chatter plant

MODELING

Universidad Politécnica de Valencia
Valencia UPV iGEM 2017

iGEM UPV VALENCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Tabla de contenido

1. INTRODUCTION

Modeling our gene circuits is of paramount importance since it provides us with deep insight and prediction capability of the biological processes taking place in ChatterPlant. This way, we aim to break the traditional wall imposed by trial-error approaches, which often result in unnecessary or inconclusive experiments.

Mathematical models in synthetic biology contribute not only to generate empirically contrastable hypothesis but also to manage laboratory time and hardware resources efficiently. Furthermore, because of the understanding and foresight provided by modeling, tuning and programing ChatterPlant (link a la chatter) according to the necessities of certain situations is possible more than ever.

1.1. DETERMINISTIC AND STOCHASTIC MODELS

In ChatterPlant we analyze the dynamic behavior of our biological system (link explicación circuito) considering the biochemical species involved in a certain set of reactions. According to the degree of approximation to capture the dynamic behavior, we can differentiate two approaches:

1. **Deterministic**. Deterministic models do not take into account the natural randomness of the reactions. For each chemical species, the amount of molecules transformed within reactions only depends on the initial amount of molecules, reaction rates and stoichiometry relations.
   The type of deterministic model that we implement is a system of Ordinary Differential Equations (ODEs) (link a este modelo).
2. **Stochastic**. Inherent noise due to random events plays a relevant role in the dynamics. As a deterministic model does not capture noise, we use stochastic linear differential equations (link a este modelo).

2. OVERVIEW

2.1. WHAT WE ARE MODELING

The SynBio-based design integrated in ChatterPlant is composed by two gene circuits. The analysis of both their single performance and their interaction with several factors (e.g. the cell medium, environment and ChatterBox), is basic to reach one of our main goals: a new sustainable and efficient agriculture system (link a human?).

Our model comprises of:

1. Optogenetic circuit. How long has to remain the light pulse in order to get a certain protein amount? How could be optimized the energetic supply of LEDs light in order to maximize a cycle of protein production? These questions among others arose to our minds as we were designing the LED's system.

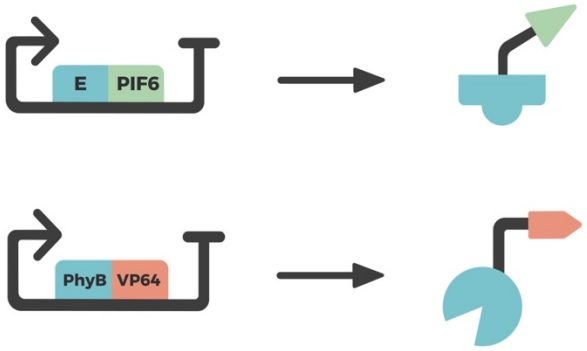
2. Sensor circuit. Bearing in mind the time span during which the biological sensor is transiently ON, it is mandatory to design a sampling rate test according to the plant's periodic security necessities.

2.2. MODELING SOFTWARE MODULES

We start building the genetic circuits from basic modules, coupling them to generate the mathematical model of the whole system. As UPV_iGEM is an interdisciplinary team, most of the models generated in ChatterPlant are included in the modeling software tool (**link modeling software tool**) and are represented by modules in an artistic graphic interface, for the purpose of introducing researchers to a more realistic conception of the engineering in biology, meanly, SynBio.

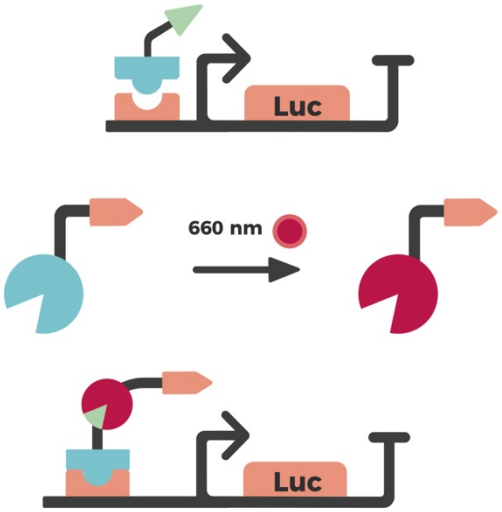
3. HUMAN-PLANT: OPTOGENETIC CIRCUIT

Two constitutive modules express the E-PIF6 and PhyB-VP64 fusion proteins that regulate the output expression.



Constitutive modules representation of the fusion proteins E-PIF6 and PhyB-VP64.

E-PIF6 binds to the promoter's operator. When red light (660 nm wavelength) LEDs are switched on, PhyB changes its conformation (PhyB*) and binds to PIF6. Consequently, the transcription of the desired protein starts because of the RNAp recruitment by VP64.



Expression regulated by the transcriptional factors.

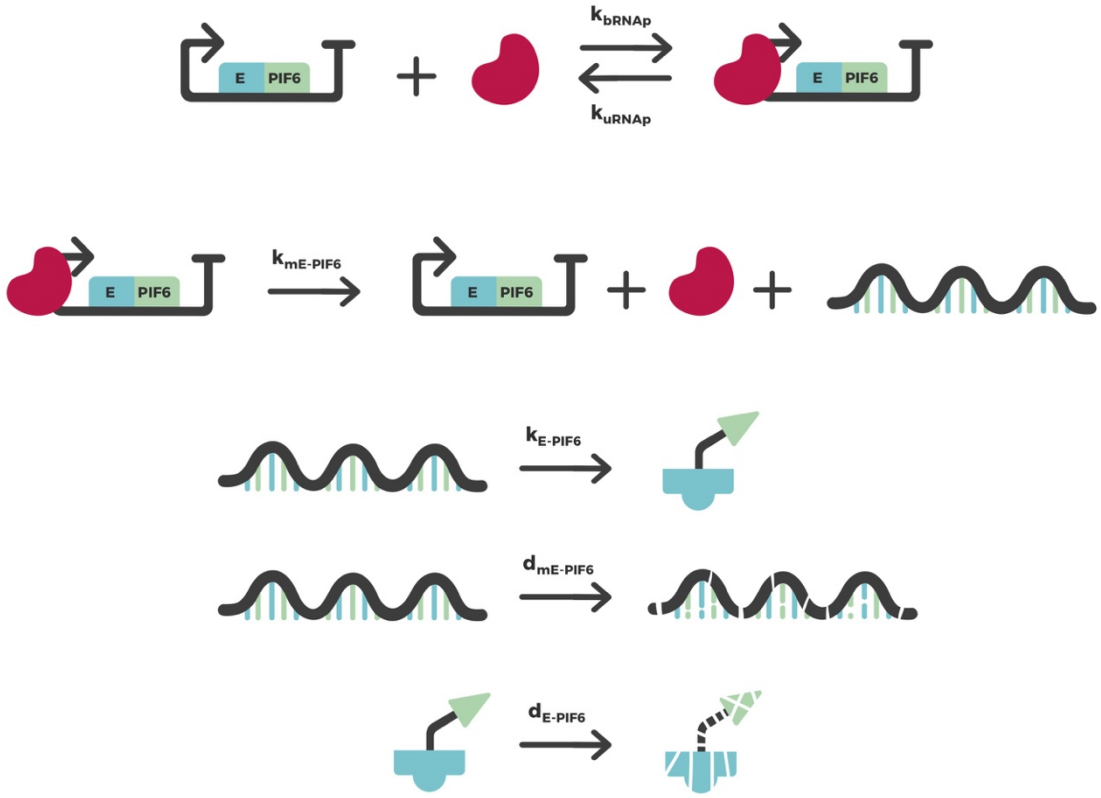
Far red light (740 nm wavelength) reverts PhyB* to its natural conformation (PhyB). This change stops de transcriptional activity of the third optogenetic circuit's module.

Switch off.

