

A COMPUTATIONAL MODEL OF GENE EXPRESSION IN AN INDUCIBLE SYNTHETIC CIRCUIT

FRANCESCA CERONI

*Department of Electronics, Computer Sciences and Systems (DEIS)
Cellular and Molecular Engineering Laboratory, University of Bologna, via Venezia 52
Cesena, 47023, ITALY*

SIMONE FURINI

*Department of Medical Surgery and Bioengineering
University of Siena, viale Mario Bracci 12, Siena, 53100, Italy*

SILVIO CAVALCANTI

*Department of Electronics, Computer Sciences and Systems (DEIS)
Cellular and Molecular Engineering Laboratory, University of Bologna, via Venezia 52
Cesena, 47023, ITALY*

Abstract. Synthetic biology aims to the rational design of gene circuits with predictable behaviours. Great efforts have been done so far to introduce in the field mathematical models that could facilitate the design of synthetic networks. Here we present a mathematical model of a synthetic gene-circuit with a negative feedback. The closed loop configuration allows the control of transcription by an inducer molecule (IPTG). *Escherichia coli* bacterial cells were transformed and expression of a fluorescent reporter (GFP) was measured for different inducer levels. Computer model simulations well reproduced the experimental induction data, using a single fitting parameter. Independent genetic components were used to assemble the synthetic circuit. The mathematical model here presented could be useful to predict how changes in these genetic components affect the behaviour of the synthetic circuit.

1. Introduction

Synthetic Biology is a novel discipline defined as the engineering of biology, i.e. the synthesis of systems based on biological material, which display prefixed functions. This engineering perspective may be applied at all hierarchical levels of biology, from individual molecules to whole cell, tissues and organisms (1). A major focus of the discipline is the synthesis of genetic components and gene circuits with predictable behaviours (2), either to endow cells with novel functions, or to study analogous natural systems (3). In the last years many gene circuits have been developed to achieve a fine regulation of gene expression and protein synthesis (4-6). Among them, auto-regulated networks with negative feedback loops have been invoked as a way to control and decrease transcriptional noise (7), conferring stability to gene expression, and high sensitivity to induction by an extracellular stimulus (8). A crucial element in these networks is the presence of promoter sequences with one or more operator sites that can be recognized by regulatory molecules, the transcription factors (TF). The position of the operator sequences inside the promoter region has a strong impact on how transcription factors control gene-transcription. Notably, the presence of an operator downstream of the TATA box region is responsible for transcription repression (9). This property can be used to assemble synthetic promoters that are repressed by the same TF, but which have different constitutive transcriptional strengths. Here we present a computational study of a synthetic device where gene expression is controlled by regulated promoters. The device includes two parts (Fig. 1): (i) an auto-regulated

generator of the LacI repressor protein (LacI-supplier); and (ii) a LacI-inverter, which uses a GFP protein as reporter. The mathematical model describes the dynamical interactions between these parts and accurately reproduces the device response to IPTG induction.

2. Materials and Methods

2.1. Gene circuit scheme

The LacI and GFP genes were placed both under the control of a synthetic promoter repressed by the LacI protein (Fig. 1). The promoters were designed by assembling the natural Lac operator sequence O₂ downstream of two constitutive promoters (P_{LacI} and P_{GFP}), with the P_{GFP} transcriptional strength greater than the P_{LacI} one. The LacI repressor protein can bind to the operator site O₂ preventing the binding of DNA-polymerase to the P_{LacI} promoter. As a result, the LacI transcription is auto-regulated by a negative feedback. The LacI protein amount produced by the LacI-supplier also controls the GFP protein transcription by binding to the operator site O₂ placed downstream of the constitutive promoter P_{GFP}. When the LacI molecule concentration increases, the GFP reporter decreases, thus this part was called the LacI-inverter. The LacI-supplier was cloned in a medium copy number plasmid, while the LacI-inverter was cloned in a high copy number plasmid. This allows the amplification of the GFP reporter. The inducer molecule, IPTG (Isopropyl β-D-1-thiogalactopyranoside), inhibits the repressor activity for the operator site O₂ enhancing the transcription rate of the GFP protein. The inducer is considered as the external input of the device and regulates GFP expression.

