



RESEARCH

Open Access

# Rational design of modular circuits for gene transcription: A test of the bottom-up approach

Francesca Ceroni<sup>1,2\*</sup>, Simone Furini<sup>3</sup>, Emanuele Giordano<sup>1,4</sup>, Silvio Cavalcanti<sup>1,2†</sup>

## Abstract

**Background:** Most of synthetic circuits developed so far have been designed by an ad hoc approach, using a small number of components (i.e. LacI, TetR) and a trial and error strategy. We are at the point where an increasing number of modular, inter-changeable and well-characterized components is needed to expand the construction of synthetic devices and to allow a rational approach to the design.

**Results:** We used interchangeable modular biological parts to create a set of novel synthetic devices for controlling gene transcription, and we developed a mathematical model of the modular circuits. Model parameters were identified by experimental measurements from a subset of modular combinations. The model revealed an unexpected feature of the lactose repressor system, i.e. a residual binding affinity for the operator site by induced lactose repressor molecules. Once this residual affinity was taken into account, the model properly reproduced the experimental data from the training set. The parameters identified in the training set allowed the prediction of the behavior of networks not included in the identification procedure.

**Conclusions:** This study provides new quantitative evidences that the use of independent and well-characterized biological parts and mathematical modeling, what is called a bottom-up approach to the construction of gene networks, can allow the design of new and different devices re-using the same modular parts.

## Background

Synthetic biology has evolved to the point where the design of gene circuits with complex functionalities has become a real option [1]. Inside living cells, complex behaviors arise from molecular interplays in complicated regulatory networks. In the first instance, the ability to isolate single elements from these regulatory networks - and to use them as independent modules - makes synthetic biology possible [2]. Synthetic gene circuits are thus created by assembling elementary modules together. The increasing complexity of these synthetic gene networks asks for a rational approach to design gene circuits [3]. A possible strategy to tackle this complexity is the *bottom-up* approach [4-7]. In bottom-up design, the behavior of a complicated system is predicted from the characteristics of its elementary parts. Such a prediction requires well-characterized

mathematical models of these elementary parts, and of how they behave when assembled together. In the present study, we tested if it is possible to predict by mathematical modeling the behavior of a modular gene circuit, using as inputs the properties of its elementary parts.

In the last decade, many elementary devices have been implemented both in prokaryotic and in eukaryotic cells [8-11], including logical gates [12], toggle switches [13], oscillators [14,15], band pass filters [14] and counters [16]. However, most of the circuits developed so far have been designed by an *ad hoc* approach, using specific gene components and a trial and error strategy. In order to make the design of synthetic systems easier, great efforts are today directed to extend to synthetic biology the engineering concepts of standardization, modularity, and abstraction [17,18]. In this context, the Registry of Standard Biological Parts <http://www.parts-registry.org>, maintained by the Massachusetts Institute of Technology (MIT), distributes thousands of standard parts, named BioBricks. The BioBricks are only standardized in terms of how these individual parts can be

\* Correspondence: francesca.ceroni2@unibo.it

† Contributed equally

<sup>1</sup>Laboratory of Cellular and Molecular Engineering, University of Bologna, I-47521 Cesena, Italy

Full list of author information is available at the end of the article

physically assembled into more complicated systems, whilst most of them remain not tested in terms of their functionality and biological behavior [19].

The option to use modular parts in large-scale networks will be highly facilitated by a detailed characterization of their functional properties shared by the synthetic biology community. This implies the standardization of tools, techniques, and measurement units used by different laboratories [19-21], and the definition of mathematical models for the single parts. In this direction, a repository of modular modeling components has recently been published to facilitate the mathematical modeling of biological parts, and to help the design process in synthetic biology [22]. Once mathematical models of the elementary components are available, the following step is to use them to predict the behavior of larger-scale networks. Only if the behavior of gene circuits can be predicted from the properties of their elementary parts, a bottom-up design of synthetic devices is feasible [23]. While this remains a central problem of synthetic biology, few quantitative tests of how modular systems behave have been published. One example is the work by Ellis et al [24] where a library of regulated promoters is characterized and used to design more complicated networks with predictable outcomes. This quantitative study supports the case of a bottom-up approach to gene circuit design.

In this study we used, as a prototypal model, a synthetic device based on the lactose repressor system. This synthetic gene network includes: (i) a negative feedback circuit for the synthesis of lactose repressor (LacI) molecules; (ii) and a reporter circuit, for the synthesis of green fluorescent protein (GFP) controlled by LacI (Figure 1). Transcriptional control is realized by regulated promoters, assembled from a constitutive promoter and an operator site docking for LacI. The prokaryotic promoter can be divided into three sub-regions: (i) distal (upstream of the -35 sequence), (ii) core (between -35 and -10) and (iii) proximal (downstream of the -10) [25]. LacI can regulate gene transcription by binding to operator sites placed in any of these regions [26]. However, we decided to insert our operator sites only into the proximal promoter region, avoiding modifications of the integrity of the core region. This choice preserves the basal transcriptional level of the promoter, and most importantly its modularity. The operator sites in the synthetic devices were used as modular elements. Nine different devices were assembled using three operator sites with different binding affinities for LacI, alternatively inserted into two plasmid vectors. Three of these synthetic circuits were used to define the mathematical model, and to identify its parameters. The mathematical model was then used to predict the behavior of the remaining untested

circuits, providing a quantitative test of modular design in synthetic biology.