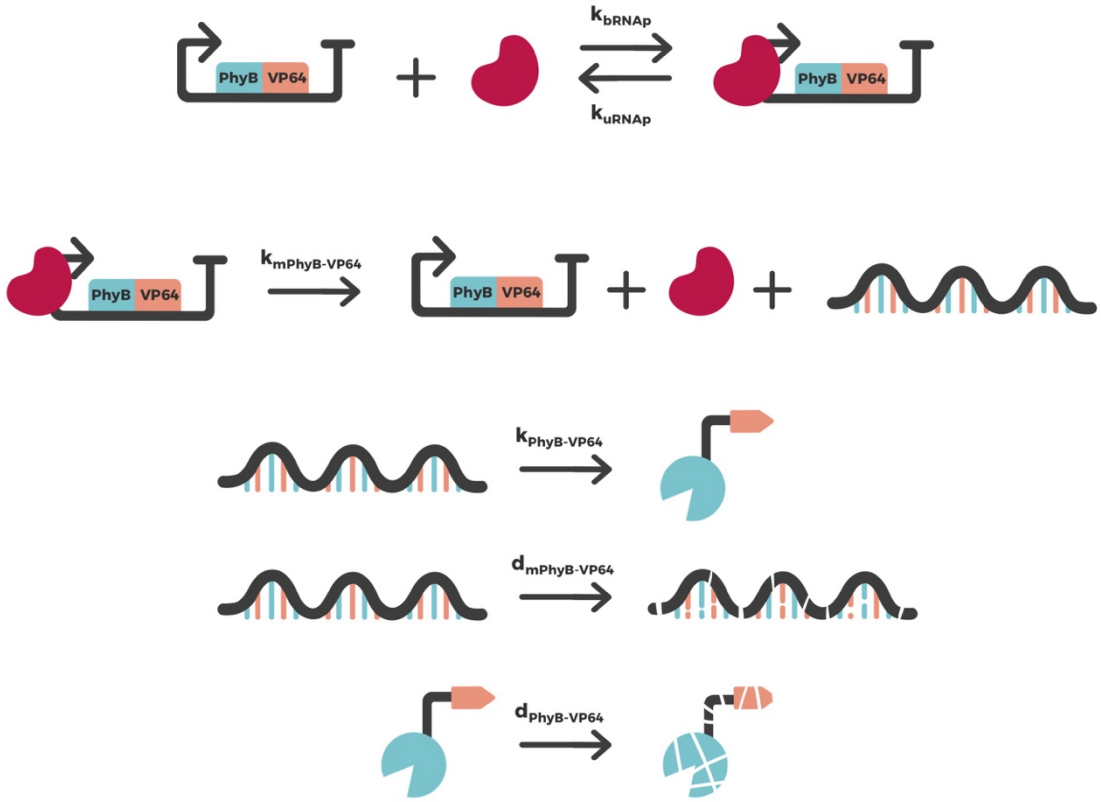
3.1. DETERMINISTIC

3.1.1. REACTIONS

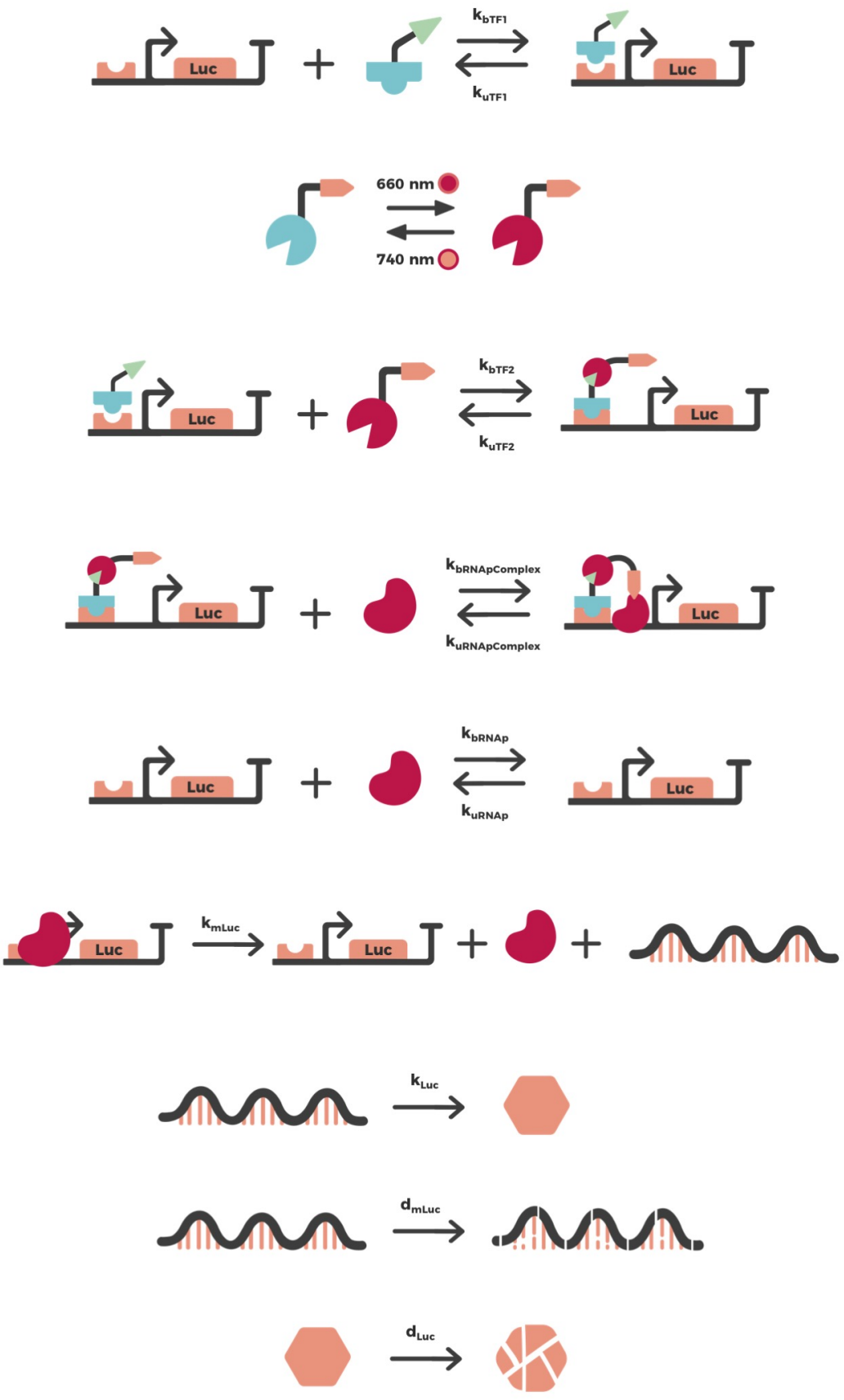
Now we take into account the principal reactions in each module representing them both graphic design and formal reactions.



E-PIF6 expression



PhyB-VP64 expression


Luc
k_{bTF1}
k_{uTF1}
660 nm
740 nm
k_{bTF2}
k_{uTF2}
$k_{bRNApComplex}$
$k_{uRNApComplex}$
k_{bRNAp}
k_{uRNAp}
k_{mLuc}
k_{Luc}
d_{mLuc}
d_{Luc}

Regulated expression

REACTIONS

Constitutive module A=E-PIF6

$$gA + RNAp \underset{k_{uRNAp}}{\overset{k_{bRNAp}}{\rightleftharpoons}} gA \cdot RNAp$$

$$gA \cdot RNAp \xrightarrow{k_{mA}} gA + RNAp + mA$$

$$mA \xrightarrow{k_A} mA + A$$

$$mA \xrightarrow{d_{mA}} \emptyset$$

$$A \xrightarrow{d_A} \emptyset$$

Constitutive module B=PhyB-VP64

$$gB + RNAp \underset{k_{uRNAp}}{\overset{k_{bRNAp}}{\rightleftharpoons}} gB \cdot RNAp$$

$$gB \cdot RNAp \xrightarrow{k_{mB}} gB + RNAp + mB$$

$$mB \xrightarrow{k_B} mB + B$$

$$mB \xrightarrow{d_{mB}} \emptyset$$

$$B \xrightarrow{d_B} \emptyset$$

Regulated module

$$gC + A \underset{k_{uTF1}}{\overset{k_{bTF1}}{\rightleftharpoons}} gC \cdot A \triangleq C_C^{TF1}$$

$$B \underset{740nm}{\overset{660nm}{\rightleftharpoons}} B^*$$

$$C_C^{TF1} + B^* \underset{k_{uTF2}}{\overset{k_{bTF2}}{\rightleftharpoons}} C_C^{TF1} \cdot B^* \triangleq C_C^{TF2}$$

$$C_C^{TF2} + RNAp \underset{k_{uRNApComplex}}{\overset{k_{bRNApComplex}}{\rightleftharpoons}} C_C^{TF2} \cdot RNAp \triangleq C_C^{RNAp}$$

$$gC + RNAp \underset{k_{uRNAp}}{\overset{k_{bRNAp}}{\rightleftharpoons}} gC \cdot RNAp \triangleq C_C^{RNAp,0}$$

$$C_C^{RNAp} \xrightarrow{k_{mC}} RNAp + C_C^{TF2} + mC$$

$$C_C^{RNAp,0} \xrightarrow{k_{mC}} RNAp + gC + mC$$

$$mC \xrightarrow{k_C} mC + C$$

$$mC \xrightarrow{d_{mC}} \emptyset$$

$$C \xrightarrow{d_C} \emptyset$$

3.1.2. ASSUMPTIONS

Considerations in the model:

1. The c_{RNAp} constant considers that the cell has the sufficient free RNAp in excess to be utilized by all the active genes that are transcribing simultaneously in the cell, including the gene of interest. Under this conception, the free RNAp vary in an almost unappreciable way in time, so can be defined as the $c_{RNApFree}$ constant and consequently the sum of the RNAp linked to the DNA and the free RNAp as the c_{RNAp} constant.
2. The RNAp binding-unbinding reactions to the promoter are much faster than the elongation and degradation reactions, so can be considered in the equilibrium state.
3. Transcription reaction is faster than translation reaction, so can be considered in the equilibrium state.
4. The conformation change is instantaneous.

3.1.3. FINAL EQUATIONS

After a mathematical development (download here for more information), we obtained the following equations, which define the constitutive and regulated expression respectively (where sub P is a generalization to name the protein).

$$\dot{y}_P = k_P \cdot \frac{k_{mPe} \cdot c_{nP}}{d_{mP}} - d_P \cdot y_P$$

$$\dot{y}_P = K_P \cdot \frac{\alpha_0 + \alpha_1 \cdot y_1 \cdot y_2}{\beta_0 + \beta_1 \cdot y_1 + \beta_2 \cdot y_1 \cdot y_2} - d_P \cdot y_P$$

Constitutive expression production is directly proportional to the translation rate (k_P), to the transcription effective rate (k_{mPe}), to the gene copy number (c_{nP}) and inverse proportional to the mRNA degradation rate (d_{mP}). The protein degradation is defined by the protein degradation rate in the cellular medium.