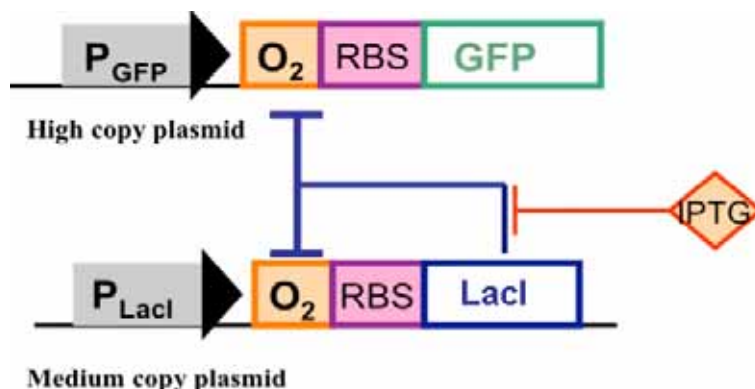


Figure 1 Gene network scheme

2.2. BioBricks and Bacterial strain

The constitutive promoters P_{GFP} (BBa_J23100) and P_{LacI} (BBa_J23118) were taken from the Registry of standard biological parts (10, 11) and were both cloned upstream of the same natural Lac operator sequence O₂ (aaatgtgagcgagtaacaacc). The GFP gene with a degradation tag was

placed under the control of the P_{GFP} regulated promoter on a high copy number plasmid (pSB1A2) with Ampicillin resistance and a pUC19-derived pMB1 replication origin. The Lac repressor coding sequence with a degradation tag was placed under the control of the P_{LacI} regulated promoter on a medium copy number plasmid (pSB3K3) with Kanamycin resistance and a p15A pMR101-derived replication origin. The two vectors were co-transformed in *Escherichia coli* Dh5 α cells (closed loop configuration) and grown under a two antibiotic selection. The same cells were also transformed with the LacI-inverter plasmid only (open loop configuration).

2.3. Growth conditions and media

Cells were grown in flasks at 37 °C in M9 minimal medium with glucose as the main carbon source, supplemented with casamino acids, thiamine and the proper antibiotics. Cells were induced for GFP expression by overnight growth with different IPTG concentrations (1 μ M, 2 μ M, 6 μ M, 10 μ M, 60 μ M, 75 μ M, 100 μ M), while cultures with uninduced cells were prepared by overnight growth in the absence of IPTG. During each experimental run, after overnight growth, 100 μ l from each cell culture were transferred into a multiwell plate for fluorescence analysis with the Victor 2 plate reader (Perkin Elmer). Wells on the plate edges were filled with fresh M9 medium in order to close off the samples from the external environment and to limit the thermal dispersion. For each well both fluorescence and optical density were measured. The green fluorescence protein values were estimated by normalizing the total fluorescence on the corresponding optical density.

2.4. Mathematical Model

The mathematical model describes gene expression from two independent co-transformed plasmids. The subscript R and G will be used for variables related to the LacI repressor and the the GFP reporter, respectively. Each plasmid can switch among three functional states: (i) bound to RNA-polymerase; (ii) bound to repressor protein; (iii) and free. The conservation of plasmid number in a single cell gives:

$$\begin{aligned} N_R &= L_R + R_R + P_R \\ N_G &= L_G + R_G + P_G \end{aligned} \quad (1)$$

where N_R and N_G are the plasmid numbers in each cell for the medium-copy and low-copy respectively; P_R, R_R, L_R are the number of plasmids bound to RNA polymerase, to repressor protein or free. When a free plasmid binds to a RNA polymerase molecule (P), an activated complex (P_R, P_G) forms that can initiate transcription and mRNA generation. The following differential equations describe the process:

$$\begin{aligned}\frac{dP_R}{dt} &= k_{RP}^f L_R P - (k_{RP}^b + \alpha_{RNA}) P_R = \frac{1}{\tau_{P_R}} \left(\frac{L_R P}{K_{RP}} - P_R \right) \\ \frac{dP_G}{dt} &= k_{GP}^f L_G P - (k_{GP}^b + \alpha_{RNA}) P_G = \frac{1}{\tau_{P_G}} \left(\frac{L_G P}{K_{GP}} - P_G \right).\end{aligned}\quad (2)$$

