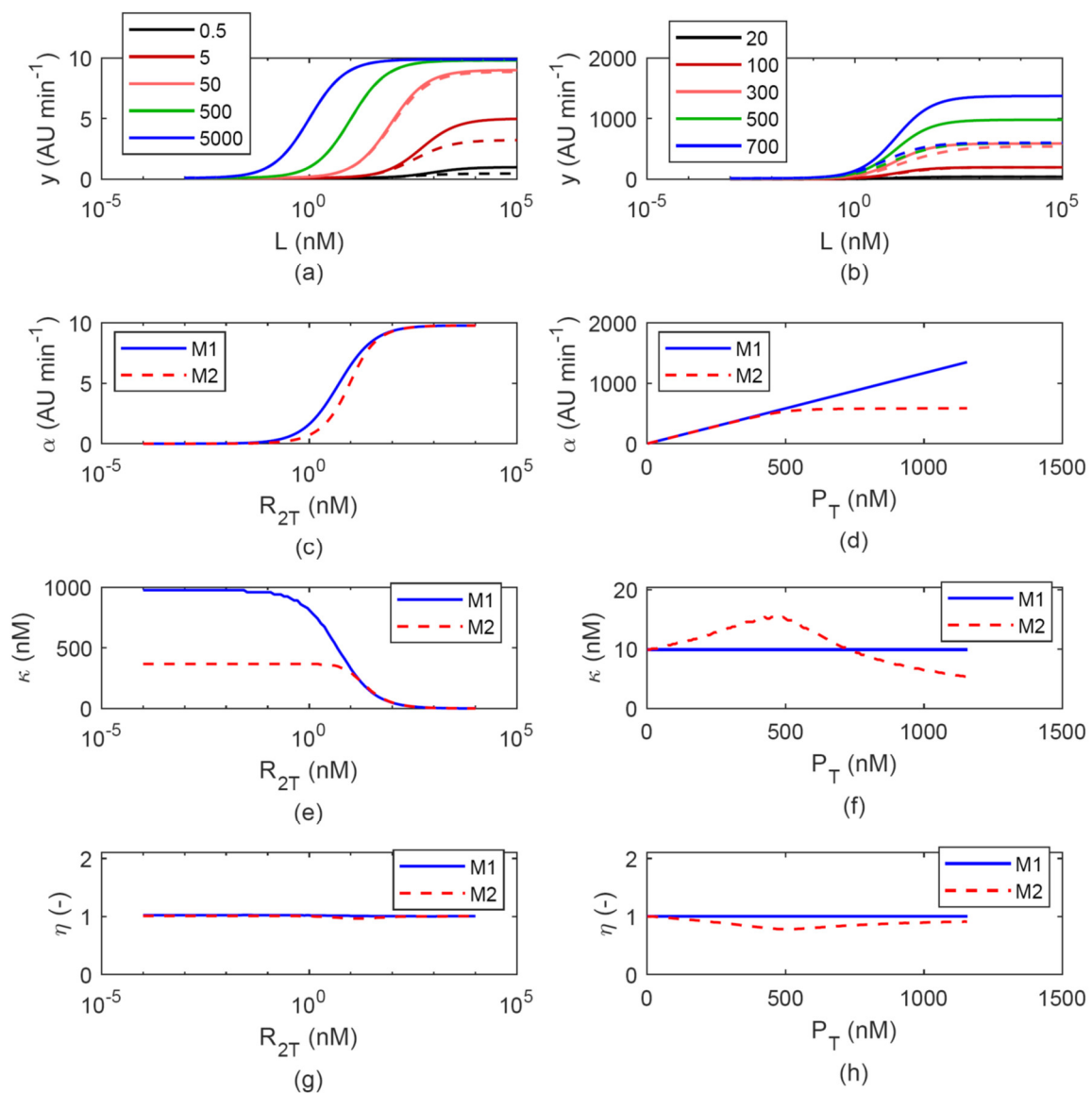
Figure 2. Comparison between M1 and M2. Panels (**a**) and (**b**) report the output activation curves of M1 (solid lines) and M2 (dashed lines) for different values of R_{2T} (panel (a), as indicated in the legend, expressed as nM) and P_T (panel (b), as indicated in the legend, expressed as per cell copy number). The (**c**)–(**d**), (**e**)–(**f**), and (**g**)–(**h**) panel pairs report the R_{2T}- and P_T-dependent trend of α, κ and η, respectively, with solid and dashed lines representing M1 and M2, respectively. In panels (a),(c),(e),(g) the promoter copy number value was set to 5, while in panels (b),(d),(f),(h) the LuxR concentration was set to 500 nM.

Assuming that a constant load (E) also affects protein expression, α showed a further decrease, and κ showed an increase compared with M1, that is, activation curves showed a systematically higher switch point than in a situation without load.