Regulated expression is proportional to the K_p (proportional to the translation rate (k_p), to the transcription rate (k_m), to the gene copy number (c_{nP}) and inverse proportional to the mRNA degradation rate (d_{mP}).

$$K_P = k_P \cdot \frac{k_{mP} \cdot c_{nP}}{d_{mP}}$$

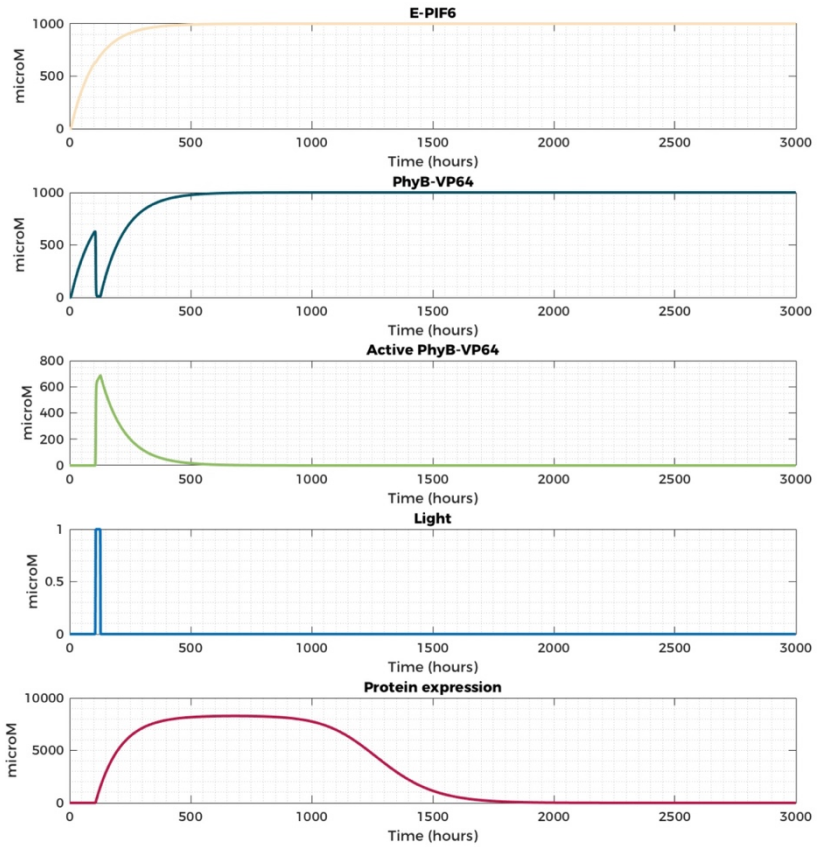
Alphas and betas are defined by rates (see supplementary). When red light is off (y_2=0), the expression becomes as:

$$\dot{y}_3 = K_P \cdot \frac{\alpha_0}{\beta_0 + \beta_1 \cdot y_1} - d_P \cdot y_3$$

The basal expression depends on the E-PIF6 production. With the model obtained we discovered that the leakage can be reduced with a strong promoter on E-PIF6.

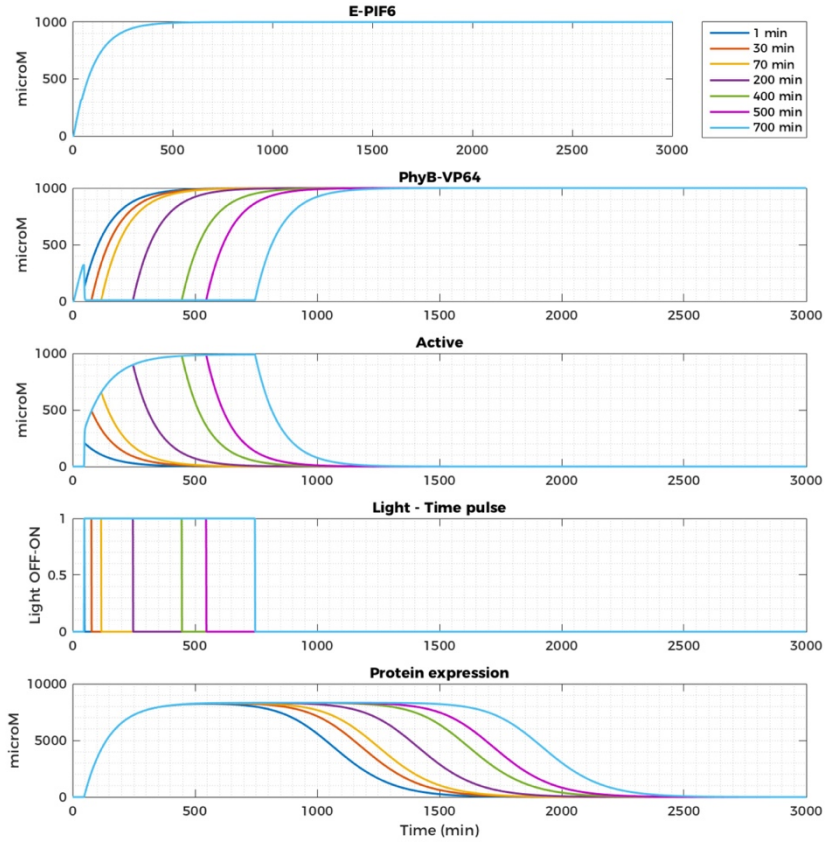
3.1.4. SIMULATIONS AND CONCLUSIONS

The optogenetic model is implemented in MATLAB. We start simulating for a pulse red light of 20 minutes.



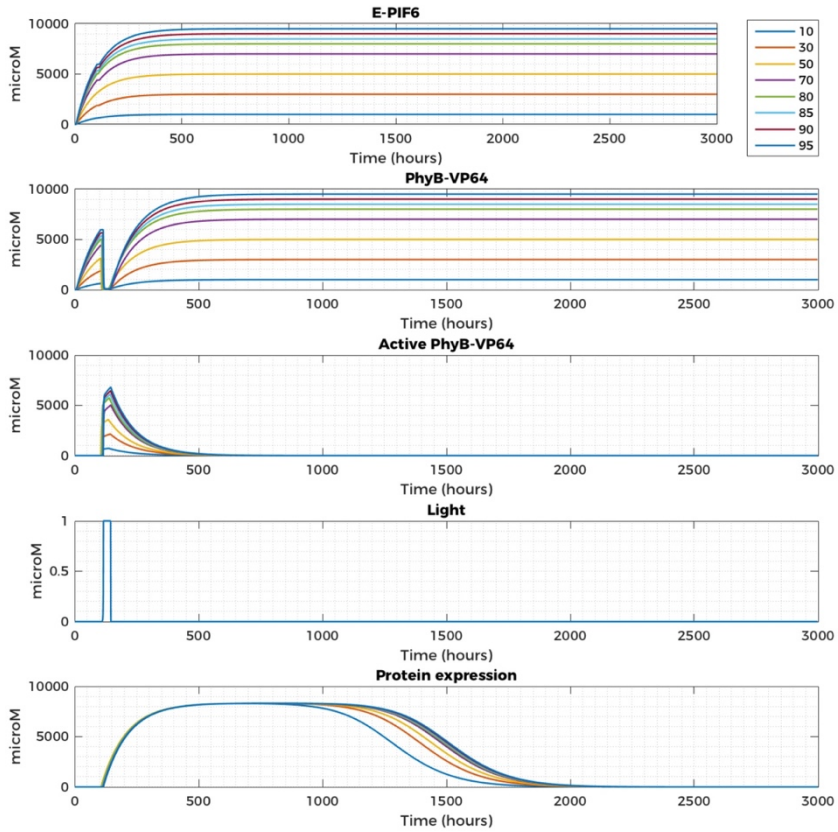
E-PIF6 is expressing constitutively as well as PhyB-VP64. When red light is ON, as we considered the conformation change instantaneously, the active PhyB-VP64 takes the current concentration from the inactive form and continues increasing because of its constitutive expression. Therefore, the desired output protein starts expressing. When red light turns off, the active PhyB-VP64 degrades with its degradation rate, while the inactive form continues expressing. Since the switch is off, the output protein total expression stops (we have to consider both basal and active form degrading as they are transcriptional factors) and fall down by its own degradation rate.

To study and to optimize energetic resources in our hardware, we simulate the dynamic optogenetic circuit's behavior for a different values of light pulse time, taking as consideration that LEDs are in their maximum power thus the experimental optogenetic data was obtained in these conditions, and our model is characterized from these data.



When the light pulse time is greater, the protein from optogenetic circuit remains during more time. The difference between 1 minute and 30-70 minutes is small, nonetheless, compared with the range of 400-700 minutes, the difference is evident. Depends on the controller's decision and the needs in the specific time how many time has to be activated the LEDs system.

As we thought not only depending on the energetic resources but also on the genetic constructions, a simulation at different gene copy number both E-PIF6 and PhyB-VP64 has done.



As can be seen in the graph, from 70 to 95 gene copy number the result in the desired output expression is identical. So for a genetic efficiency, the variation of gene copy number has to be in the range from 1 to 70.

Here below there is a resume table of all reaction rates, where P is a generalization to name the protein.