Tables 1 and 2 summarize the variables and the parameters used in the mathematical model. The second expression of the differential equations was preferred because it highlights the time-constant of the process. We hypothesized that the number of polymerase molecules available for heterologous transcription is constant, which gives:

$$P_0 = P + P_R + P_G, \quad (3)$$

where P_0 is the total number of RNA-polymerase molecules. mRNA molecule synthesis and degradation were described by first order equations:

$$\begin{aligned}\frac{dM_R}{dt} &= \alpha_{RNA_R} P_R - \beta_{RNA_R} M_R = \frac{1}{\tau_{RNA_R}} (M_{R,\infty} P_R - M_R) \\ \frac{dM_G}{dt} &= \alpha_{RNA_G} P_G - \beta_{RNA_G} M_G = \frac{1}{\tau_{RNA_G}} (M_{G,\infty} P_G - M_G)\end{aligned}\quad (4)$$

Equations analogous to (2), (3) and (4) were used to describe the binding of the mRNA molecules M_R and M_G , to ribosomes and the synthesis of the two proteins R and G :

$$\begin{aligned}\frac{dB_R}{dt} &= k_{RB}^f M_R B - (k_{RB}^b + \alpha_{PROT}) B_R = \frac{1}{\tau_{B_R}} \left(\frac{M_R B}{K_{RB}} - B_R \right) \\ \frac{dB_G}{dt} &= k_{GB}^f M_G B - (k_{GB}^b + \alpha_{PROT}) B_G = \frac{1}{\tau_{B_G}} \left(\frac{M_G B}{K_{GB}} - B_G \right),\end{aligned}\quad (5)$$

$$\begin{aligned}\frac{dR_T}{dt} &= \alpha_{PROT_R} B_R - \beta_{PROT_R} R_T = \frac{1}{\tau_{PROT_R}} (P_{R,\infty} B_R - R_T) \\ \frac{dG}{dt} &= \alpha_{PROT_G} B_G - \beta_{PROT_G} G = \frac{1}{\tau_{PROT_G}} (P_{G,\infty} B_G - G)\end{aligned}\quad (6)$$

$$B_0 = B + B_R + B_G. \quad (7)$$

Note that in equations (6) the symbol R_T is used for the repressor in place of R . R_T is the number of all the repressor molecules, while R is the number of free repressors, R_G is the number of repressors bound to the high-copy number plasmids, R_R the repressors bound to the low-copy number plasmids and R_I the repressors bound to inducer molecules. Thus:

$$R_T = R + R_I + R_R + R_G. \quad (8)$$

The binding process between the free repressor, R , and the inducer, I , was described by the equation:

$$\frac{dR_I}{d} = k_{RI}^f R I^n - k_{RI}^b R_I = \frac{1}{\tau_I} \left(\frac{R I^n}{K_I} - R_I \right), \quad (9)$$

where n is the binding cooperativity.

Finally, the following equations complete the mathematical model describing the free repressor binding to the operator site O_2 on the plasmids:

$$\begin{aligned}\frac{dR_R}{dt} &= k_{RR}^f L_R R - k_{RR}^b R_R = \frac{1}{\tau_{R_R}} \left(\frac{L_R R}{K_{RR}} - R_R \right) \\ \frac{dR_G}{dt} &= k_{GR}^f L_G R - k_{GR}^b R_G = \frac{1}{\tau_{R_G}} \left(\frac{L_G R}{K_{GR}} - R_G \right)\end{aligned}\quad (10)$$

Model equations were implemented in Simulink (Mathworks), and were numerically integrated starting from null initial conditions for the state variables. Model parameters describe well-characterized physical processes involving standard biological molecules, thus parameter values were taken from the literature (see Table 2 for details). The only parameter used to fit the experimental data was the repressor-operator binding constant. This was identified by fitting experimental data on GFP production, measured as fluorescence in the cell population. Parameter identification was performed by the Matlab routine `fminsearch`.