In all the described cases, the Hill coefficient showed values slightly lower than 1 upon plasmid copy number increase and almost unchanged values (close to 1) in case of R_{2T} variation.
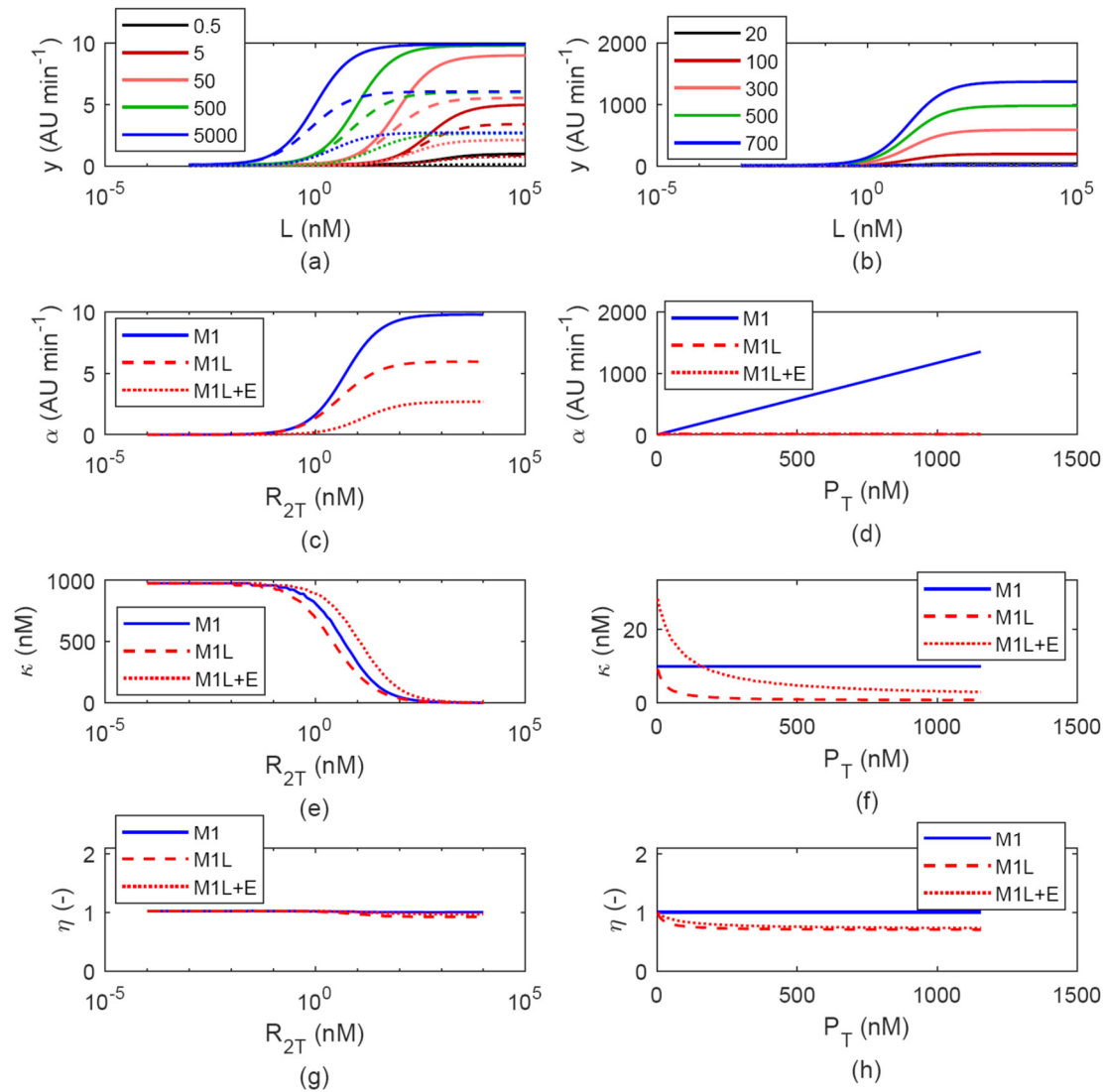
Figure 3. Comparison between M1 and M1L. Panels (**a**) and (**b**) report the output activation curves of M1 (solid lines) and M1L (dashed lines) and M1L with external load, indicated as M1L+E (dotted lines), for different values of R_{2T} (panel (a), as indicated in the legend, expressed as nM) and P_T (panel (b), as indicated in the legend, expressed as per cell copy number). The (**c**)–(**d**), (**e**)–(**f**), and (**g**)–(**h**) panel pairs report the R_{2T}- and P_T-dependent trend of α, κ and η, respectively, with solid, dashed, and dotted lines representing M1, M1L, and M1L+E, respectively. In panels (a),(c),(e),(g) the promoter copy number value was set to 5, while in panels (b),(d),(f),(h) the LuxR concentration was set to 500 nM.

We then analyzed the in vivo data from a previously published experiment, in which the activation curve of a medium copy plasmid-borne lux inducible device was characterized in absence and presence of a second co-transformed plasmid, considered as an additional constant load for the host [31]. Transfer functions were measured in four conditions (with/without load, low/high expression of IPTG-driven LuxR) and a summary of the resulting Hill function parameters is reported in Table 2. As the M1L model predicts, the additional load affects the transfer function parameters upon both low and high LuxR expression level conditions: α showed a decrease in presence of E, while κ increased. It is worth noting that a simpler model including cell load only on RFP expression (and not on LuxR), as previously adopted to analyze in vivo data [23], fails to describe the joint increase of α and decrease of κ upon LuxR overexpression, captured by the M1L model and observed in experimental data resembling the modeled situation (data not shown).

Analogous results were obtained from the simulations via M2L (data not shown).

The reported simulations showed that cell load can quantitatively affect all the Hill parameters of the activation curves, thereby demonstrating the relevance of model assumptions in the analysis of dose–response curves of inducible systems.

Table 2. Parameters estimated from in vivo experiments. Measurements were carried out using MG1655-Z1 as host strain.

Condition	α (%) [a]	κ (nM)	η (-)	Reference
Medium copy [b], no IPTG	47	194.01	1.01	[31]
Medium copy + E [c], no IPTG	27	474.1	0.98	[31]
Medium copy, IPTG = 500 μM	100	0.77	1.54	[31]
Medium copy + E, IPTG = 500 μM	68	0.99	1.29	[31]
Low copy [d], no ATc	83	34.29	0.98	This study
Low copy, ATc = 2.5 ng/ml	92	11.32	1	This study
Low copy, ATc = 5 ng/ml	97	3.73	1.15	This study
Low copy, ATc = 50 ng/ml	100	1.52	1.39	This study

[a] Percent expression, relative to the maximum expression value obtained in the same study; [b] BioBrick™ construct BBa_J107063 in the pSB3K3 medium-copy vector, IPTG-inducible LuxR expression cassette; [c] Additional load provided by the OL1 low-copy plasmid, described in the same paper; [d] BioBrick™ construct BBa_J107032 in the pSB4C5 low-copy vector, ATc-inducible LuxR expression cassette.