MODULE	REACTION NAME	RATE	VALUE	UNITS	SOURCE
CONSTITUTIVE	Binding/unbinding RNAp to promoter	k_{bRNAp}	100	$\mu M^{-1} \cdot min^{-1}$	Assumed
		k_{uRNAp}	10	min^{-1}	Assumed
	Transcription	k_{mP}	6	min^{-1}	Assumed
	Translation	k_p	1	min^{-1}	Assumed
	mRNA degradation	d_{mP}	0.1	min^{-1}	Assumed
	Protein degradation	d_P	0.01	min^{-1}	Zúrich 2014 iGEM
REGULATED	Binding/unbinding transcriptional factor 1 (TF1) to operator	k_{bTF1}	100	$\mu M^{-1} \cdot min^{-1}$	Assumed
		k_{uTF1}	10	min^{-1}	Assumed
	Binding/unbinding transcriptional factor 2 (TF2) to operator-TF1 complex	k_{bTF2}	100	$\mu M^{-1} \cdot min^{-1}$	Assumed
		k_{uTF2}	10	min^{-1}	Assumed
	Binding/unbinding RNAp to operator-TF1-TF2 complex	$k_{bRNApComplex}$	100	$\mu M^{-1} \cdot min^{-1}$	Assumed
		$k_{uRNApComplex}$	10	min^{-1}	Assumed
	Binding/unbinding RNAp to operator	k_{bRNAp}	10	$\mu M^{-1} \cdot min^{-1}$	Assumed
		k_{uRNAp}	100	min^{-1}	Assumed
	Transcription	k_{mP}	10	min^{-1}	Assumed
	Translation	k_P	1	min^{-1}	Assumed
	mRNA degradation	d_{mP}	0.1	min^{-1}	Assumed
	Protein degradation	d_p	0.01	min^{-1}	Zúrich 2014 iGEM

MODULE	REACTION NAME	PARAMETER	VALUE	UNITS	SOURCE
RECOMBINASE INTERACTION	PhiC31 dimerization rate	k_{bPhi}	60	$\mu M^{-1} \cdot min^{-1}$	Smith, M. C. M. et.al.
	Dimerization dissociation constant PhiC31.	K_{ii}	0.3	µM	Smith, M. C. M. et.al.
	PhiC31 undimerization rate	$k_{uPhi} = k_{bPhi} \cdot K_{ii}$	18	min^{-1}	Smith, M. C. M. et.al.

RECOMBINASE-REPORTER INTERACTION	PhiC31 binding to BP reporter rate	k_{bPhiBP}	60	$\mu M^{-1} \cdot min^{-1}$	Smith, M. C. M. et.al.
	Equilibrium constant PhiC31-reporter BP	K_{b1}	50	μM^{-1}	Smith, M. C. M. et.al.
	PhiC31 unbinding to BP reporter rate	$k_{uPhiBP} = k_{bPhiBP} / K_{b1}$	1.2	min^{-1}	
	Recombination rate	k_r	1	min^{-1}	Smith, M. C. M. et.al.
	PhiC31 unbinding to LR reporter	k_{uPhiLR}	10	$\mu M^{-1} \cdot min^{-1}$	Smith, M. C. M. et.al.
	Equilibrium constant PhiC31-reporter LR	K_{b2}	0.0047	μM	Smith, M. C. M. et.al.
	PhiC31 biding to LR reporter	$k_{bPhiLR} = K_{b2} \cdot k_{uPhiLR}$	0.047	min^{-1}	[1]
GP3-REPORTER INTERACTION	Gp3 binding to BP reporter rate	k_{bGP3LR}	60	$\mu M^{-1} \cdot min^{-1}$	Smith, M. C. M. et.al.
	Equilibrium constant Gp3-reporter BP	K_{b3}	40	μM^{-1}	Smith, M. C. M. et.al.
	Gp3 unbinding to BP reporter rate	$k_{uGP3LR} = k_{bGP3LR} / K_{b3}$	1.5	min^{-1}	Smith, M. C. M. et.al.
	Recombination rate	k_r	1	min^{-1}	Smith, M. C. M. et.al.
	Gp3 unbinding to BP reporter	k_{uGP3BP}	60	$\mu M^{-1} \cdot min^{-1}$	Smith, M. C. M. et.al.
	Equilibrium constant Gp3-reporter BP	K_{b4}	0.0053	μM	Smith, M. C. M. et.al.
	Gp3 biding to BP reporter	$k_{bPhiLR} = K_{b2} \cdot k_{uPhiLR}$	0.0047	min^{-1}	Smith, M. C. M. et.al.

3.1.2. MODEL AND CONSIDERATIONS

For a better explanation of the mathematical's model, we establish the dynamic balances for each module separately for finally couple them in the circuit's design.

3.1.2.1. CONSTITUTIVE MODULE

Chemical species are defined by the next variables:

$$[gA] = x_1; \quad [RNAp] = x_2; \quad [gA \cdot RNAp] = x_3; \quad [mA] = x_4; \quad [A] = x_5$$

Reactions occur at speeds proportional to the product of the reactant concentrations and to the proportionality factor, known as the reaction rate:

$$v_1 = k_{1A} \cdot x_1 \cdot x_2; \quad v_{-1} = k_{-1A} \cdot x_3; \quad v_2 = k_{mA} \cdot x_3; \quad v_3 = k_A \cdot x_4; \quad v_4 = d_{mA} \cdot x_4; \quad v_5 = d_A \cdot x_5$$

Mass Action Kinetic gives:

$$\dot{x}_1 = -v_1 + v_{-1} + v_2 = -k_{1A} \cdot x_1 \cdot x_2 + k_{-1A} \cdot x_3 + k_{mA} \cdot x_3$$

$$\dot{x}_2 = -v_1 + v_{-1} + v_2 = -k_{1A} \cdot x_1 \cdot x_2 + k_{-1A} \cdot x_3 + k_{mA} \cdot x_3$$

$$\dot{x}_3 = +v_1 - v_{-1} - v_2 = k_{1A} \cdot x_1 \cdot x_2 - k_{-1A} \cdot x_3 - k_{mA} \cdot x_3$$

$$\dot{x}_4 = +v_2 - v_4 = k_{mA} \cdot x_3 - d_{mA} \cdot x_4$$

$$\dot{x}_5 = +v_3 - v_5 = k_A \cdot x_4 - k_A \cdot x_5$$

Since:

$$\dot{x}_1 + \dot{x}_3 = 0$$

$$\dot{x}_2 + \dot{x}_3 = 0$$

we can define the next two constants:

$$x_1 + x_3 = c_{nA}$$

$$x_2 + x_3 = c_{RNAp}$$

The first invariant defines the gene copy number as the sum of the DNA free and DNA linked to the RNApol.

The second invariant assumes that the cell has the sufficient free RNAp in excess to be utilized by all the active genes that are transcribing simultaneously in the cell, including the gene of interest. Under this conception, the free RNAp vary in an almost inappreciable way in time, so can be defined as the $c_{RNApFree}$ constant and consequently the sum of the RNAp linked to de DNA and the free RNAp as the c_{RNAp} constant.

$$x_2 \approx c_{FreeRNAp}$$

Now we can consider the variation of x_3 null because it is defined as the result of constants difference, as x_3 is defined as the result of constants difference, we can consider its variation null:

$$x_3 = c_{RNAp} - c_{FreeRNAp}$$

$$0 = k_{1A} \cdot x_1 \cdot x_2 - k_{-1A} \cdot x_3 - k_{mA} \cdot x_3$$

$$x_3 = \frac{x_1 \cdot x_2}{\frac{k_{-1A} + k_{mA}}{k_{1A}}}$$

Considering that:

$$x_1 = c_{nA} - x_3$$
$$x_2 \approx c_{FreeRNAp}$$

we obtain

$$x_3 = \frac{(c_{nA} - x_3) \cdot c_{FreeRNAp}}{\frac{k_{-1A} + k_{mA}}{k_{1A}}} = \frac{c_{nA}}{\frac{1}{c_{FreeRNAp}} \cdot \frac{k_{-1A} + k_{mA}}{k_{1A}}} - \frac{x_3}{\frac{1}{c_{FreeRNAp}} \cdot \frac{k_{-1A} + k_{mA}}{k_{1A}}}$$

$$x_3 = \frac{c_{nA}}{1 + \frac{1}{c_{FreeRNAp}} \cdot \frac{k_{-1A} + k_{mA}}{k_{1A}}}$$

From the transcription balance and x_3:

$$\dot{x}_4 = k_{mA} \cdot \frac{c_{nA}}{1 + \frac{1}{c_{FreeRNAp}} \cdot \frac{k_{-1A} + k_{mA}}{k_{1A}}} - d_{mA} \cdot x_4$$

where the effective transcription rate is defined as:

$$k_{mAe} = \frac{k_{mA}}{1 + \frac{1}{c_{FreeRNAp}} \cdot \frac{k_{-1A} + k_{mA}}{k_{1A}}}$$

The model shows that de mRNA synthesis increases if the promoter has much affinity for the RNAp (k_{1A} high) up to a limit given by the rate of transcription and the gene copy number. In addition, it is observed that if there are many cellular processes consuming free RNAp, the synthesis mRNA will decrease.