Table 1. Model variables

Symbol	Definition
$L_{R/G}$	Free plasmids, i.e. not bound to Dna-polymerase or repressor
$P_{R/G}$	Plasmids bound to Dna-polymerase
P	Free polymerase, i.e. not bound to any plasmid
$M_{R/G}$	mRNA
$B_{R/G}$	Ribosomes bound to $M_{R/G}$
B	Free ribosomes, i.e. not bound to any mRNA molecule
$R_{R/G}$	Plasmids bound to the repressor
R	Free repressor, LacI
R_I	IPTG bound repressor
R_T	Total repressor molecules, free and bound
G	Reporter protein, GFP (green fluorescence protein)

Table 2. Model parameters

Symbols	Definition and Comments	Value
N_R	Number of medium-copy plasmids	15 (11)
N_G	Number of high-copy plasmids	150 (10)

P_0	Number of polymerase molecules	1500 (12)
B_0	Number of ribosomes	6800 (12)
α_{PROT_R}	LacI synthesis rate. The parameter was calculated as the peptide chain elongation rate multiplied by the number of LacI residues.	2.85 min ⁻¹ (12)
α_{PROT_G}	GFP synthesis rate. The parameter was calculated as the peptide chain elongation rate multiplied by the number of GFP residues.	1.92 min ⁻¹ (12)
α_{RNA_R}	LacI mRNA molecule synthesis rate. The parameter was calculated as the mRNA elongation rate multiplied by the number of nucleotides in LacI.	3.17 min ⁻¹ (13)
α_{RNA_G}	GFP mRNA molecule synthesis rate. The parameter was calculated as the mRNA elongation rate multiplied by the number of nucleotides in GFP.	4.63 min ⁻¹ (13)
$\beta_{PROT_R/G}$	Protein degradation rate. This parameter was estimated through experimental measurements of GFP degradation rate. Analogous degradation rate was assumed for the LacI protein.	2.13*10 ⁻² min ⁻¹
$\beta_{RNA_R/G}$	mRNA degradation rate	0.19 min ⁻¹ (14)
$k_{GB}^{f/b}$	Forward and backward kinetic rates for the reaction between the reporter gene mRNA molecules and the ribosomes	
$k_{RB}^{f/b}$	Forward and backward kinetic rates of the reaction between the repressor protein mRNA molecules and the ribosomes	
$k_{GR}^{f/b}$	Forward and backward kinetic rates of the reaction between free repressor molecules and the operator O ₂ on the high-copy number plasmid	
$k_{RR}^{f/b}$	Forward and backward kinetic rates of the reaction between free repressor molecules and the operator O ₂ on the low-copy number plasmid	
$k_{RP}^{f/b}$	Forward and backward kinetic rates of the reaction between DNA polymerase molecules and the promoter on the low-copy number plasmids	
$k_{GP}^{f/b}$	Forward and backward kinetic rates of the reaction between DNA polymerase molecules and the promoter on the high-copy number plasmids	
$k_{RI}^{f/b}$	Forward and backward kinetic rates of the reaction between free repressor and inducer molecules	
$\tau_{P_{R/G}}$	$\frac{1}{k_{R/GP}^b + \alpha_{RNA_R/G}}$	1.7*10 ⁻³ min (8)
K_{RP}	$\frac{k_{RP}^b + \alpha_{RNA_R}}{k_{RP}^f}$	1.5*10 ⁴ (15)
K_{GP}	$\frac{k_{GP}^b + \alpha_{RNA_G}}{k_{GP}^f}$ The ratio between GFP expression with the two promoters, P _{GFP} and P _{LacI} , was determined experimentally as 1.4. The same ratio was assumed between K_{RP} and K_{GP}	0.7*10 ⁴
$\tau_{RNA_R/G}$	$\frac{1}{\beta_{RNA_R/G}}$	5.32 min