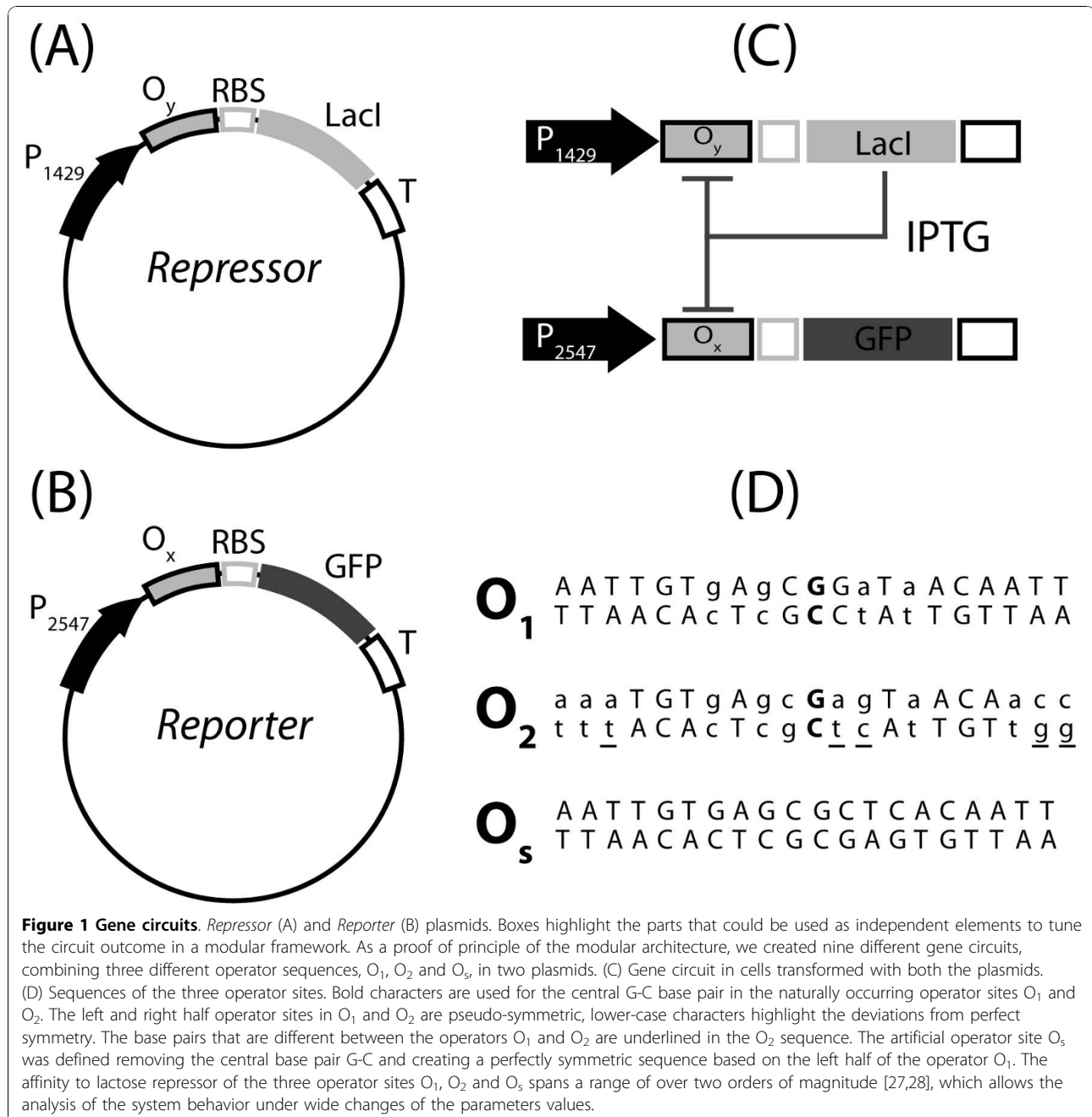
## Results

### Definition of the gene-circuits

*E. coli* DH5 $\alpha$  cells were transformed with two different plasmids, here named *Repressor* and *Reporter* (Figure 1). The gene circuit encoded in the *Repressor* plasmid has a negative feedback structure as an operator site for LacI is inserted downstream of a promoter sequence controlling the expression of the lactose repressor gene itself (Figure 1A). The concentration of LacI in the cytoplasm also controls the transcription of the GFP gene on the *Reporter* gene circuit (Figure 1B). The synthesis of GFP in a cell transformed with both *Repressor* and *Reporter* plasmids can be tuned by changing the concentration of the gratuitous inducer IPTG (Figure 1C). We created nine different gene circuits, using all the possible combinations of three operator sites ( $O_1$ ,  $O_2$  and  $O_s$ , see the sequences in Figure 1D) in the *Repressor* and *Reporter* plasmids. The affinity to lactose repressor of the operator sites  $O_1$ ,  $O_2$  and  $O_s$  spans a range of over two orders of magnitude, which allows the creation of systems with widely different characteristics [27,28]. The symbol  $O_xO_y$  is used to identify the gene circuit with the operator sequence  $O_x$  on the *Reporter* plasmid, and the operator sequence  $O_y$  on the *Repressor* plasmid. If cells are transformed with the *Reporter* plasmid alone, the symbols  $O_1$ ,  $O_2$ , or  $O_s$  are used, according to which operator site is present.

### Test of the modular architecture

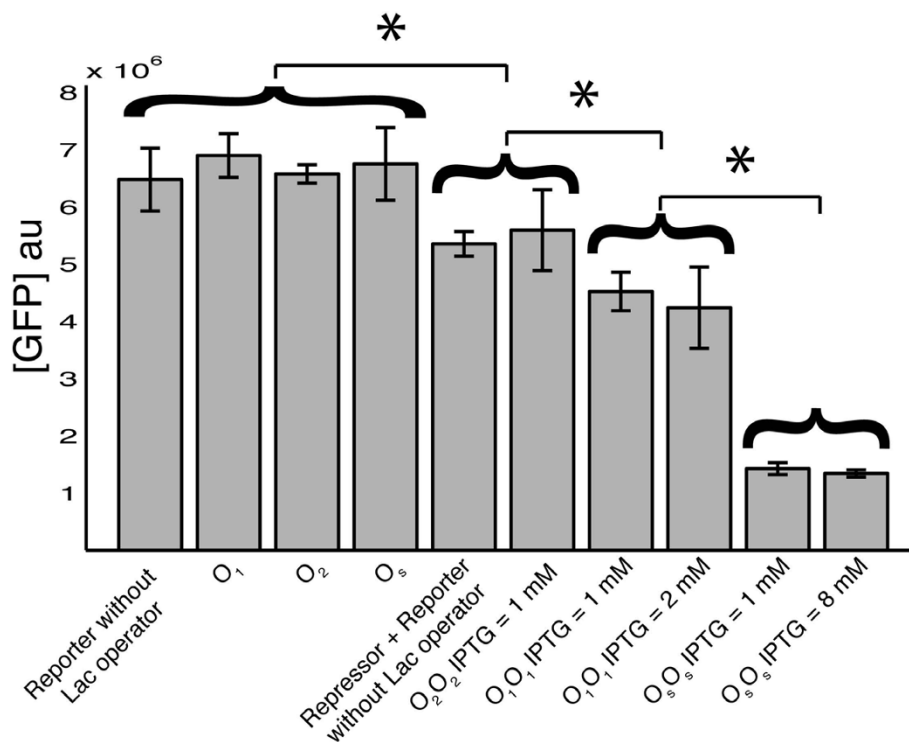
Cells transformed only with *Reporter* plasmids showed equilibrium values of the normalized fluorescence equal to (see the Methods section for the definition of normalized fluorescence,  $[GFP]_{au}$ )  $6.89 \times 10^6$ ,  $6.55 \times 10^6$  and  $6.75 \times 10^6$  arbitrary units (au) for plasmids with the  $O_1$ ,  $O_2$  and  $O_s$  operator sequence respectively (Figure 2). In absence of *Repressor* plasmids, no molecules of LacI are synthesized, thus no LacI-dependent inhibition is exerted on GFP transcription. In accordance, cells transformed with *Reporter* plasmids with different operator sequences yielded analogous fluorescence levels. This result supports the case of a modular architecture of the gene circuit, where promoter and operator sequences can be used as independent modular parts. In agreement, cells transformed only with a *Reporter* plasmid lacking the operator sequence showed equilibrium fluorescence of  $6.47 \times 10^6$  au (Figure 2). This value is statistically equivalent to those measured in cells transformed with a *Reporter* plasmid containing one of the three operator sequences. This suggests that the presence of the operator sequence alone does not affect



transcription rates. Thus, a unique value of the GFP transcription rate,  $\alpha_G^M$ , was used in the mathematical model, independently of the specific operator sequence present in the gene circuit (see the Methods section for the definition of the mathematical model, and Table 1 for the list of the parameters).