Assuming that the transcription reaction is faster than the translation reaction, we obtain:

$$\dot{x}_4 = 0$$
$$0 = k_{mAe} \cdot c_{nA} - d_{mA} \cdot x_4$$
$$x_{4eq} = \frac{k_{mAe} \cdot c_{nA}}{d_{mA}}$$

Finally, the protein expression of the constitutive module can be defined as comes:

$$\dot{x}_5 = k_A \cdot x_4 - d_A \cdot x_5 = k_A \cdot \frac{k_{mAe} \cdot c_{nA}}{d_{mA}} - d_A \cdot x_5$$

where the effective translation rate is defined as:

$$k_{Ae} = k_A \cdot \frac{k_{mAe}}{d_{mA}}$$

$$gC + x_1 \underset{k_{uTF1}}{\overset{k_{bTF1}}{\rightleftharpoons}} gC \cdot x_1 \triangleq C_C^{TF1}$$
$$B \underset{740nm}{\overset{660nm}{\rightleftharpoons}} B^*$$
$$C_C^{TF1} + x_2 \underset{k_{uTF2}}{\overset{k_{bTF2}}{\rightleftharpoons}} C_C^{TF1} \cdot x_2 \triangleq C_C^{TF2}$$
$$C_C^{TF2} + y \underset{k_{uRNAp}}{\overset{k_{bRNAp}}{\rightleftharpoons}} C_C^{TF2} \cdot y \triangleq C_C^{y}$$
$$gC + y \underset{k_{uRNAp}}{\overset{k_{bRNAp}}{\rightleftharpoons}} gC \cdot y \triangleq C_C^{y,0}$$
$$C_C^{y} \xrightarrow{k_{mC}} y + C_C^{TF2} + mC$$
$$C_C^{y,0} \xrightarrow{k_{mC}} y + gC + mC$$
$$mC \xrightarrow{k_C} mC + C$$
$$mC \xrightarrow{d_{mC}} \emptyset$$
$$x_3 \xrightarrow{d_C} \emptyset$$

$$\dot{C}_C^{TF1} = k_{bTF1} \cdot gC \cdot x_1 - k_{uTF2} \cdot C_C^{TF1} - k_{bTF2} \cdot C_C^{TF1} \cdot x_2 + k_{uTF2} \cdot C_C^{TF2}$$
$$\dot{C}_C^{TF2} = k_{bTF2} \cdot C_C^{TF1} \cdot x_2 - k_{uTF2} \cdot C_C^{TF2} - k_{bRNAp} \cdot C_C^{TF2} \cdot y + k_{uRNAp} \cdot C_C^{y} + k_{mC} \cdot C_C^{y}$$
$$\dot{C}_C^{y,0} = k_{bRNAp} \cdot gC \cdot y - k_{bRNAp} \cdot C_C^{y,0} - C_C^{y,0} \cdot k_{mC}$$
$$\dot{C}_C^{y} = k_{bRNAp} \cdot C_C^{TF2} \cdot y - k_{uRNAp} \cdot C_C^{y} - k_{mC} \cdot C_C^{y}$$

$$C_C^{TF1} = \frac{k_{bTF1}}{k_{uTF1}} \cdot gC \cdot x_1 = \frac{1}{\frac{k_{uTF1}}{k_{bTF1}}} \cdot gC \cdot x_1 \triangleq \frac{1}{k_{nC2}} \cdot gC \cdot x_1$$

$$C_C^{TF2} = \frac{k_{bTF2}}{k_{uTF2}} \cdot C_C^{TF1} \cdot x_2 = \frac{1}{\frac{k_{uTF1}}{k_{bTF2}}} \cdot C_C^{TF1} \cdot x_2 \triangleq \frac{1}{k_{nC3}} \cdot C_C^{TF1} \cdot x_2$$

$$C_C^{y,0} = \frac{k_{bRNAp}}{k_{uRNAp} + k_{mC}} \cdot gC \cdot y \triangleq \frac{1}{k_{y,0}} \cdot gC \cdot y$$

$$C_C^{y} = \frac{k_{bRNAp}}{k_{uRNAp} + k_{mC}} \cdot C_C^{TF2} \cdot y \triangleq \frac{1}{k_{y}} \cdot gC \cdot y \cdot \frac{x_1 \cdot x_2}{k_{nC2} \cdot k_{nC3}}$$

Total number of gene C is equal to free gene C and to gene occuped by TFs and RNAp:

$$gC + C_C^{TF1} + C_C^{TF2} + C_C^{y} + C_C^{y,0} = c_C$$

$$gC = \frac{c_C}{1 + \frac{y}{k_{y,0}} + \frac{x_1}{k_{nC2}} \left(1 + \frac{x_2}{k_{nC3}} \cdot \left(1 + \frac{y}{k_y}\right)\right)}$$

Considering transcription rate greater than translation rate:

$$\dot{m}_C = k_{mC} \cdot C_C^{y,0} + k_{mC} \cdot C_C^{y} - d_{mC} \cdot m_C$$

$$\dot{m}_C = 0$$

$$m_C = \frac{k_{mC}}{d_{mC}} \cdot (C_C^{y,0} + C_C^{y})$$

$$\dot{x}_3 = k_C \cdot m_C - d_C \cdot x_3$$

$$\dot{x}_3 = k_C \cdot \frac{k_{mC}}{d_{mC}} \cdot (C_C^{y,0} + C_C^{y}) - d_C \cdot x_3 = k_C \cdot \frac{k_{mC}}{d_{mC}} \cdot \left(\frac{1}{k_{y,0}} \cdot gC \cdot y + \frac{1}{k_y} \cdot gC \cdot y \left(\frac{x_1 \cdot x_2}{k_{nC2} \cdot k_{nC3}}\right)\right) - d_C \cdot x_3$$

$$\dot{x}_3 = k_C \cdot \frac{k_{mC}}{d_{mC}} \cdot c_C \cdot \frac{\frac{y}{k_{y,0}} + \frac{y}{k_y} \cdot \left(\frac{x_1}{k_{nC2}}\right) \cdot \left(\frac{x_2}{k_{nC3}}\right)}{1 + \frac{y}{k_{y,0}} + \frac{x_1}{k_{nC2}} \cdot \left(1 + \frac{x_2}{k_{nC3}} \cdot \left(1 + \frac{y}{k_y}\right)\right)} - d_C \cdot x_3$$

$$\dot{x}_3 = K_C \cdot \frac{\alpha_0 + \alpha_1 \cdot x_1 \cdot x_2}{\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_1 \cdot x_2} - d_C \cdot x_3,$$

$$K_C = k_C \cdot \frac{k_{mC}}{d_{mC}} \cdot c_C$$

$$\alpha_0 = \frac{y}{k_{y,0}}$$

$$\alpha_1 = \frac{y}{k_y} \cdot \frac{1}{k_{nC2}} \cdot \frac{1}{k_{nC3}} = \beta_1 \cdot \frac{y}{k_y} \cdot \frac{1}{k_{nC3}}$$

$$\beta_0 = 1 + \alpha_0$$

$$\beta_1 = \frac{1}{k_{nC2}}$$

$$\beta_2 = \frac{1}{k_{nC2}} \cdot \left(\frac{1}{k_{nC3}} \cdot \left(1 + \frac{1}{k_y} \cdot y \right) \right) = \beta_1 \cdot \left(\frac{1}{k_{nC3}} + \frac{\alpha_1}{\beta_1} \right) = \frac{\beta_1}{k_{nC3}} + \alpha_1$$

3.2. STOCHASTIC

Like we know, all biological systems are affected by inherent noise coming from our environment. Not only at the macroscopic scale but also in the microscopic scale. Our genes expression depends on a lot of factors that define how our world evolves.

So, the noise is a property of all biological systems and we have to be concerned about that. To take into account the randomness associated to natural system, including plants cells, we had the necessity of use stochastic models.

With the objective of obtain a more realistic description of our optogenetic circuit, we decided to generate a stochastic model of our constructions. Due to the uncertainly related to these genetic modules, we decided to develop a model capable of going further than deterministic model.

To continue, we are going to look over the theory behind this type of models and then we will be able to understand all about these curious models, including our own results.

In stochastic models, people usually works with the chemical master equations (CME) which try to explain the evolution of a system with probabilistic rates. This master

equations are groups of differential equations that describe the switching of each chemical. Unfortunately, CME only can be solved in some practical cases (simple and linear cases). The CME expression can be obtained defining all balance equation for each probabilistic distribution.

In these cases, in instead of using deterministic reaction rates, we use probabilistic reaction rates. These rates define the probability per time unit of one chemical changes to other or disappear.

3.2.1. EXTENSION FROM DETERMINISTIC MODEL

Let´s see it with an example:

We have the next reactions for transcription.

$$\text{DNA} \xrightarrow{r} \text{DNA} + \text{mRNA}$$
$$\text{mRNA} \xrightarrow{d} \emptyset$$

The probability of transcription is ***r*** and the probability of mRNA degradation is ***d***. If we consider more than one gene copy, we would have a probability equals to ***number of copies * r***, instead of ***r*** and the same for ***d***. Now, we search for the expression of the probability of having ***n*** copies of mRNA at ***t+∂t*** time. We consider all the possible exclusive events that could occur in that reaction:

- Have ***n-1*** copies and the reaction of production occurs in ***∂t***.
- Have ***n+1*** copies and the reaction of degradation occurs in ***∂t***.
- Have ***n*** copies and no reaction.

So, the probability of a specific reaction is the product between the probabilistic reaction rate and the time that have been consumed. The probability mentioned before will be:

$$p(n, t+\delta t) = p(n-1, t) \cdot r \cdot \delta t + p(n+1, t) \cdot (n+1) \cdot d \cdot \delta t + p(n, t) \cdot [1 - n \cdot d \cdot \delta t - r \cdot t]$$

Considering an infinitesimal time fraction, the CME would be:

$$\frac{\delta p(n,t)}{\delta t} = d[p(n+1,t)(n+1) - p(n,t)n] + r[p(n-1,t) - p(n,t)]$$

So, we would obtain an ODE equation for each possible value of n (copies of chemical). Solve the previous equation is a computational challenge, however there are solving methods that allow us to obtain equal results without that resource consume.