$M_{R,\infty}$	$\frac{\alpha_{RNA_R}}{\beta_{RNA_R}}$		16.9
$M_{G,\infty}$	$\frac{\alpha_{RNA_G}}{\beta_{RNA_G}}$		24.6
$\tau_{B_{R/G}}$	$\frac{1}{k_{R/GB}^b + \alpha_{PROT_R/G}}$		$7.4 \cdot 10^{-3}$ min (13)
$K_{R/GB}$	$\frac{k_{R/GB}^b + \alpha_{PROT_R/G}}{k_{R/GB}^f}$		$4.4 \cdot 10^{-3}$ (13)
$\tau_{PROT_R/G}$	$\frac{1}{\beta_{PROT_R/G}}$		47 min
$P_{R,\infty}$	$\frac{\alpha_{PROT_R}}{\beta_{PROT_R}}$		90
$P_{G,\infty}$	$\frac{\alpha_{PROT_G}}{\beta_{PROT_G}}$		134
n	IPTG inducer binding cooperativity		2
τ_I	$\frac{1}{k_{RI}^f}$		22.4 min (16)
K_I	$\frac{k_{RI}^b}{k_{RI}^f}$		1680 (17)
$\tau_{R_{R/G}}$	$\frac{1}{k_{G/RR}^b}$		$1 \cdot 10^{-3}$ min
$K_{R/GR}$	$\frac{k_{RR}^b}{k_{RR}^f}$	This is the only parameter of the mathematical model adjusted to fit the experimental data.	0.3

3. Results

Figure 2 shows how GFP transcription can be modulated using different IPTG concentrations. A good agreement between the mathematical model and the experimental data was obtained by tuning the parameter K_{RR} , which models the binding affinity of the repressor for the operator on the medium-copy number plasmid. Since the same operator sequence was used both in the high copy number and low copy number plasmid, parameters K_{RR} and K_{GR} were assumed equal. The

fluorescence observed at different IPTG concentrations was divided by the fluorescence measured in the open loop configuration (bacteria transformed only with the LacI-inverter circuit), thus defining a normalized measurement of GFP expression levels.

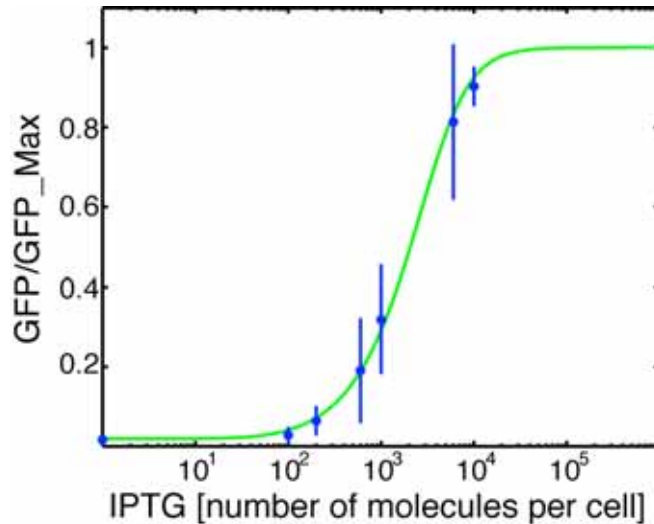


Figure 2 Dose Response curve. GFP expression was normalized to the value in the open-loop gene circuit, both for the experimental data (blue points, bars for standard deviation) and for the results of the numerical simulations (green line).