Cells transformed with the *Reporter* circuit alone reached the maximum fluorescence level in our experimental conditions. In an ideal modular system, this maximum fluorescence is conserved in cells transformed with both *Reporter* and *Repressor* plasmids, if the

*Reporter* plasmids lack the lactose operator sequence. To test this issue, we co-transformed cells with a *Repressor* plasmid, containing any of the three different operator sites, and with a *Reporter* plasmid lacking the operator site. Cells co-transformed with these two plasmids (*Repressor* plasmid and *Reporter* plasmid lacking the operator site) yielded a fluorescence value of  $5.44 \times 10^6$  au, significantly lower than the maximum fluorescence observed in cells transformed with either of the *Reporter* plasmids alone (Figure 2). The average fluorescence in cells co-transformed with the *Repressor* plasmid



**Figure 2 Fluorescence levels.** The normalized fluorescence, equation 2, in arbitrary units, is shown for cells transformed with: *Reporter* plasmid without the operator sequence (n = 6); *Reporter* plasmids with the O<sub>1</sub> (n = 5), O<sub>2</sub> (n = 8), and O<sub>s</sub> (n = 10) operator sequence; *Repressor* plasmids with any of the three operator sequences and *Reporter* plasmid lacking the lactose operator sequence (n = 10); O<sub>2</sub>O<sub>2</sub> induced with 1 mM IPTG (n = 13); O<sub>1</sub>O<sub>1</sub> induced with 1 mM (n = 9) and 2 mM (n = 5) IPTG; O<sub>s</sub>O<sub>s</sub> induced with 1 mM (n = 4) and 8 mM (n = 3) IPTG. In absence of *Repressor* plasmids, the normalized fluorescence is the same if the *Reporter* plasmid does not include the operator sequence, or if it includes any of the three operator sequences tested. Co-transformation with a *Repressor* plasmid and a *Reporter* plasmid without the operator site causes a decrease in the maximum fluorescence, even if no LacI activity is exerted. The same fluorescence was observed in the O<sub>2</sub>O<sub>2</sub> gene circuit at saturating concentrations of IPTG. On the other hand, the maximum fluorescence in the O<sub>1</sub>O<sub>1</sub> and O<sub>s</sub>O<sub>s</sub> gene circuits is statistically different from the maximum value.

and the *Reporter* plasmid without the operator site, was used to define the maximum fluorescence level of the gene-circuit ( $GFP_{max}$ ). The relative fluorescence,  $GFP_{\%}$ , in the gene circuits O<sub>x</sub>O<sub>y</sub>, was defined as:

$$GFP_{\%} = \frac{[GFP]_{au}}{GFP_{max}} \quad (1)$$

Only in the O<sub>2</sub>O<sub>2</sub> gene circuit a relative fluorescence close to 100% was obtained upon IPTG induction. Indeed, the fluorescence observed in the O<sub>2</sub>O<sub>2</sub> gene circuit in presence of 1 mM IPTG was not significantly different from the maximum fluorescence produced by the gene circuit with the *Repressor* plasmid and the *Reporter* plasmid without the operator site. On the other hand, fluorescence values induced by 1 mM IPTG in the O<sub>1</sub>O<sub>1</sub> and O<sub>s</sub>O<sub>s</sub> gene circuits were statistically different from  $GFP_{max}$ . Increasing the IPTG concentration up to 2 mM in O<sub>2</sub>O<sub>2</sub>, and up to 8 mM in O<sub>s</sub>O<sub>s</sub>, did not change the fluorescence value, suggesting that

the removal of LacI-mediated inhibition of transcription is already saturated at 1 mM IPTG in both the gene circuits (Figure 2).

In absence of IPTG, the gene-circuits O<sub>2</sub>O<sub>2</sub>, O<sub>1</sub>O<sub>1</sub> and O<sub>s</sub>O<sub>s</sub> showed relative inductions of  $2.22 \times 10^{-2}$ ,  $4.62 \times 10^{-3}$  and  $2.85 \times 10^{-3}$  respectively (Figure 3A). The relative inductions of the three gene circuits with the same operator sequence on both plasmids were used to identify the model parameters. A single parameter,  $\alpha_G^M$ , was used in the fitting procedure of the gene circuits O<sub>2</sub>O<sub>2</sub> and O<sub>1</sub>O<sub>1</sub>. The excellent agreement with the experimental data suggests that the mathematical model correctly reproduce the behavior of the gene networks in absence of IPTG (Figure 3A). A second parameter,  $K_s^L$ , was used to fit the data of the O<sub>s</sub>O<sub>s</sub> gene circuit, thus, it is not surprising that the mathematical model well reproduces this data.

To test the validity of the modular approach, selected combinations of *Reporter* and *Repressor* plasmids with different operator sites were tested, namely O<sub>2</sub>O<sub>s</sub>, O<sub>1</sub>O<sub>s</sub>,