3.1.3. Evaluation of LuxR-HSL Complex Formation Assumptions: M(1-2) vs. M(1-2)T

By assuming that the LuxR dimer has two binding sites for HSL molecule, we defined models including a hetero-tetramer formation step (M1T and M2T, depending on the LuxR abundance assumption as above). We also assumed a non-cooperative behavior for the HSL ligand binding, i.e., the two successive HSL binding events occur with the same probability and $K_4 = K_1/4$, as described in Equations (34)–(36). Considering the Adair equation (Equation (48)), describing the fraction (F) of HSL-bound sites of LuxR over the total number of sites [40], the resulting Hill coefficient (computed as in Equation (47)) is always 1 under the non-cooperativity assumption, for any parameter and R_{2T} value

$$F = \frac{Q + 2 \times Q_2}{2 \times R_2 + 2 \times Q + 2 \times Q_2} = \frac{1}{2} \times \frac{K_1 \times L + 0.5 \times K_1^2 \times L^2}{1 + K_1 \times L + 0.25 \times K_1^2 \times L^2} \tag{48}$$

Differently from Equation (48), the expressions of Q_2 (which can be calculated from Equations (17),(31) and (33)) and y show LuxR level-dependent Hill coefficients. For both Q_2 and y, such features can be observed in the closed-form expressions which can be obtained from M1T (Section 2.2.4). In particular, model simulations of M1T showed that the Hill coefficient increases as a function of R_{2T}, which represented the main difference from the respective model without hetero-tetramerization assumption (M1), while the other parameters showed analogous trends as above (Figure 4). The simulation of M2T also showed this feature on the Hill coefficient, together with the same trends described in Section 3.1.1 due to the removal of LuxR abundance assumption.
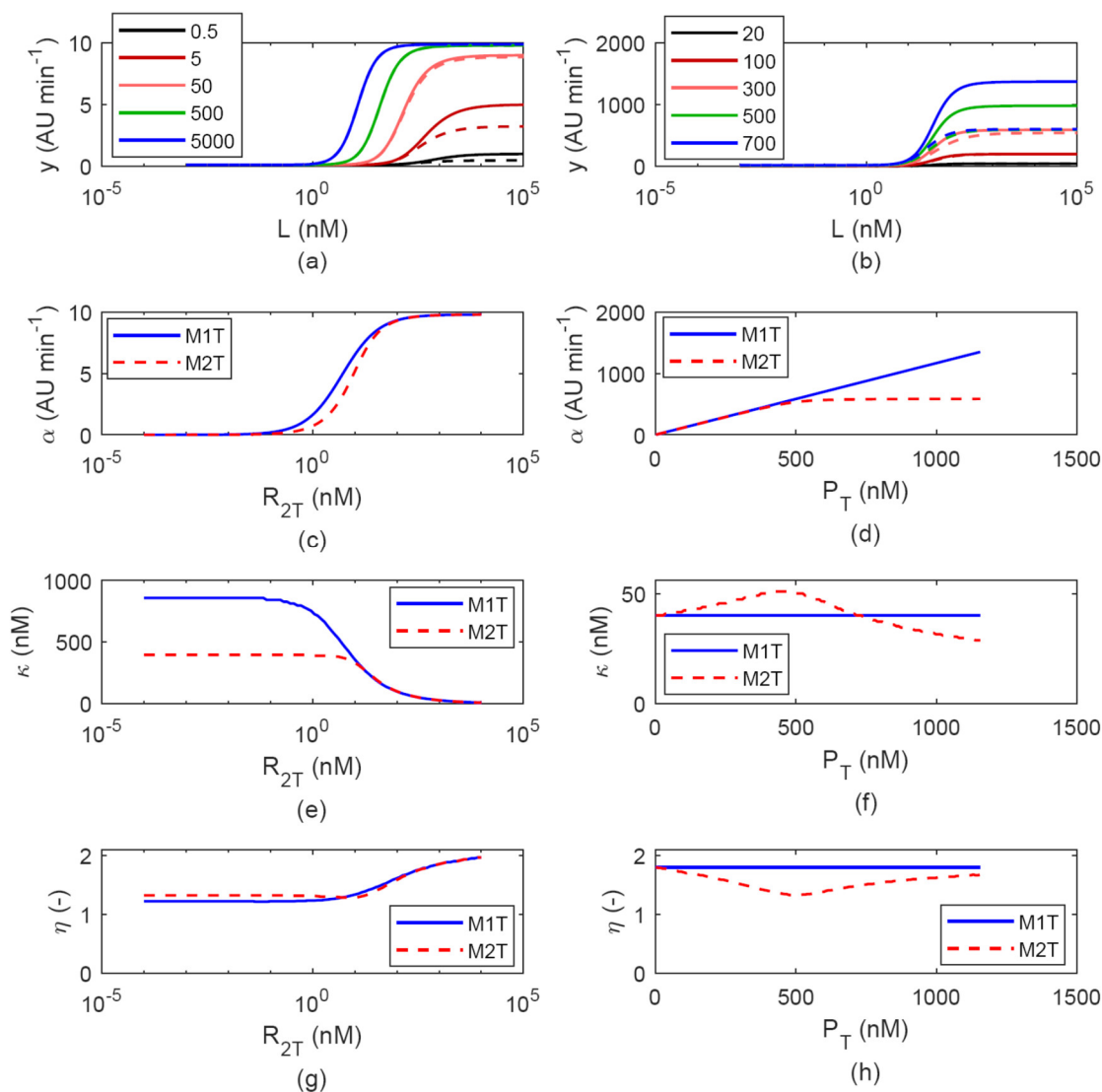
Figure 4. Comparison between M1T and M2T. Panels (**a**) and (**b**) report the output activation curves of M1T (solid lines) and M2T (dashed lines) for different values of R_{2T} (panel (a), as indicated in the legend, expressed as nM) and P_T (panel (b), as indicated in the legend, expressed as per cell copy number). The (**c**)–(**d**), (**e**)–(**f**), and (**g**)–(**h**) panel pairs report the R_{2T}- and P_T-dependent trend of α, κ and η, respectively, with solid and dashed lines representing M1T and M2T, respectively. In panels (a),(c),(e),(g) the promoter copy number value was set to 5, while in panels (b),(d),(f),(h) the LuxR concentration was set to 500 nM.