It is important to note that the mean of solutions in stochastic model coincides with the deterministic dynamic of the mean concentration. We also use the Chemical Langevin Equation that approximates the CME with stochastic differential equations of an order equal or greater than species number.

Assuming a ***∂t*** time period, we suppose that the reaction will occur several times between t and t+***∂t***. The number of reactions in a ***∂t*** period is equal to the product between probability of have n copies of mRNA in t time with the probabilistic reaction rate (p(n,t)*r - reaction propension) and the time period ***∂t***. Like biological reactions are unpredictable, we consider a Poisson probabilistic distribution to describe the number of times that occurs a reaction in a ***∂t*** period.

The ʎ parameter of Poisson distribution (expected events) are equals to the reaction frequency (p(n,t)*r*∂t) . Langevin approximation consider ʎ sufficiently high to assume a normal distribution with mean ʎ and variance ʎ, instead of Poisson distribution.

Finally, the number of reactions in ***∂t*** in a specific reaction will be:

$$n(\delta t) = \lambda + \sqrt{(\lambda)N(0,1)}$$

Which is the mean of possible reaction in ***∂t*** plus the uncertainly associated (product of standard deviation in the number of reactions and gaussian normal, describing noise with mean 0 and standard deviation 1).

Therefore, the sum of copies of ARNm in our cell in t+***∂t*** could be approximated by the next expression:

$$n_{mRNA}(t+\delta t) = n_{mRNA}(t) + (r - dn_{n_{mRNA}})\delta t + \\ + \sqrt{r \cdot \delta t} N_1(0,1) - \sqrt{d \cdot \delta t \cdot n_{mRNA}} N_2(0,1)$$

This equation can be generalized for any chemical expression in our cell. Only we would have to change the value of probabilistic rates like r and d. Also, we have to consider different normal distribution on each reaction.

Constitutive module

In all modules we have on the one hand the mRNA production and on the other hand the protein production that depends on the mRNA production. Obviously, each protein is associated with a different mRNA.

In the constitutive one, we have the next equations:

$$mRNA(k) = mRNA(k-1) + (km \cdot cn - dm \cdot mRNA(k-1)) \cdot \delta t + \\ + \sqrt{km \cdot cn \cdot \delta t} N(0,1) - \sqrt{dm \cdot mRNA(k-1) \cdot \delta t} N(0,1)$$

The probabilistic production rate is composed of the transcription rate and the number of copies of our gene $(\mathrm{km} \cdot \mathrm{cn})$. Obviously, to obtain the number of produced chemicals we have to obtain the product of the produced chemicals in one gene and the number of genes for that reaction.

The transcription rate is the same as in deterministic model:

$$km = \frac{k_{mA}}{\left(1 + \frac{1}{cRNA_{polfree}}\right)\left(\frac{k_{-1A}+k_{mA}}{k_{1A}}\right)}$$

So, the equation for the production of our constitutive proteins is:

$$C(k) = C(k-1) + (K \cdot mRNA(k-1) - dp \cdot C(k-1))\,\delta t + \\ + \sqrt{K \cdot mRNA(k-1) \cdot \delta t} N(0,1) - \sqrt{dp \cdot C(k-1) \cdot \delta t} N(0,1)$$

Like we could see is the same structure as in the previous equation. In this case, the production depends on the amount of mRNA and the degradation depends on the quantity of protein.

We would have these two previous equations for each constitutive protein. In our case, we have four equations: two for E-PIF6 and two for PhyB-VP64.

For the PhyB-VP64 active form, we defined a new chemical in our model with the next equation:

$$PhyB - VP64_{active} = PhyB - VP64(k-1) - dp \cdot PhyB - VP64(k-1) \cdot t - \\ - \sqrt{dp \cdot PhyB - VP64(k-1) \cdot \delta t} N(0,1)$$

We have only considered the degradation of this chemical because its formation depends on the light stimulus. In our model we can define when we want to apply this stimulus and during how many time. If the light is on, all the inactivated PhyB_VP64 change its conformation (be transformed to activated PhyB_VP64) instantaneously and when we turn the light off, inactivated PhyB_VP64 starts to be increased.

Regulated module

For our target protein in our optogenetic circuit we have also two equations, one for the ARN and one for itself. In this case, the complexity come defined by the production rates. Following, we introduce both equations:

$$mRNA_{TP}(k) = mRNA_{TP}(k-1) + (Kk \cdot cn - dm \cdot mRNA_{TP}(k-1)) \cdot \delta t + \\ + \sqrt{Kk \cdot cn \cdot \delta t} N(0,1) - \sqrt{dm \cdot mRNA_{TP}(k-1) \cdot \delta t} N(0,1)$$

This Kk probabilistic production depends on EPIF6 and activated PhyB_VP64 rate is defined by:

$$Kk(k) = \frac{K_C \cdot \frac{y}{k_{y0}} + K_C \cdot \frac{y}{k_y \cdot k_{nC2} \cdot k_{nC3}} \cdot EPIF6(K-1) \cdot PhyB - VP64_{activated}(k-1)}{1 + \frac{y}{k_{y0}} + \frac{EPIF6(k-1)}{k_{nC3}}\left(1 + \frac{PhyB - VP64_{activated}(k-1)}{k_{nC3}}\left(1 + \frac{y}{k_y}\right)\right)}$$

The final protein C equation would be the same as before:

$$C(k) = C(k-1) + (K \cdot mRNA_C(k-1) - dp \cdot C(k-1))\,\delta t + \\ + \sqrt{K \cdot mRNA_C(k-1) \cdot \delta t} N(0,1) - \sqrt{dp \cdot C(k-1) \cdot \delta t} N(0,1)$$

3.2.2. SIMULATIONS AND CONCLUSIONS

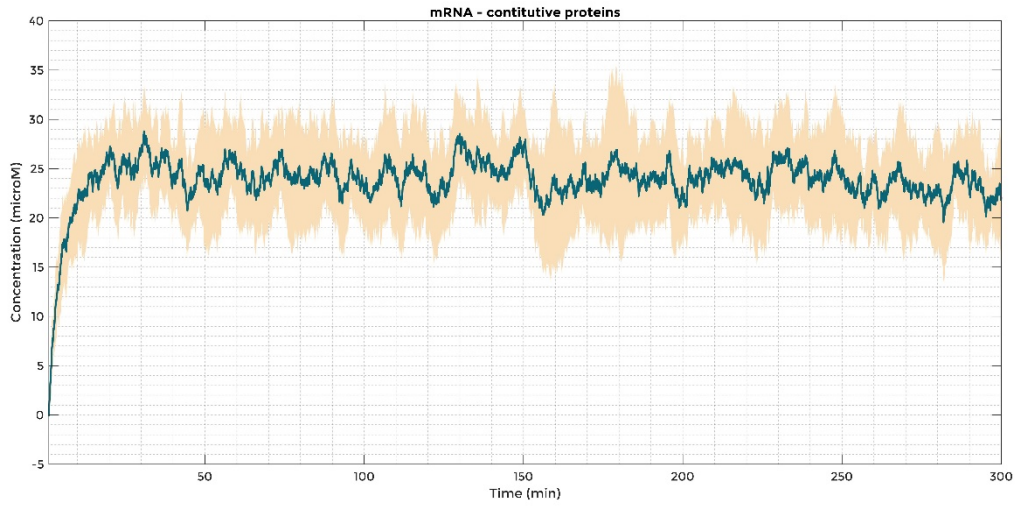


Figure 1. mRNA expression in constitutive module

In this figure we can see the variation associated to the expression of mRNA for constitutive proteins in our system due to the gaussian noise added to the expression.