**Table 1 Variables and parameters of the mathematical model**

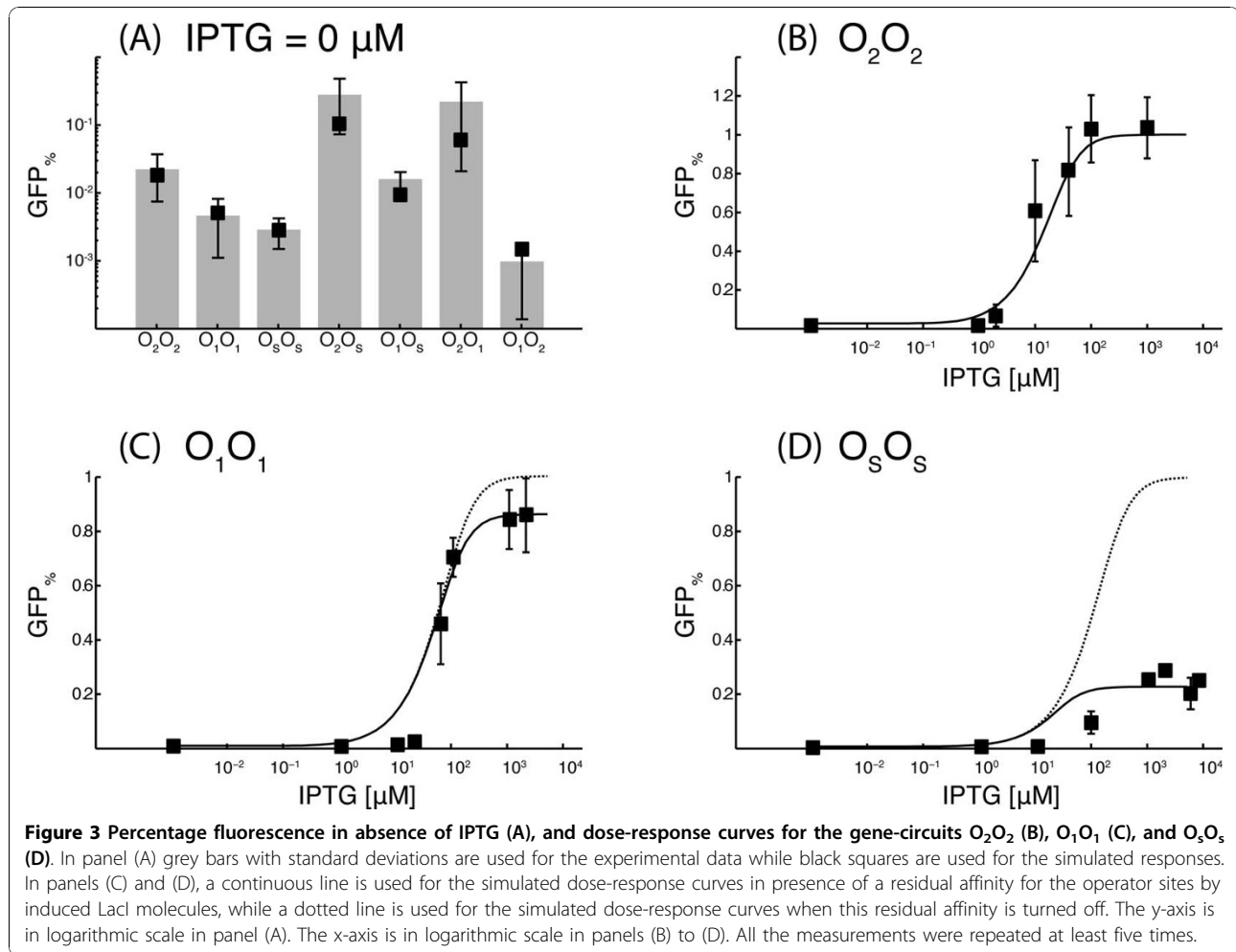
	Description	Value	Units
$I$	IPTG concentration		mM
$G$	Concentration of GFP protein		molecules per cell
$L^F$	Free Lacl molecules, i.e. not bound to operator sites or IPTG molecule		molecules per cell
$L^I$	Lacl molecules bound to IPTG		molecules per cell
$M_G$	mRNA molecules of GFP		molecules per cell
$M_L$	mRNA molecules of Lacl		molecules per cell
$D_{G/L}^F$	Free <i>Repressor/Reporter</i> plasmids		plasmids per cell
$D_{G/L}^L$	<i>Repressor/Reporter</i> plasmids bound to Lacl molecules		plasmids per cell
$D_{G/L}^I$	<i>Repressor/Reporter</i> plasmids bound to induced Lacl molecules		plasmids per cell
$D_G^0$	Number of <i>Reporter</i> plasmids per cell	80	plasmids per cell
$D_L^0$	Number of <i>Repressor</i> plasmids per cell	$D_G^0 / 3.75^{\S} = 21.33$	plasmids per cell
$\lambda_{G/L}$	Protein degradation rate	0.0214 <sup>\S</sup>	min <sup>-1</sup>
$\lambda_{G/L}^M$	mRNA degradation rate	0.271 [39]	min <sup>-1</sup>
$\alpha_G$	GFP rate of synthesis	$540\lambda_G = 11.54$ [40]	min <sup>-1</sup>
$A_L$	Lacl rate of synthesis	$\alpha_G/4 = 2.88$	min <sup>-1</sup>
$\alpha_G^M$	GFP transcription rate	0.56 <sup>\dagger</sup>	min <sup>-1</sup>
$\alpha_L^M$	Lacl transcription rate	$\alpha_G^M / 1.23^{\S} = 0.45$	min <sup>-1</sup>
$K_x^L$	Equilibrium binding constant of the complex Lacl-O <sub>x</sub>	$K_1^L$	0.13 [41] molecules per cell
		$K_2^L$	1.63 [41] molecules per cell
		$K_s^L$	0.0394 <sup>\dagger</sup> molecules per cell
$K_x^I$	Equilibrium binding constant for the binding of induced Lacl molecule to the operator sequence O <sub>x</sub>	$K_1^I$	25336 <sup>\dagger</sup> molecules per cell
		$K_2^I$	$\infty$ molecules per cell
		$K_s^I$	313 <sup>\dagger</sup> molecules per cell
$K^{LI}$	Equilibrium binding constant for the binding IPTG-Lacl	0.2890 <sup>\dagger</sup>	mM
$n$	Cooperativity of the binding Lacl-IPTG	1.8688 <sup>\dagger</sup>	
$\tau^{LI}$	Time constant of Lacl binding to the operator sequences	0.02 [43]	min
$\tau^{DI}$	Time constant of induced-Lacl binding to the operator sequences	*	min
$\tau^{DL}$	Time constant of the binding Lacl-IPTG	*	min

\* Only the steady-state behaviors of the gene circuits are analyzed, thus arbitrary values can be used for these time constants. <sup>\dagger</sup>Values defined through the fitting procedure. <sup>\S</sup>Values obtained by experimental measurements. References are included for the values retrieved from the literature.

O<sub>2</sub>O<sub>1</sub>, and O<sub>1</sub>O<sub>2</sub>. The remaining combinations were discarded because the operator site in the *Reporter* plasmid was too strong compared to the operator site in the *Repressor* plasmid. The mathematical model predicts well the relative inductions of these four gene circuits in absence of IPTG (Figure 3A), with no further adjustment of the model parameters. The predictions of the model are closer than one standard deviation from the experimental data for all the gene circuits.

The dose-response curves of the gene circuits O<sub>2</sub>O<sub>2</sub>, O<sub>1</sub>O<sub>1</sub> and O<sub>s</sub>O<sub>s</sub> were fitted using five parameters. Two parameters,  $K^{LI}$  and  $n$ , describe the binding of IPTG to Lacl and are the same for the three gene circuits. The other parameters,  $K_2^I$ ,  $K_1^I$  and  $K_s^I$  describe the binding of induced Lacl molecules to the respective operator sites. Since the gene circuit O<sub>2</sub>O<sub>2</sub> reaches 100%

induction at saturating concentration of IPTG, the parameter  $K_2^I$  can be considered equal to infinite for any practical reason. On the other hand, the maximum fluorescence in the gene circuit O<sub>1</sub>O<sub>1</sub> is statistically different from  $GFP_{max}$ . In absence of Lacl molecules, gene circuits with the O<sub>1</sub> or O<sub>2</sub> operator sequence on the *Reporter* plasmids showed the same fluorescence. Thus, the diminished maximum fluorescence of the O<sub>1</sub>O<sub>1</sub> gene circuit can only result from a Lacl-dependent impairment of transcription, persisting at saturating concentration of IPTG. A marginal affinity of the induced Lacl,  $L^I$ , to the operator sequence can explain the experimental data of the O<sub>1</sub>O<sub>1</sub> gene circuit. Figure 3C shows the simulated dose-response curve if induced Lacl molecules have (continuous line), or have not (dotted line) a residual affinity for the operator sites.



Introducing the marginal affinity  $K_1^I$  decreases the distance from the experimental data from 0.33 to 0.25. The effect of this marginal affinity is much more evident in the O<sub>s</sub>O<sub>s</sub> gene circuit, where only ~25% of GFP synthesis was achieved at saturating concentrations of IPTG (Figure 3D). According to the estimated parameters, the operator sites O<sub>1</sub> and O<sub>s</sub> bind the free LacI molecules respectively  $\sim 2 \times 10^5$  and  $\sim 1 \times 10^4$  times stronger than the induced LacI molecules.