Although a number of previous experimental works on the lux system showed a Hill coefficient around 1 or slightly lower in the tested conditions [20,33,37], in some works a higher number was reported [31,41]. While Hill coefficient values lower than 1 could be due to burden effects and/or violation of the LuxR abundance assumption, as described by the models illustrated in Sections 3.1.1 and 3.1.2, higher values could not be described by those models in any tested case. Using preliminary data from a previous study [31] and a novel ad-hoc experiment (Table 2), we showed that the Hill coefficient increases upon increase of LuxR level, consistently with the simulations of M1T and M2T.

By also considering cell load in these more detailed models (obtaining the M1TL and M2TL models), the same trends and conclusions illustrated in Section 3.1.2 could be observed (data not shown). Importantly, the resulting Hill coefficient value, as well as the other parameter values, could be tuned by the joint contribution of different effects, e.g., cell load, LuxR abundance assumption

violation, and hetero-tetramerization, with the latter exerting an increase of η for increasing R_{2T} levels, and the other effects causing a decrease of its value.

3.2. Model Identifiability

The usability of the M1 (Section 3.2.1), M2 (Section 3.2.2), M1T (Section 3.2.3), and M2T (Section 3.2.4) models was evaluated by studying their structural and practical identifiability, to eventually understand if their parameters can be reliably estimated and which experiments are needed for this task.

3.2.1. M1

The M1 model is a priori identifiable: two experiments, in which the transfer function is measured for different LuxR levels (tuned by n_2), are needed to properly estimate model parameters in Equation (23). In particular, assuming that the Michaelis–Menten function parameters defined in Equations (24)–(26) are known from the fitting of experimental data, the following expressions can be written (Equations (49)–(54))

$$\bar{\delta} = P_U \times n_1 \times \hat{k}_{m0} \tag{49}$$

$$\bar{\alpha} = \frac{P_U \times n_1 \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)}{1 + 1/\left(\hat{K}_3 \times \overline{n_2}\right)} \tag{50}$$

$$\bar{\kappa} = 1/\left(K_1 \times \left(1 + \hat{K}_3 \times \overline{n_2}\right)\right) \tag{51}$$

$$\bar{\bar{\delta}} = P_U \times n_1 \times \hat{k}_{m0} \tag{52}$$

$$\bar{\bar{\alpha}} = \frac{P_U \times n_1 \times \left(\hat{k}_{m0} + \hat{k}_{mL}\right)}{1 + 1/\left(\hat{K}_3 \times \overline{\overline{n_2}}\right)} \tag{53}$$

$$\bar{\bar{\kappa}} = 1/\left(K_1 \times \left(1 + \hat{K}_3 \times \overline{\overline{n_2}}\right)\right) \tag{54}$$

where the single and double bars on Hill function parameters and copy numbers indicate the parameters of the first and second experiment, respectively. These expressions demonstrate that LuxR must not be very low or very high (compared with $1/\hat{K}_3$) in both experiments to enable identifiability ($n_2 << 1/\hat{K}_3$ or $n_2 >> 1/\hat{K}_3$). These expressions also demonstrate that two or more experiments in which n_1 changes (while keeping n_2 constant) do not lead to a structurally identifiable model.

The a posteriori identifiability was also confirmed by fitting synthetic data generated from parameters in Table 1 (for n_2 = 50 and 500 AU, and r_T = 1) and a realistic number of data points. As expected, this model is practically identifiable since it led to reasonably low estimation errors and CV for its four unknown parameters (Figure 5).

3.2.2. M2

The a priori identifiability of the M2 model could not be studied due to its complex expression. Only a posteriori identifiability could be addressed. Since M2 has the same parameters as M1 with the addition of r_T (which cannot be estimated separately from K_3 in M1), its proper estimation in M2 is intuitively possible only when LuxR is not abundant compared with promoter concentration, otherwise the M2 expression would become identical to M1. Accordingly, the synthetic experiments were simulated with n_2 = 0.005, 0.5, and 5 AU (with r_T = 100 nM/AU). The simultaneous estimation of its five parameters did not lead to a structurally identifiable model due to the high REE and CV, even by considering more activation curves with different LuxR levels (e.g., n_2 = 0.05 which was added to the ones above—data not shown). For this reason, we investigated a two-stage procedure in which synthetic data obtained by setting n_2 = 0.5 and 5 AU (corresponding to R_{2T} = 50 and 500 nM) were fitted with M1 (first stage). This fitting is expected to provide reliable estimates (as shown in Section 3.2.1) because LuxR is highly abundant and the assumptions of M1 are not violated. Since