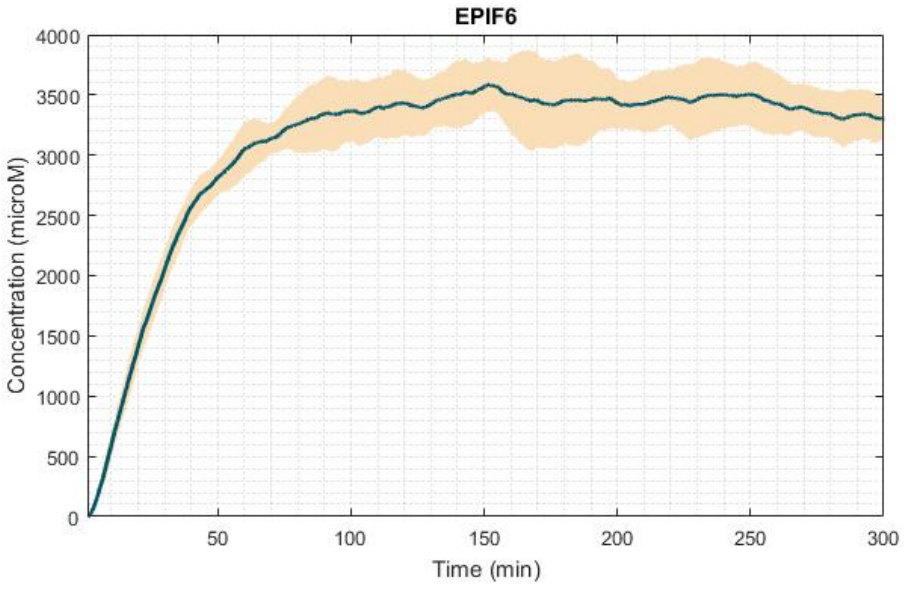


Figure 2. EPIF6 and constitutive expression

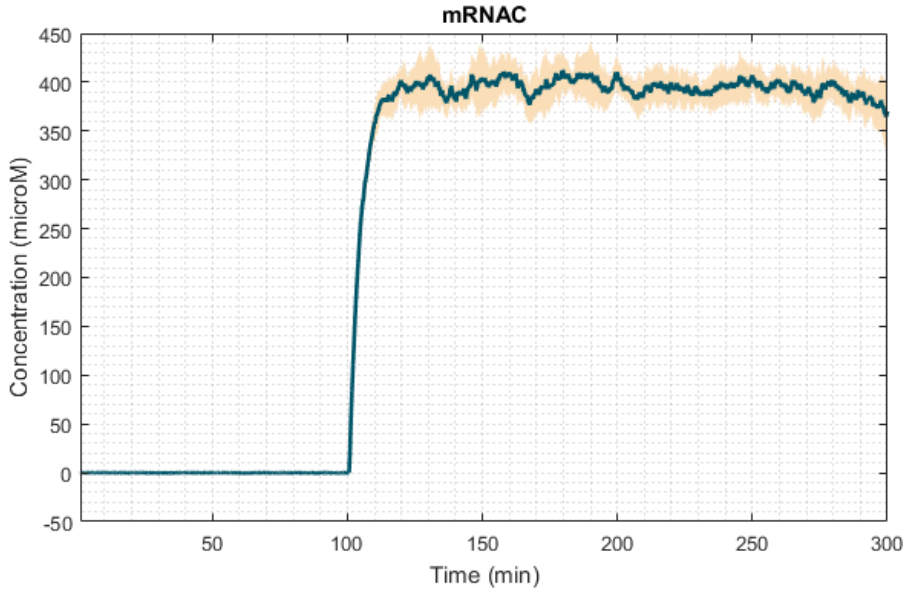


Figure 3. mRNA expression in regulated module

Due to the oscillating production of mRNA we obtain a similar variation on the production of the protein that depends on it.

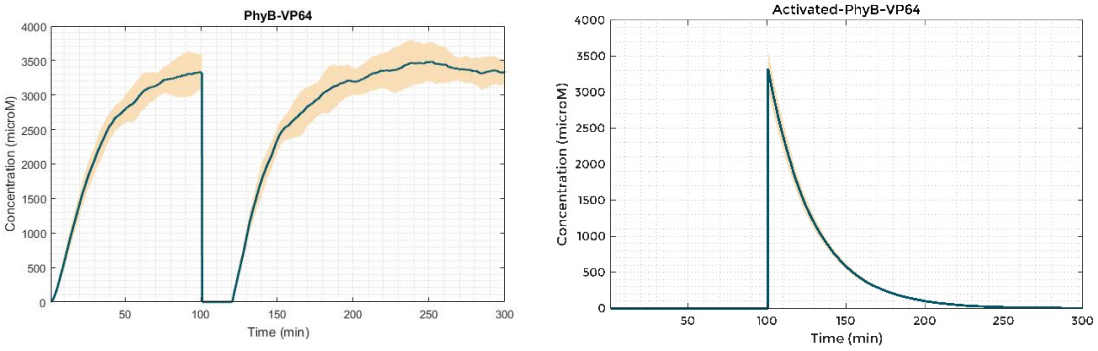


Figure 4 and 5. PhyB-VP64, activated PhyB-VP64 evolution

In this image, we realize about the behavior of our model when we put the red light on. All the inactivated chimeric protein (PhyB-VP64) changes its conformation instantaneously because of its dynamic is fast enough to assume it.

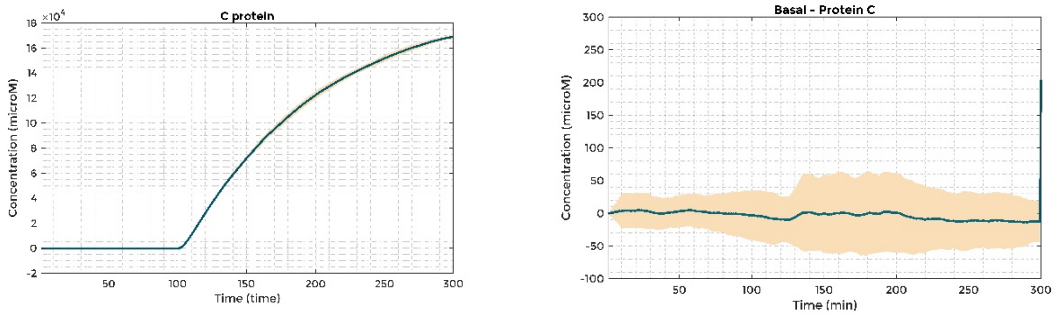


Figure 6 and 7. C protein regulated expression and basal expression.

We can confirm visually the dependency between the production of our target protein (C) and the presence of both activated PhyB-VP64 and EPIF6. In this case, we simulate a light stimulus in minute 100 and with a duration of 20 min. It should be noted that apparently our optogenetic construction has a filter behavior. We can see that the noise variation in RNA messenger is not showed on the production of our target protein (C).

3.4. OPTIMIZATION AND VALIDATION

For the optimization of our optogenetic model we have used luciferase results obtained in lab. The optimization consists of comparing the values obtained with our model with empirical values. For that, we implemented a genetic algorithm to generate random values for the parameters of our model. Some of them are constricted to a fixed value and others are constricted to a gap of values ensuring an optimal and logical solution.

Also, we found an optimal seed (starting point) to calculate the value of our parameters. Then the algorithm applied several iterations (generations) to reduce the relative error between experimental data and model data.

In order to have an easily comprehension of changes in our model when we change parameters value, we decided to simplify the model taking into account some assumptions.

$$\dot{x_4} = K_C \frac{\frac{y}{k_{y0}} + \frac{y}{k_y} \cdot \frac{x_1}{k_{nC2}} \cdot \frac{x_3}{k_{nC3}}}{1 + \frac{y}{k_{y0}} + \frac{x_1}{k_{nC2}} \left(1 + \frac{x_2}{k_{nC3}} \left(1 + \frac{y}{k_y}\right)\right)} - d_c \cdot x_4$$

Figure x. Final equation of optogenetic circuit

$$\dot{x_4} = K_C \frac{\alpha_0 + \frac{x_1 \cdot x_3}{K}}{\alpha_0 + 1 + \frac{x_1 \cdot x_3}{K}} - d_c \cdot x_4$$

$$\dot{x_4} = K_C \frac{K \cdot \alpha_0 + x_1 \cdot x_3}{K + x_1 \cdot x_3} - d_c \cdot x_4$$

Figure x. Simplification of our model in two steps.

Then we proved that these assumptions didn't eliminate simulation capability of our model and we ensure that we didn't loose of information.

Finally, we obtained these results:

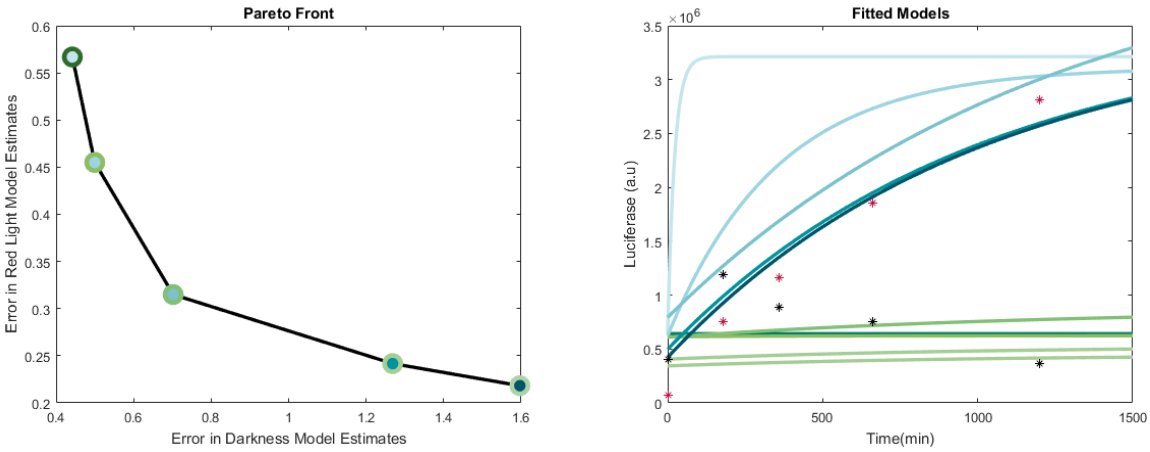


Figure 7. Here we can see how the algorithm defined the Pareto front with different errors in the constitutive expression (darkness) and in the regulated one (red light presence).

In the image, lines represent the model fitted to the points (*), which are results obtained in this experiment. In this case, parameters correspond to the set which minimizes the red light model relative error. The election of any other set of parameters is possible and responds to different subjective criteria, which means that there is not a unique optimal solution.