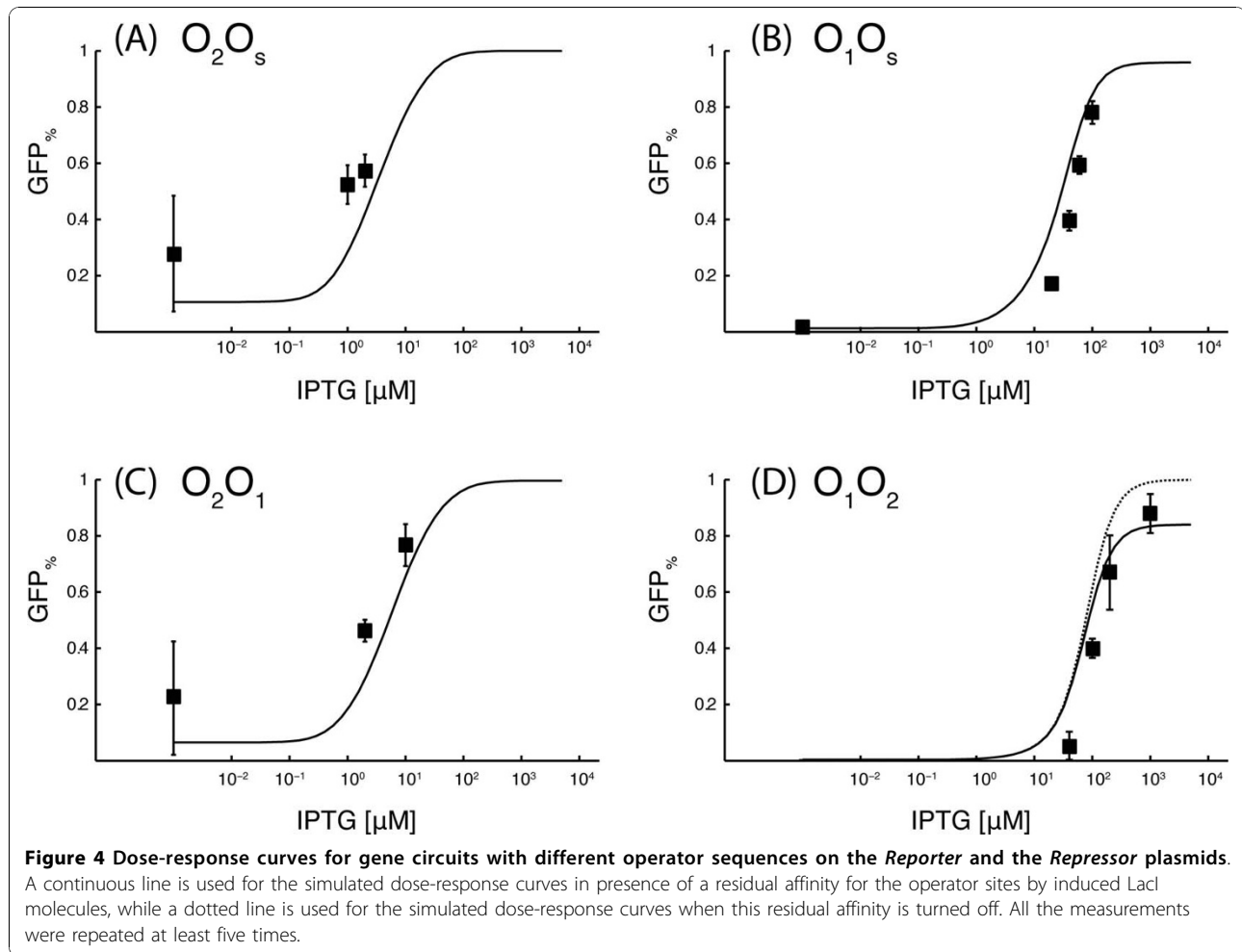
The predicted dose-response curves for the O<sub>2</sub>O<sub>s</sub>, O<sub>1</sub>O<sub>s</sub>, O<sub>2</sub>O<sub>1</sub> and O<sub>1</sub>O<sub>2</sub> gene circuits are in reasonable agreement with the experimental data (Figure 4). It is important to remark that the experimental data for these four circuits were not used in the fitting procedure, and thus the simulated curves in Figure 4 have to be intended as purely predictive.

## Discussion

The design of new synthetic gene devices will be highly facilitated by a modular structure of gene circuits and a bottom-up strategy. While the synthetic biology

community is rapidly adopting a modular architecture, quantitative studies of how multi-part systems behave are still missing. Here, we analyzed gene transcription in a modular gene circuit, and tested if the properties of the complete device can be predicted from the properties of its elementary parts. The major findings are that: (i) each modular part may behave in a different way if it is either isolated or included in a more complex gene circuit; and (ii) mathematical modeling can, however, predict the outcome of the device, if the elementary parts are characterized in the proper experimental conditions. These experimental conditions must be as close as possible to the gene circuit where the elementary parts should be used, in order to compensate for the non-ideal modularity of the system.

In an ideal modular system, the insertion of a new component should not affect the properties of the other parts, as each of them should work independently. In the real world, this is never the case. The properties of the different parts in a biological system are obviously affected by other parts. This poses major problems for



the design of synthetic gene circuits with a bottom-up approach. Even the simplified system analyzed in this study is not strictly modular. In fact, cells transformed with both the *Repressor* and the *Reporter* plasmids have lower fluorescence than those transformed with the *Reporter* plasmid alone, even though the LacI-dependent repression of gene transcription is completely suppressed by removing the operator sequence from the promoter. The decreased fluorescence in cells transformed with both the plasmids is likely due to the extra burden that the cells have to sustain in presence of a second plasmid and antibiotic [29,30]. Since the same synthetic device can have different outcomes depending on the gene transcription rates, the deviations from a perfect modular behavior have to be carefully taken into account in the design of new gene circuits. Thus, if we want to use a bottom-up approach we need to characterize the elementary parts in a gene circuit reproducing to the closest extent the final synthetic device.

Here we characterized the elementary parts measuring fluorescence in cells transformed with *Reporter* and

*Repressor* plasmids having the same operator sequence and then we used the properties of these systems to predict the behavior of gene circuits with different operator sequences in the two plasmids. The analysis of the response of the circuits with the same operator in the *Reporter* and in the *Repressor* revealed an unexpected feature about the functionality of LacI. Experimental data are well reproduced by the mathematical model only if we hypothesize a residual affinity for the operator sites by the induced LacI molecules. This possibility has already been proposed [31,32], but to the best of our knowledge never measured experimentally. A direct measure of this residual affinity is hampered by its low value compared to the affinity of the operator for the free LacI molecules (4 and 5 orders of magnitude higher respectively for the  $O_s$  and  $O_1$  operator sequences). Our data provide a first experimental measurement of these residual affinities in *E. coli*, which can be useful in mathematical modeling of the lactose repressor system.

An alternative explanation for the decrease of the maximum fluorescence in the  $O_1O_1$  and  $O_sO_s$  gene

circuits is the saturation of the intracellular IPTG concentration at a value around 1 mM. However, no data to support this hypothesis are present in literature for such a value of IPTG concentration. Indeed, IPTG can move through cell membranes both by a diffusive and by an active transport [33], the latter being more efficient for the inward motion. Under these conditions, the intracellular concentration of IPTG should be equal, if not higher, than the extracellular concentration.