M1 and M2 share the same $\hat{k}_{m0}$, $\hat{k}_{mL}$ and K_1 parameters, their estimated values were fixed in the second stage, in which the M2 model was used to fit the data with n₂ = 0.005 (corresponding to R_{2T} = 0.5 nM) and, as before, 0.5 and 5 AU, to estimate K_3 and r_T, exploiting at least one condition in which LuxR is not abundant compared with the P_lux promoter. This procedure led to a structurally identifiable condition for the M2 model, since it could estimate its parameters with reasonably low estimation error and CV, even if it had higher REE compared with M1 but much lower CV (Figure 5).
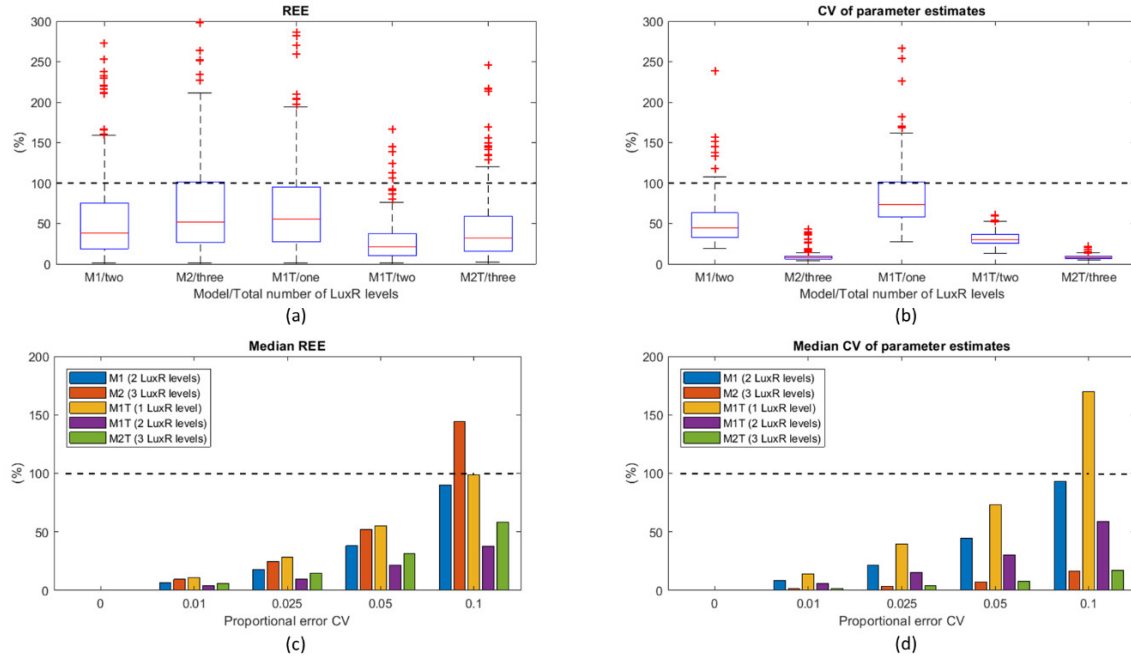
Figure 5. Relative estimation error (REE) and uncertainty of parameter estimates (CV) in a posteriori identifiability. Panels (**a**) and (**b**) report the distribution of REE (a) and CV (b) among 200 runs starting from different simulated data, with a 5% proportional error. The model and the number of LuxR levels included in the fitted data are specified for each boxplot. The red line represents the median of the distribution. Panels (**c**) and (**d**) report the median of the REE (c) and CV (d) distribution as a function of the proportional error entity, from 0 to 10%. The models and specific LuxR levels are described in the legend. In all the panels, the dashed horizontal line indicates the 100% value to facilitate the interpretation of the graphs.

3.2.3. M1T

The M1T model is a priori identifiable from only one experiment. In particular, assuming that the coefficients of the rational function in Equation (37) are known (Equation (55)), the Equations (56)–(60) system can be solved, with the only constraint that no solution is obtained for $n_2 << 1/\overline{K_5}$

$$y = \frac{a' + b' \times L + c' \times L^2}{1 + d' \times L + e' \times L^2} \tag{55}$$

$$a' = P_U \times n_1 \times \hat{k}_{m0} \tag{56}$$

$$b' = P_U \times n_1 \times \hat{k}_{m0} \times K_1 \tag{57}$$

$$c' = P_U \times n_1 \times \hat{k}_{m0} \times K_1^2/4 + P_U \times n_1 \times \hat{k}_{mL} \times K_1^2/4 \times \hat{K}_5 \times n_2 \tag{58}$$

$$d' = K_1 \tag{59}$$

$$e' = K_1^2/4 + K_1^2/4 \times \hat{K}_5 \times n_2 \tag{60}$$

The a posteriori identifiability was investigated by using only one activation curve with n_2 = 50 AU (with r_T = 1). Despite the identifiability was successfully confirmed, the REE and CV were both higher than in M1 (Figure 5). We also investigated the availability of a second activation curve with

n_2 = 500 AU, as in the case of M1. In this case, REE and CV (Figure 5) were systematically lower than in M1T with one LuxR level and also in M1.

3.2.4. M2T

Since the M2T model structure did not enable the study of a priori identifiability, only the a posteriori one was investigated. Similar consideration to M2 also persist for M2T when different activation curves were fitted with this model to estimate its five parameters, leading to an a posteriori non-identifiability (data not shown). For this reason, the same two-stage identification procedure described in Section 3.2.2 was adopted, with the exception that the first estimation step was carried out via M1T to estimate its four parameters, and then M2T was used to estimate K_5 and r_T from the synthetic data as described in Section 3.2.2 by fixing the $\hat{k}_{m0}$, $\hat{k}_{mL}$, and K_1 parameters to the previously estimated values. Results showed that, following the described procedure, the model is a posteriori identifiable with low REE and CV (Figure 5).