To simulate the regulated promoter response to changes in the LacI repressor level (Fig. 3), we forced the model by opening the negative feedback and using R as an input. R was slowly increased from 0 to 1000 molecules and the quasi-static level of the reporter protein G was computed. Simulation shows that the regulated promoter is repressed at 50% by ~100 molecules of LacI (see Fig. 3). This value closely agrees with the experimental determination of LacI repressor binding constant on the P_{Lac} promoter (70 molecules).

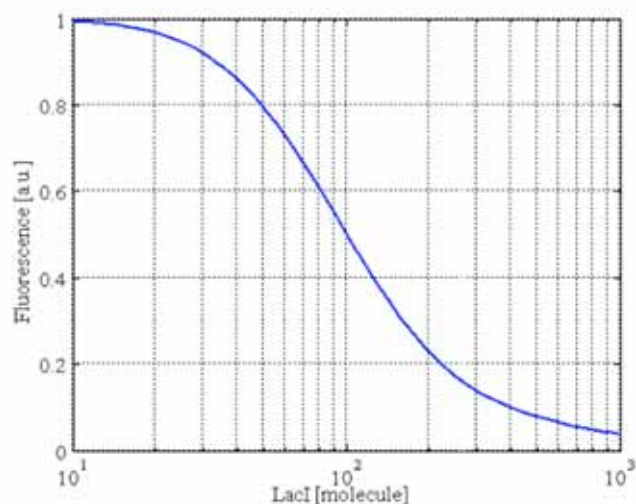


Figure 3. Fluorescence decreasing in presence of the LacI repressor. The 50% of repression is achieved with 100 repressor molecule.

To characterize the dynamic response of the device we considered the rise-time, (i.e. the delay between the initial and half maximal production). The rise-time was estimated by simulating the transient of the gene-circuit from null initial conditions ($M_R = R = 0$ and $M_G = G = 0$), in both open and close loop configurations (Fig. 4). In open-loop, when the GFP promoter is unregulated, the circuit reaches the steady state after 180 min, while in close-loop there is an initial over-shoot at 20 min due to the latency in the R protein synthesis. After 80 min the transient is extinct.

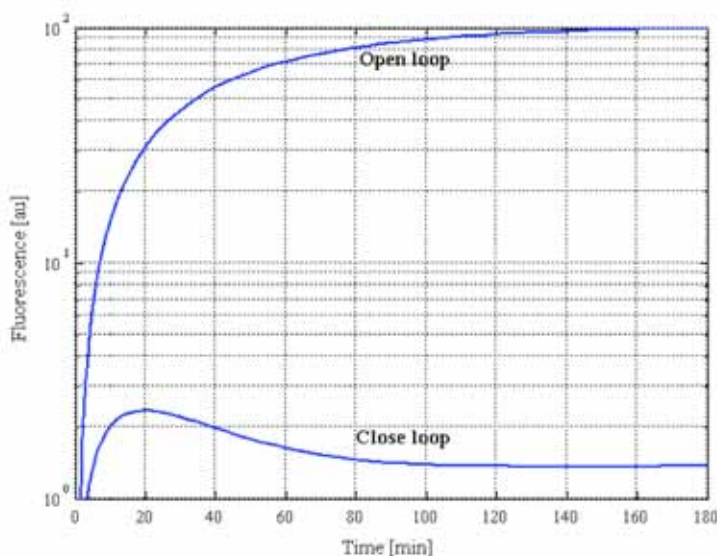


Figure 4. Dynamic response of close- and open-loop configurations. Open loop refers to the cells transformed with the Lac-inverter plasmid only, while closed loop configuration refers to cells co-transformed with both LacI-supply and the LacI-inverter plasmids