The blockade of gene transcription by induced LacI molecules was essential to reproduce the non-complete induction of the  $O_1O_1$  and  $O_sO_s$  circuits. When this term is included in the mathematical model, the theoretical dose-response curves of the three circuits with *Reporter* and *Repressor* plasmids having the same operator sequence are in agreement with the experimental data. Small deviations are observed for the  $O_sO_s$  circuit. Here, the computed IPTG concentration giving half the maximal fluorescence is lower than the experimental value. This discrepancy may result from a poor determination of the  $O_sO_s$  fluorescence in absence of IPTG. Due to the high affinity of LacI to the  $O_s$  operator site, the fluorescence in absence of IPTG is close to cellular auto-fluorescence. Thus, its experimental value can be measured only with low accuracy, which is reflected in a poor determination of the parameter  $K_s^L$  of the mathematical model. Lower values of  $K_s^L$  give a better fitting of the raising phase of the dose-response curve of the  $O_sO_s$  circuit. However, finding the best fitting of the experimental data is out of the scope of this study. More detailed models of the lactose repressor system have been proposed [34,35] and these models may likely provide a better agreement to the experimental data. The choice of an extremely simple model was inspired both by practical and theoretical considerations. First, more detailed models have a higher number of parameters. If these parameters have to be assigned in a non-arbitrary way, more experimental measurements are needed. An excessive burden in the characterization of the elementary parts will make a bottom-up approach to gene circuit design not efficient. Second, mathematical models with many parameters can suffer from the over-fitting problem, i.e. the experimental data used in the fitting procedure are well reproduced because the model describes even the noise in the experimental data. If over-fitting occurs, the parameters of the model cannot be used to analyze slightly different systems. Since this is exactly what we wanted to do here, we preferred to use a simple mathematical model with fewer parameters.

In this context, the agreement between the theoretical and the experimental data of the  $O_1O_s$ ,  $O_2O_s$ ,  $O_2O_1$  and  $O_1O_2$  gene circuits is remarkable. The agreement is extremely good in absence of IPTG, while small deviations are observed in the dose-response curves. Even if

the adopted circuit is simple, it allows the study of an important problem in synthetic biology, i.e. the regulation of gene transcription. In fact, quantitative control of gene expression has always been a central problem for the design of gene networks. While designing the logical scheme of a network is relatively simple, the actual behavior of the network depends on the gene transcription rates, with different transcription rates causing different static and dynamic outcomes [36]. The present study analyzes the crucial aspect of gene transcription control in a quantitative way, and characterizes a modular system for the control of transcription in gene networks. Moreover, the adopted gene circuit has a negative feedback structure, which is shared by a plethora of synthetic devices, and also plays a central role in diverse natural processes, from pattern formation to development. The negative feedback structure speeds up the time for the response of the gene network, and also stabilize its behavior [37]. Thus, the adopted architecture of the network could be a useful solution for gene transcription control in synthetic devices.

## Conclusions

Our results support a bottom-up approach to gene circuit design, but they also highlight the importance of characterizing the elementary blocks in the appropriate experimental conditions. If this is certainly a severe constraint for a bottom-up approach to gene circuit design, it does not rule out this possibility. Synthetic devices work by the interplay of several gene circuits, and usually the hard step in the design of new devices is to tune the transcription of different genes in order to get the desired output. If molecular interplays are shut down one-by-one - as it has been done here by removing the operator sequence in the *Reporter* plasmid - the elementary parts can be characterized in an isolated environment. Then, this information can be used to facilitate the design of new synthetic devices. Moreover, it is important to remember that cell transformation by a second plasmid is a strong test for the modular architecture. If changes are minimal perturbations for the biological system, the deviation from a perfect modularity will likely be smaller, and a functional characterization of the elementary parts in standardized experimental conditions could be adequate to help in the design of new gene circuits.

## Methods

### Plasmid construction

All the biological parts were taken from the Registry of Standard Biological parts, except the lactose operator sites that were synthesized (GeneArt) in the BioBrick standard format. The *Reporter* gene circuit was cloned in a high copy number plasmid (pSB1A2) containing



Ampicillin resistance and a pUC19-derived pMB1 replication origin. The *Repressor* gene circuit was cloned in a medium copy number plasmid (pSB3K3) containing Kanamycin resistance and a pMR101-derived p15A replication origin. The constitutive promoters P<sub>2547</sub> (BBa\_J23100) and P<sub>1429</sub> (BBa\_J23118) were cloned upstream of the lactose operator sequences, respectively in the high and medium copy number plasmid. Three lactose operator sequences were used: O<sub>1</sub>, aattgtgagcgga-taacaatt; O<sub>2</sub>, aaatgtgagcgagtaacaacc; and O<sub>s</sub>, aattgt-gagcgctcacaatt [27,38]. The GFP gene (BBa\_J04031), with a LVA degradation tag, was placed downstream of the P<sub>2547</sub>O<sub>x</sub> regulated promoters in the high copy number plasmid. The LacI-coding sequence with LVA degradation tag (BBa\_C0012) was placed downstream of the P<sub>1429</sub>O<sub>x</sub> regulated promoter in the medium copy number plasmid. The same ribosome binding site (RBS) sequence (BBa\_B0034) was cloned upstream of all the protein coding sequences. A double transcriptional terminator T (BBa\_B0015) was placed downstream of each transcriptional unit. Moreover, a double terminator is present at both sides of the plasmid multiple cloning site in order to prevent random transcriptions. In order to test the modular approach and to better identify the parameters of the mathematical model, we also built: (i) *Reporter* plasmids lacking the operator sequence; (ii) high and medium copy number plasmids both with a non-tagged GFP protein (BBa\_I13504) cloned downstream of the P<sub>1429</sub> promoter; (iii) *Reporter* plasmids with the P<sub>1429</sub> promoter lacking the operator sequence.

#### Fluorescence measurements

Dh5α *E. coli* cells were grown at 37°C in 5 ml of M9 minimal medium, supplemented with casamino acids, thiamine hydrochloride and the appropriate antibiotics. Glucose was the main carbon source (Sigma). Fluorescence measurements were performed after an overnight growth in presence of the desired amount of Isopropyl β-D-1-thiogalacto-piranoside (IPTG; Sigma). 100 μl samples were transferred from each cell culture into a multi-well plate and measured in a Wallac VICTOR<sup>2</sup> reader (Perkin Elmer). Both fluorescence (Fluo; Ex 501/Em 511 nm) and optical density (OD; 600 nm) were measured. Normalized fluorescence, [GFP]au, in arbitrary unit (au), was estimated as:

$$[GFP]au = \frac{Fluo - Fluo_{DH5\alpha}}{OD}, \quad (2)$$

where *Fluo* and *OD* are respectively the fluorescence and the optical density of the sample, and *Fluo*<sub>DH5α</sub> is the average auto-fluorescence in samples of non-

transformed Dh5α cells. These measurements, referred to as overnight growth, correspond to samples evaluated at time equal to 12 h, with an average OD of 0.36 ± 0.10. This time point was determined by preliminary time-course measurements (Additional File 1, Figure S1) where cells were grown up to 14 hours and sampling done every 30 min for different cell cultures; namely: (i) cells transformed with a single plasmid; (ii) cells co-transformed with two plasmids; (iii) cells co-transformed with two plasmids, and induced by IPTG. Throughout these measurements, normalized fluorescence reached a steady-state value after ~11 h, for all the tested cell populations, and this value was maintained stable for the following 3 hours measured. This steady state behavior justifies the static measurements used for the remaining analyses. The linear relationship between absolute fluorescence and optical density (Additional File 1, Figure S1) justifies the normalization procedure performed by equation 2.