4. Discussion

Accurate predictive mathematical models are needed to support the bottom-up design of complex biological systems in synthetic biology. In an effort towards the development of computational tools to overcome this need, different mechanistic models were herein proposed for the lux inducible system. Their mechanistic structure was expected to increase the details of the described system, thereby making the model more generalizable to context variations than traditional empirical equations.

Eight different mechanistic models, based on different assumptions on regulatory protein abundance, ligand binding, and resource usage (and their combinations), were studied in silico and compared, with the final goal of understanding the impact of the underlying assumptions on the transfer function of the circuit. One of the models was strongly inspired by a previously published work [20], while the others represent novel computational tools.

The different assumptions summarized above affected the simulated output of the models (i.e., dose–response curve of recombinant protein production as a function of inducer concentration), thereby demonstrating that significant diversity in model output could be generated by different features. The initially considered model (M1) assumed that one HSL binding event to the LuxR dimer occurred for the activation of the complex, and that LuxR concentration was much higher than the one of P$_{lux}$, with a resulting mathematical expression equivalent to a Michaelis–Menten function model.

When the assumption on the LuxR abundance was removed (M2), the model gave different outputs compared with M1 when the assumption was violated. In particular, the increase of P$_{lux}$ copy number was predicted to result into a linearly increasing expression by M1, while M2 showed a saturating trend, thereby demonstrating that the removal of this simplifying assumption can lead to an intuitively more realistic behavior of the inducible system upon DNA copy number changes. In this situation and with the structural parameter values used in our study, the output showed maximum activity variations up to 2.3-fold between M1 and M2 for biologically plausible values of DNA and protein concentrations.

When a limited resource framework was assumed (M1L), the expression of LuxR and output protein were globally affected by cell load due to the expressed output protein (in an HSL-dependent fashion) and/or to a load outside the inducible circuitry. As previously demonstrated in vivo and *in silico*, the output of inducible systems is affected by the resource usage of the expressed proteins [23]. In the M1L model, we showed that an external load causes a decrease in maximum output and an increase of the half-maximum HSL concentration. This result could not be predicted by some of the traditionally used models in which the transcriptional regulator is assumed to be constant and is not modeled [23,34,42]. In this situation and with the structural parameter values used in our study, the output showed parameter variations up to 2.5-fold between M1 and M1L for biologically plausible values of resource usage parameters, and DNA and protein concentrations. The predicted effects were consistent with previously published experimental data of our group, in which the same

inducible system with and without cell load (due to the presence of an additional co-transformed plasmid) showed the same effect on maximum activity and activation curve sensitivity [31].

In both M2 and M1L models, we also showed that the Hill coefficient of the output curve could decrease (compared to 1, i.e., the one of M1) up to 0.8 with the parameters used in our study.

When a different ligand binding reaction was considered (M1T), i.e., two HSL molecule binding to the LuxR dimer non-cooperatively, the output expression introduced a power over the HSL concentration term. As a result, output curves had a Hill coefficient greater than one, even if the described binding mechanism was assumed to be non-cooperative. To our knowledge, this is the first study explicitly highlighting that a Hill coefficient greater than one could occur in absence of cooperative binding in the transcriptional activator. This effect was consistent with previously published experimental data from our group [31] and others [41], as well as novel preliminary experimental data explicitly measured in this work: the Hill coefficient of the output curve increased as a function of LuxR level. In our M1T model, an increase up to 2-fold was observed for the Hill coefficient compared to M1, which relied on a different assumption on HSL binding.

The behavior of the remaining models, including combinations of the described assumptions (M2L, M2T, M1TL, M2TL), showed more complex features, but similar conclusions on the effects of the investigated assumptions could be drawn.

The M1, M2, M1T, and M2T models were also studied in terms of usability, by investigating their identifiability, to eventually understand if their parameters could be estimated from experimental data and which experimental design is recommended. In fact, in addition to simulation, parameter estimation is a crucial step in model usability that enables the re-use of well-characterized regulatory components in synthetic biology. All the models enabled parameter estimation with reasonably low error and a few constraints (described in Section 3). In general, as expected, by increasing the random noise affecting experimental data REE and CV increase. As for the individual models, M1 required two experiments with different LuxR levels, while M1T required only one experiment with a single LuxR expression value, despite its parameters could be estimated much more reliably by adding a second experiment, like in the M1 case. The M2 and M2T models required an additional experiment, compared to M1 and M1T, respectively, to be properly identified, in which a curve obtained with a LuxR level that is not much higher than P$_{lux}$ concentration had to be measured. Considering a proportional error model for the generated data (5% CV), all the models enabled the estimation of structural parameters within 2-fold compared with the true value (considering interquartile ranges of REE distribution). The M2 model showed the highest estimation error (>100%, median among 200 runs with different datasets simulated with a 10% proportional error), making it the less robust model among the tested ones in estimation tasks. The M1T model with a single LuxR level showed similar drawbacks, which could be overcome by adding activation curve data with more LuxR levels, while M2 could not be improved following the same procedure.

If used to fit the data in real experimental works, a major advantage of the M2 and M2T models is that they potentially enable the estimation of the actual LuxR intracellular concentration, which could not be estimated with the LuxR abundance assumption (M1 and M1T). However, the simultaneous estimation of all the parameters of M2 and M2T failed, thereby leading to the definition of an alternative procedure for the identification of such models: first, M1 (or M1T) was identified by fitting two activation curves data that conformed to the LuxR abundance assumption to estimate all model parameters; then, M2 (or M2T) was identified by fixing three parameters, previously estimated via M1 (or M1T), and estimating the two remaining ones by fitting the two activation curves data used in the first stage, together with an additional curve obtained for a low LuxR level. Given the same number of experiments, M1T and M2T could be identified with lower estimation error and parameter uncertainty than M1 and M2, respectively. The identifiability of models including cell load was not herein addressed, but their usability with real data was investigated previously [23]; they are not expected to include additional identifiability issues, since resource usage parameters could be estimated separately if required [23,43] and their identification can rely on a second output of the system, i.e., a cell burden monitor which acts as a proxy of cell load.