4. Discussion

In recent years an increasing number of studies in synthetic biology have been focused on constructing simple synthetic gene circuits that exhibit desired properties. In this work, we used a gene network with a negative feedback, where the Lac repressor protein (LacI) is responsible for the repression of GFP transcription from a strong promoter on a high copy number plasmid and, at the same time, it prevents its own transcription from a weak promoter on a medium copy number plasmid, auto-regulating itself. Since the usage of the same biological parts in another synthetic gene network would be greatly facilitated by their mathematical modelling, we performed an analytical analysis of the gene circuit. After the model formulation, we have cloned the device in *Escherichia coli* to compare the experimental measurements with the computer model prediction. A good fitting of the experimental data was obtained (see Fig. 2) only by tuning the LacI repressor affinity to the operator site, whereas the other model parameters were assigned according to the literature. Notably, variation of the circuit genetic components can be simulated *in-silico* by changing the appropriate model parameter. As an example, after model identification we computed the LacI-inverter response to LacI repressor (Fig. 3). This curve can be used to establish the sensitivity of the LacI-inverter to LacI changes in the region with negative slope (-4 % for each 10 LacI molecules) as well as the amount of LacI molecules (10^3) needed to get a 95% GFP repression. Moreover, the mathematical model can be used to identify which modifications should be introduced in the circuit to obtain a specific sensitivity to the inducer molecules or to change the parameter values to predict the protein expression levels from promoters with different transcriptional strengths.

We used the model to compute the dynamic behaviour of the device through time. To characterize the dynamics we considered the rise-time, (i.e. the delay from the initiation of production until half maximal product concentration is reached). According to previous observation (18), the rise-time in negatively auto-regulated transcription circuits was shorter than in non-regulated transcription. Negative feedback (also termed autogenous control) reduces the rise-time. This may help in understanding the function of negative auto-regulation, which appears in over 40% of known transcription factors in *E. coli*.

Mathematical modelling could prove to be useful in the design of synthetic gene networks. The behaviour of a gene network is highly affected by the transcription rates of the involved genes. The optimization of these transcription rates is usually a time-consuming task, which is a severe obstacle for a further evolution of synthetic biology. The gene circuit presented here does not only offer the possibility to tune gene expression in a predictable way by controlling the concentration of an extracellular inducer molecule, but it also exhibits a modular architecture, which is particularly useful for the control of gene transcription in synthetic biology. Different plasmids types, operator sequences, promoters, can be combined to obtain a set of circuits, whose characteristics can be described, and eventually predicted, with the same mathematical model, with a typically engineering approach.

5. References

1. Serrano, L., *Mol Syst Biol*, **3**:158 (2007).
2. Guido, N.J., et al., *Nature*, **439**(7078):856-860 (2006).
3. Ellis, T., et al., *Nat Biotechnol*, **27**(5):465-71 (2009).
4. Elowitz, M.B., et al., *Nature*, **403**(6767):335-338 (2000).
5. Gardner, T.S., et al., *Nature*, **403**(6767):339-342 (2000).
6. Atkinson, M.R., et al., **113**(5):597-607 (2003).
7. Dublanche, Y., et al., *Mol Syst Biol*, **2**:41 (2006).
8. Stricker, J., et al., *Nature*, **456**(7221):516-519 (2008).
9. Cox, R., et al., *Mol Syst Biol*, **3** (2007).
10. Knight, T. <http://partsregistry.org/Part:pSB1A2>. (2004).
11. Shetty, R. <http://partsregistry.org/Part:pSB3K3>. (2004).
12. Bremer, H., et al., *Modulation of chemical composition and other parameters of the cell by growth rate, in Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Neidhardt, Editor. (1996)
13. Epshtein, V., et al., *Science*, **300**(5620):801-805 (2003).
14. Bernstein, J.A., et al., *Proc Natl Acad Sci U S A*, **101**(9):2758-63 (2004).
15. Bintu, L., et al., *Current Opinion in Genetics & Development*, **15**(2):116-124 (2005).
16. Xie, X.S., et al., *Annual Review of Biophysics*, **37**(1):417-444 (2008).
17. Wilson, C.J., et al., *Biophysical Chemistry*, **126**(1-3):94-105 (2007).
18. Rosenfeld, N., et al., *J Mol Biol*, **323**(5):785-93 (2002).