#### Statistical analysis

Values are reported as mean ± standard deviation. One-way ANOVA, and the Bonferroni post-hoc test for pairwise comparisons were used for detecting differences in normalized fluorescence between multiple groups, for which normal distributions were found (Jarque-Bera test). A significance level of 95% (p < 0.05) was used for all the statistical analyses. MATLAB package (2007a, The MathWorks, Natick, MA) was used for the statistical tests.

#### Mathematical model and parameter definition

The O<sub>x</sub>O<sub>y</sub> gene-circuit – *Reporter* plasmid with the O<sub>x</sub> operator sequence and *Repressor* plasmid with the O<sub>y</sub> operator sequence – was described by the following differential equations:

$$\frac{dM_{G/L}}{dt} = \alpha_{G/L}^M D_{G/L}^F - \lambda_{G/L}^M M_{G/L}, \quad (3)$$

$$D_{G/L}^F = D_{G/L}^0 - D_{G/L}^L - D_{G/L}^I, \quad (4)$$

$$\frac{dD_{G/L}^L}{dt} = \frac{1}{\tau_{DL}} \left[ \frac{D_{G/L}^F L^F}{K_{x/y}^L} - D_{G/L}^L \right], \quad (5)$$

$$\frac{dD_{G/L}^I}{dt} = \frac{1}{\tau_{DI}} \left[ \frac{D_{G/L}^F I^I}{K_{x/y}^I} - D_{G/L}^I \right], \quad (6)$$

$$\frac{dL^I}{dt} = \frac{1}{\tau LI} \left[ L^F \left( \frac{I}{K LI} \right)^n - L^I \right] - \lambda_L L^I - \frac{dD_G^I}{dt} - \frac{dD_L^I}{dt}, \quad (7)$$

$$\frac{dL^F}{dt} = \alpha_L M_L - \lambda_L L^F - \frac{1}{\tau LI} \left[ L^F \left( \frac{I}{K LI} \right)^n - L^I \right] - \frac{dD_G^L}{dt} - \frac{dD_L^L}{dt}, \quad (8)$$

$$\frac{dG}{dt} = \alpha_G M_G - \lambda_G G. \quad (9)$$

Subscripts  $G$  and  $L$  are used for the variables/parameters of *Repressor* and *Reporter* gene-circuit respectively. Table 1 defines all the symbols used in equations 3-9. Differently from other mathematical models of the lactose repressor system, the equations adopted here take into account a possible inhibition of transcription by induced LacI molecules ( $L^I$ , LacI molecules bound to IPTG). Each promoter switch among three possible states: promoter bound to LacI,  $D_{G/L}^L$ ; promoter bound to induced LacI,  $D_{G/L}^I$ ; and free-promoter,  $D_{G/L}^F$  (Equation 4). As a comparison, simulations without residual affinity of induced lactose repressor molecules for the operator sites were also performed. The mathematical model used for these simulations is provided as supplementary material (Additional File 2). In brief, differential equations 6 and state variables  $D_{G/L}^I$  were removed ( $D_{G/L}^I \equiv 0$ ). The parameters for the two mathematical models were the same, with except of the  $K_{x/y}^I$  parameters that do not appear in the model without residual affinity.

The total number of *Reporter* plasmids per cell (high copy number plasmids,  $D_G^0$ ), was assumed equal to 80. This value only scales the other parameters of the model, but not the conclusions drawn from equations 3-9. In order to estimate the ratio  $D_G^0 / D_L^0$ , we measured the fluorescence in cells transformed with: (i) *Reporter* plasmid missing the operator sequence and with the non-tagged GFP protein (BBa\_I13504) cloned downstream of the  $P_{1429}$  promoter, and (ii) the same gene circuit cloned in the medium copy number plasmid. Cloning the *Reporter* and *Repressor* circuits respectively on a high and medium copy number plasmid gives higher sensitivity to IPTG induction, allowing a more accurate quantitative analysis. At the same time - in the context of a modular approach to gene circuit design - the types of plasmids adopted can be treated as further

interchangeable parts of the circuit. As described above, the ratio between the number of high and medium copy number plasmids per cell was estimated by a separate set of experimental measurements in simplified circuits, and then used for the analysis of all the other circuits - which is the typical approach used for the design of complex circuits using modular parts and a bottom-up strategy.

The degradation rate of mRNA,  $\lambda_{G/L}^M$ , was set to  $1/3.69 \text{ min}^{-1}$  [39], both for LacI and GFP. Protein degradation rate,  $\lambda_{G/L}$ , was measured experimentally by first inducing the  $O_1O_1$  circuit with saturating concentration of IPTG, and then monitoring the GFP decay. In detail, after overnight grown in medium containing 1 mM IPTG, cells were washed with fresh medium and let at  $37^\circ\text{C}$  for growth in IPTG-free medium. Samples were taken every 15 minutes for fluorescence measurement. The GFP translation rate was set to  $540\lambda_G$ , in order to have a ratio of 540 between the number of mRNA and protein molecules at equilibrium [40]. Since the functional form of LacI is tetrameric, the LacI transcription rate was set to  $\alpha_L = \alpha_G/4$ . This choice preserves the ratio of 540 between the number of protein chains and mRNA molecules, assuming that this number is the characteristic gain of the translation process. The equilibrium binding constant of the complexes LacI- $O_1$  ( $K_1^I$ ) and LacI- $O_2$  ( $K_2^I$ ) were set to 0.133 and 1.63 molecule/cell respectively [41] (a cell volume of  $1 \mu\text{m}^3$  was assumed [42]). The value of the mRNA transcription rate constant for the *Reporter* circuit,  $\alpha_G^M$ , was estimated by fitting the relative induction of the  $O_1O_1$  and  $O_2O_2$  circuits in absence of IPTG (See the Results section for the definition of relative induction). The ratio  $\alpha_G^M / \alpha_L^M$  was measured experimentally, comparing the cell fluorescence in cells transformed with *Reporter* plasmids with the  $P_{2547}$  or the  $P_{1429}$  promoter and without the operator sequence. Using a stronger promoter in the *Reporter* plasmids, compared to the *Repressor* plasmids, yields higher sensitivity to IPTG induction. As done for the ratio between the number of plasmids inside the cell, the strength-ratio between the two promoters was identified by experimental measurements in simplified circuits, and then used for the analysis of the remaining gene-circuits, with a typical modular approach to gene circuit design.

The value of the equilibrium binding constant for the LacI- $O_S$  complex ( $K_S^I$ ) was defined by fitting the relative induction of the  $O_SO_S$  circuit in absence of IPTG. The constants for the binding of induced LacI molecules to the three operators sites ( $K_1^I$ ,  $K_2^I$  and  $K_S^I$ ) and the parameter for the LacI-IPTG binding ( $n$  and  $K^{LI}$ ), were defined by fitting the experimental dose response curves

of the three circuits  $O_2O_2$ ,  $O_1O_1$  and  $O_5O_5$ . A separate fitting procedure of the parameters  $n$  and  $K^{LI}$  for the mathematical model without residual affinity between induced lactose repressors and operator sites was also performed, but it did not improve the agreement with the experimental data. Indeed, the model without residual affinity cannot reproduce dose-response curves with a maximum  $GFP\%$  below 100%. Fitting procedure were performed by the *fminsearch* routine of the MATLAB package (2007a, The MathWorks, Natick, MA) minimizing the percentage error.