A number of limitations may affect the usability and predictive performance of the studied models against in vivo data. In fact, although mechanistic details have been herein added to traditional models to improve their generalization performance, other assumptions may still be inaccurate in capturing the real behavior of a synthetic circuit. Among the potential crucial aspects, it is worth noting that cell systems are inherently stochastic and when the reacting molecules are present at small intracellular copy numbers stochasticity can result in large fluctuations in the behavior of single cells in a population. In addition, the non-cooperativity of the inducer-regulator binding, herein assumed, should be further investigated. More in general, despite preliminary data have been used to confirm some of the transfer function variations found in this work, a larger-scale experimental validation should be required to investigate and confirm the described effects. Such effort should experimentally validate the impact of the individual assumptions. The different LuxR-HSL binding assumptions will need such validation to select the one best describing experimental measures. However, the validation of the other assumptions (LuxR abundance and cell load) should not lead to the selection of a best-performing model since it is expected to be application-specific, e.g., a no-burden model may have high predictive power if all the genes of a circuit have low resource usage.

In summary, we have proposed different usable mechanistic models that had a significant impact on the predicted output of an inducible system, and they were also consistent with a set of preliminary experimental data. In the future, the reported models may support synthetic circuits output prediction in practical situations with unprecedented details, also facilitating the bottom-up design of complex circuits due to their generalization power. In this framework, highly relevant applications of such models in synthetic biology are, e.g., the prediction of circuit output as a function of unseen DNA, protein and inducer concentrations; estimation of protein regulator abundance; investigation of different binding mechanisms for subsequent model selection.

The proposed model definition and analytic procedures may be used to study other systems different from the LuxR/P$_{lux}$ module, considered in this work. The underlying binding reactions could be known, or they might be investigated by testing different model assumptions against experimental data for model parametrization and mechanistic understanding.

Author Contributions: Conceptualization, L.P. and P.M.; Formal analysis, L.P. and P.M.; Investigation, L.P., M.B., D.D.M., and P.M.; Methodology, L.P.and P.M.; Writing—Original Draft, L.P.; Writing—Review & Editing, M.B. and P.M.

References

1. Peccoud, J. Synthetic Biology: Fostering the cyber-biological revolution. *Synth. Biol.* **2016**, *1*, ysw001.
2. Kitney, R.I.; Freemont, P.S. Engineering biology: A key driver of the bio-economy. *IET: Eng. Biol.* **2017**, *1*, 3–6.
3. Del Vecchio, D.; Qian, Y.; Murray, R.M.; Sontag, E.D. Future systems and control research in synthetic biology. *Annu. Rev. Control* **2018**, *45*, 5–17.
4. Schaerli, Y.; Isalan, M. Building synthetic gene circuits from combinatorial libraries: Screening and selection strategies. *Mol. Biosyst.* **2013**, *9*, 1559–1567.
5. Engstrom, M.D.; Pfleger, B.F. Transcription control engineering and applications in synthetic biology. *Synth. Syst. Biotechnol.* **2017**, *2*, 176–191.
6. Pasotti, L.; Zucca, S. Advances and computational tools towards predictable design in biological engineering. *Comput. Math. Methods Med.* **2014**, *2014*, 369681, doi:10.1155/2014/369681.
7. Way, J.C.; Collins, J.J.; Keasling, J.D.; Silver, P.A. Integrating biological redesign: Where synthetic biology came from and where it needs to go. *Cell* **2014**, *157*, 151–161.

8. Zhang, F.; Carothers, M.; Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **2012**, *30*, 354–359.

9. Lo, T.M.; Chng, S.H.; Teo, W.S.; Cho, H.S.; Chang, M.W. A two-layer gene circuit for decoupling cell growth from metabolite production. *Cell Syst.* **2016**, *3*, 133–143.

10. Dragosits, M.; Nicklas, D.; Tagkopoulos, I. A synthetic biology approach to self-regulatory recombinant protein production in *Escherichia coli*. *J. Biol. Eng.* **2012**, *6*, 2.

11. Carr, S.B.; Beal, J.; Densmore, D.M. Reducing DNA context dependence in bacterial promoters. *PLoS ONE* **2017**, *12*, e0176013.

12. Qian, Y.; Huang, H.H.; Jiménez, J.I.; Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **2017**, *6*, 1263–1272.

13. Bandiera, L.; Pasini, A.; Pasotti, L.; Zucca, S.; Mazzini, G.; Magni, P.; Giordano, E.; Furini, S. Experimental measurements and mathematical modeling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits. *J. Theor. Biol.* **2016**, *395*, 153–160.

14. Gyorgy, A.; Jiménez, J.I.; Yazbek, J.; Huang, H.H.; Chung, H.; Weiss, R.; Del Vecchio, D. Isocost lines describe the cellular economy of genetic circuits. *Biophys. J.* **2015**, *109*, 639–646.

15. Ceroni, F.; Algar, R.; Stan, G.B.; Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **2015**, *12*, 415–418.

16. Arkin, A.P. A wise consistency: Engineering biology for conformity, reliability, predictability. *Curr. Opin. Chem. Biol.* **2013**, *17*, 893–901.