## Additional material

**Additional File 1: Preliminary dynamical measurements.** Time-course measurements of normalized fluorescence and optical density in cells transformed with different gene-circuits.

**Additional File 2: Mathematical model without induced LacI residual affinity.** Equations of the mathematical model without residual affinity between induced lactose repressor molecules and operator sites

## Author details

<sup>1</sup>Laboratory of Cellular and Molecular Engineering, University of Bologna, I-47521 Cesena, Italy. <sup>2</sup>Department of Electronics, Computer Science and Systems, University of Bologna, I-47521 Cesena, Italy. <sup>3</sup>Department of Medical Surgery and Bioengineering, University of Siena, I-53100 Siena, Italy. <sup>4</sup>Department of Biochemistry "G. Moruzzi", University of Bologna, I-40126 Bologna, Italy.

## Authors' contributions

FC built all the genetic constructs used in this work and performed all the experimental measurements. SF developed the mathematical model and performed the fitting procedure and parameter identification. EG advised on wet-lab procedures. SC designed the experiments conducted by FC and the model developed by SF. All authors wrote, read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 24 May 2010 Accepted: 11 November 2010

Published: 11 November 2010

## References

- Hasty J, McMillen D, Collins JJ: Engineered gene circuits. *Nature* 2002, **420**:224-230.
- Shen-Orr SS, Milo R, Mangan S, Alon U: Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 2002, **31**:64-68.
- Mukherji S, van Oudenaarden A: Synthetic biology: understanding biological design from synthetic circuits. *Nat Rev Genet* 2009, **10**:859-871.
- Andrianantoandro E, Basu S, Karig DK, Weiss R: Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* 2006, **2**:2006 0028.
- Hasty J, McMillen D, Collins JJ: Engineered gene circuits. *Nature* 2002, **420**:224-230.
- Sprinzak D, Elowitz MB: Reconstruction of genetic circuits. *Nature* 2005, **438**:443-448.
- Simpson ML: Cell-free synthetic biology: a bottom-up approach to discovery by design. *Mol Syst Biol* 2006, **2**:69.
- Drubin DA, Way JC, Silver PA: Designing biological systems. *Genes Dev* 2007, **21**:242-254.
- Boyle PM, Silver PA: Harnessing nature's toolbox: regulatory elements for synthetic biology. *J R Soc Interface* 2009, **6**(Suppl 4):S535-546.
- Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M: A tunable synthetic mammalian oscillator. *Nature* 2009, **457**:309-312.
- Haynes KA, Silver PA: Eukaryotic systems broaden the scope of synthetic biology. *J Cell Biol* 2009, **187**:589-596.
- Guet C, abreve IC, Elowitz MB, Hsing W, Leibler S: Combinatorial Synthesis of Genetic Networks. *Science* 2002, **296**:1466-1470.
- Gardner TS, Cantor CR, Collins JJ: Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 2000, **403**:339-342.
- Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J: A fast, robust and tunable synthetic gene oscillator. *Nature* 2008, **456**:516-519.
- Elowitz MB, Leibler S: A synthetic oscillator network of transcriptional regulators. *Nature* 2000, **403**:335-338.
- Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ: Synthetic Gene Networks That Count. *Science* 2009, **324**:1199-1202.
- Endy D: Foundations for engineering biology. *Nature* 2005, **438**:449-453.
- Heinemann M, Panke S: Synthetic biology—putting engineering into biology. *Bioinformatics* 2006, **22**:2790-2799.
- Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, de Mora K, Gliberman AL, Monie DD, Endy D: Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 2009, **3**:4.
- Marchisio MA, Stelling J: Computational design tools for synthetic biology. *Curr Opin Biotechnol* 2009, **20**:479-485.
- Canton B, Labno A, Endy D: Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* 2008, **26**:787-793.
- Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, Hallinan J, Wipat A: Standard Virtual Biological Parts: A Repository of Modular Modeling Components for Synthetic Biology. *Bioinformatics* 2010, **26**.
- Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ: A bottom-up approach to gene regulation. *Nature* 2006, **439**:856-860.
- Ellis T, Wang X, Collins JJ: Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* 2009, **27**:465-471.
- Lisser S, Margalit H: Determination of common structural features in *Escherichia coli* promoters by computer analysis. *Eur J Biochem* 1994, **223**:823-830.
- Cox R, Surette M, Elowitz M: Programming gene expression with combinatorial promoters. *Mol Syst Biol* 2007, **3**.
- Sadler JR, Sasmor H, Betz JL: A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc Natl Acad Sci USA* 1983, **80**:6785-6789.
- Gilbert W, Gralla J, Majors AJ, Maxam A: Lactose operator sequences and the action of lac repressor. In *Protein-ligand interactions*. Edited by: Sund H, Blauer G, Walter de Gruyte, Berlin; 1975:193-207.
- Flores S, de Anda-Herrera R, Gosset G, Bolivar FG: Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol Bioeng* 2004, **87**:485-494.
- Glick BR: Metabolic load and heterologous gene expression. *Biotechnol Adv* 1995, **13**:247-261.
- Daber R, Lewis M: Towards evolving a better repressor. *Protein Eng Des Sel* 2009, **22**:673-683.
- Saiz L, Vilar JM: Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res* 2008, **36**:726-731.
- Noel JT, Pilyugin SS, Narang A: The diffusive influx and carrier efflux have a strong effect on the bistability of the lac operon in *Escherichia coli*. *Journal of Theoretical Biology* 2009, **256**:14-28.
- Santillan M, Mackey MC: Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophys J* 2004, **86**:1282-1292.
- van Hoek M, Hogeweg P: The effect of stochasticity on the lac operon: an evolutionary perspective. *PLoS Comput Biol* 2007, **3**:e111.
- Atkinson MR, Savageau MA, Myers JT, Ninfa AJ: Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia coli*. 2003, **113**:597-607.
- Rosenfeld N, Elowitz MB, Alon U: Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 2002, **323**:785-793.
- Gilbert W, Maxam A: The nucleotide sequence of the lac operator. *Proc Natl Acad Sci USA* 1973, **70**:3581-3584.

39. Bernstein JA, Lin PH, Cohen SN, Lin-Chao S: **Global analysis of Escherichia coli RNA degradosome function using DNA microarrays.** *Proc Natl Acad Sci USA* 2004, **101**:2758-2763.
40. Lu P, Vogel C, Wang R, Yao X, Marcotte EM: **Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.** *Nat Biotechnol* 2007, **25**:117-124.
41. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R: **Transcriptional regulation by the numbers: applications.** *Curr Opin Genet Dev* 2005, **15**:125-135.
42. Phillips R, Kondev J, Theriot J: **Physical Biology of the Cell.** *Garland Science* 2008.
43. Dunaway M, Olson JS, Rosenberg JM, Kallai OB, Dickerson RE, Matthews KS: **Kinetic studies of inducer binding to lac repressor-operator complex.** *J Biol Chem* 1980, **255**:10115-10119.

doi:10.1186/1754-1611-4-14

**Cite this article as:** Ceroni *et al.*: Rational design of modular circuits for gene transcription: A test of the bottom-up approach. *Journal of Biological Engineering* 2010 **4**:14.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