17. Ang, J.; Harris, E.; Hussey, B.J.; Kil, R.; McMillen, D.R. Tuning response curves for synthetic biology. *ACS Synth. Biol.* **2013**, *2*, 547–567.

18. Trabelsi, H.; Koch, M.; Faulon, J.L. Building a minimal and generalizable model of transcription factor-based biosensors: Showcasing flavonoids. *Biotechnol. Bioeng.* **2018**, *115*, 2292–2304.

19. Endler, L.; Rodriguez, N.; Juty, N.; Chelliah, V.; Laibe, C.; Li, C.; Le Novere, N. Designing and encoding models for synthetic biology. *J. R. Soc. Interface* **2009**, *6* (Suppl. 4), S405–S417.

20. Carbonell-Ballestero, M.; Duran-Nebreda, S.; Montañez, R.; Solé, R.; Macía, J.; Rodríguez-Caso, C. A bottom-up characterization of transfer functions for synthetic biology designs: Lessons from enzymology. *Nucleic Acids Res.* **2014**, *42*, 14060–14069.

21. Mileyko, Y.; Joh, R.I.; Weitz, J.S. Small-scale copy number variation and large-scale changes in gene expression. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16659–16664.

22. Canton, B.; Labno, A.; Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **2008**, *26*, 787–793.

23. Pasotti, L.; Bellato, M.; Casanova, M.; Zucca, S.; Cusella De Angelis, M.G.; Magni, P. Re-using biological devices: A model-aided analysis of interconnected transcriptional cascades designed from the bottom-up. *J. Biol. Eng.* **2017**, *11*, 50.

24. Gabor, A.; Villaverde, A.F.; Banga, J.R. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Syst. Biol.* **2017**, *11*, 54.

25. Berset, Y.; Merulla, D.; Joublin, A.; Hatzimanikatis, V.; van der Meer, J.R. Mechanistic modeling of genetic circuits for ArsR arsenic regulation. *ACS Synth. Biol.* **2017**, *6*, 862–874.

26. Rydenfelt, M.; Cox, R.S., 3rd; Garcia, H.; Phillips, R. Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. *Phys. Rev. E* **2014**, *89*, 012702.

27. Bintu, L.; Buchler, N.E.; Garcia, H.G.; Gerland, U.; Hwa, T.; Kondev, J.; Kuhlman, T.; Phillips, R. Transcriptional regulation by the numbers: Applications. *Curr. Opin. Genet. Dev.* **2005**, *15*, 125–135.

28. Bintu, L.; Buchler, N.E.; Garcia, H.G.; Gerland, U.; Hwa, T.; Kondev, J.; Phillips, R. Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **2005**, *15*, 116–124.

29. Stevens, A.M.; Greenberg, E.P. Quorum sensing in Vibrio fischeri: Essential elements for activation of the luminescence genes. *J. Bacteriol.* **1997**, *179*, 557–562.

30. Balagadde, F.K.; Song, H.; Ozaki, J.; Collins, C.H.; Barnet, M.; Arnold, F.H.; Quake, S.R.; You, L. A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.* **2008**, *4*, 187.

31. Pasotti, L.; Bellato, M.; Politi, N.; Casanova, M.; Zucca, S.; Cusella De Angelis, M.G.; Magni, P. A synthetic close-loop controller circuit for the regulation of an extracellular molecule by engineered bacteria. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *13*, 248–258.

32. Zucca, S.; Pasotti, L.; Politi, N.; Casanova, M.; Mazzini, G.; Cusella De Angelis, M.G.; Magni, P. Multi-faceted characterization of a novel LuxR-repressible promoter library for *Escherichia coli*. *PLoS ONE* **2015**, *10*, e0126264.

33. Zucca, S.; Pasotti, L.; Mazzini, G.; Cusella De Angelis, M.G.; Magni, P. Characterization of an inducible promoter in different DNA copy number conditions. *BMC Bioinform.* **2012**, *13*, S11.

34. Danino, T.; Mondragon-Palomino, O.; Tsimring, L.; Hasty, J. A synchronized quorum of genetic clocks. *Nature* **2010**, *463*, 326–330.

35. Pasotti, L.; Quattrocelli, M.; Galli, D.; Cusella De Angelis, M.G.; Magni, P. Multiplexing and demultiplexing logic functions for computing signal processing tasks in synthetic biology. *Biotechnol. J.* **2011**, *6*, 784–795.

36. Qian, Y.; Del Vecchio, D. Effective interaction graphs arising from resource limitations in gene networks. In Proceedings of the 2015 American Control Conference (ACC), Chicago, IL, USA, 1–3 July 2015; pp. 4417–4423.

37. Pasotti, L.; Politi, N.; Zucca, S.; Cusella De Angelis, M.G.; Magni, P. Bottom-up engineering of biological systems through standard bricks: A modularity study on basic parts and devices. *PLoS ONE* **2012**, *7*, e39407.

38. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **2013**, *35*, 1050–1055.

39. Magni, P.; Sparacino, G. Parameter estimation. In *Modeling Methodology for Physiology and Medicine*, 2nd ed.; Elsevier: Amsterdam, Netherlands, 2014; pp. 83–110, doi:10.1016/B978-0-12-411557-6.00005-7.

40. Stefan, M.I.; Le Novere, N. Cooperative binding. *PLoS Comput. Biol.* **2013**, *9*, e1003106.

41. Colton, D.M.; Stabb, E.V.; Hagen, S.J. Modeling analysis of signal sensitivity and specificity by Vibrio fischeri LuxR variants. *PLoS ONE* **2015**, *10*, e0126474.

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

43. Borkowski, O.; Bricio, C.; Murgiano, M.; Rothschild-Mancinelli, B.; Stan, G.B.; Ellis, T. Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* **2018**, *9*, 1497.