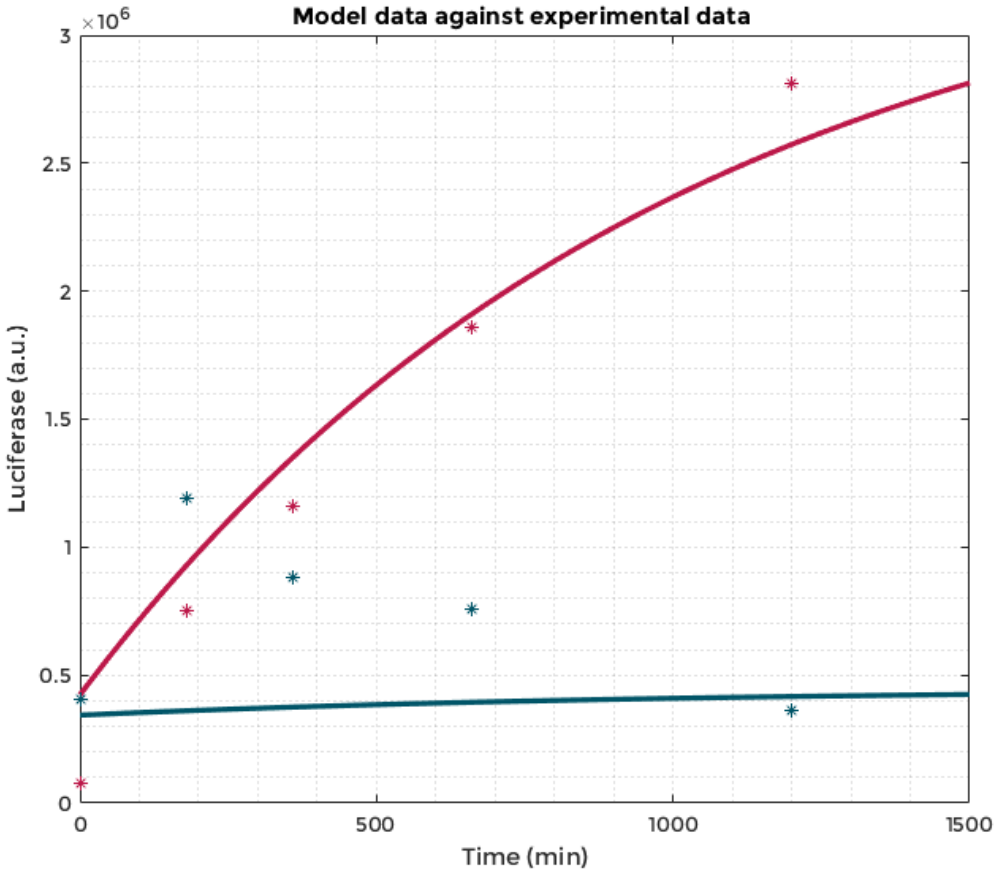


Figure 8. Representation of experimental data (points) and adjusted model (lines).

Therefore, we got a model able to predict the expression of a target protein in our optogenetic construction. The following table contains estimated values of parameters according to different importance criteria.

REACTION NAME	RATE	VALUE	VALUE	VALUE	VALUE	VALUE	UNITS
Production rate	K_{EPIF}	1000	1000	805,571598	1000	50	1/min
Copy number	n_{EPIF}	200	200	200	200	200	copies
Degradation rate	d_{EPIF}	0,28087894	0,2289368	0,2159143	0,34827246	0,34827246	1/min
Production rate	K_{PhyB}	342,661465	892,530021	930,132781	964,169666	494,420677	Molecules/min
Copy number	n_{PhyB}	200	200	200	200	200	copies
Degradation rate PhyB	d_{PhyB}	0,001	0,001	0,15215729	0,001	0,37161292	1/min
Proportion of production with red light	k_{change}	0,1	3,80769205	24,9399062	0,1	42,4837085	No dimension
Degradation rate activated PhyB-VP64	d_{act}	0,2434054	0,22550566	0,12431998	0,00150999	0,001	1/min
Luciferase translation rate	K_c	145773,044	8826,19341	3500	3500	3500	1/min
ARN polimerase	y	1	1	1	1	1	number
Basal binding rate	k_{y0}	1	1	1	1	1	No dimension
Binding rate polymerase – translation complex	k_y	1	1	1	1	1	No dimension
Binding rate – EPIF-6 operator	K_{nc2}	1	1	1	1	1	No dimension
Binding rate – activated PhyB-VP64 , EPIF6	K_{nc3}	1	235,917311	610,375621	1	749,916619	No dimension
Target protein degradation	d_c	0,04536059	0,00283412	0,0008	0,001	0,001	1/min
Variables compound	$alpha$	0,2	0,2	0,2	0,15040884	0,12793009	1/min

Figure 9. Table with different parameters values for each point on Pareto front from

3.5. CONCLUSION

- Human plant communication is feasible and has been confirmed by our in vivo and in silico results.

- To optimize plant response to red light, maintaining the same gene copy number for PhyB-VP64 and E-PIF6, translation rates for PhyB-VP64 need to be 8-fold higher than E-PIF's.

- Given an importance criteria, the model could be used to tune the LEDs in order to obtain a desired amount of protein and economize energy resources.

4. PLANT-HUMAN: SENSOR CIRCUIT

A constitutive module expresses the integrase-recombinase PhiC31 which recombines the specific attachment sites of the reporter in BP state to LR state (BxP → LxR reaction, figure X), keeping this state because of the constitutive expression.

When the dispenser puts on dexamethasone to the plant, the integrase-excisionase gp3 (also called recombination directionality factor - RDF) expresses through the transitory effect of the dexamethasone. Gp3 with the presence of PhiC31 recombines the specific attachment sites to BP state, then the inducible promoter is addressed to produce the transcription product of the specific color under the respective stress inductor that switch on the expression in the reporter assembly.

For more information, visit the biological design.

4.1. DETERMINISTIC

4.1.1. REACTIONS

For this part of Chatterplant we have made two models to obtain information about the performance of the PhiC31 recombinase and Gp3. At first, we decided to produce a simple model to understand how PhiC31 works and how does the reporter assembly acts with the performance of PhiC31. For this part we consider these reactions:

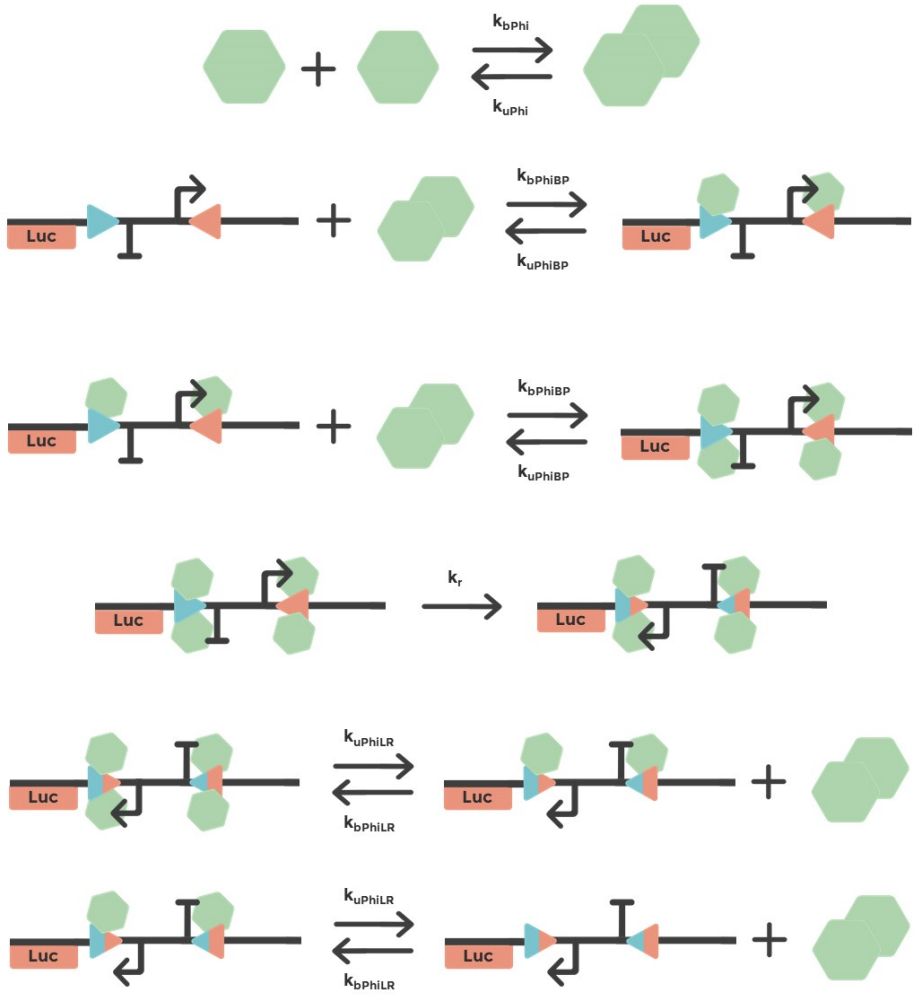


Figure x. Graphic description of the behavior of our PhiC31 simple model with reaction parameters.

In the image we can see that PhiC31 (green hexagon) forms dimers and then each dimer joint with our register assembly. In this case, the experiments in lab used luciferase so our model is inspired on them. When PhiC31 is connected with our register, it acts and recombines the sites. Then PhiC31 leaves the register dimer by dimer. We would have expression of our target protein (luciferase) when the sites are recombined (LR).

To continue, we show the formal reactions that correspond to the previous image:

$$Phi + Phi \underset{k_{uPhi}}{\overset{k_{bPhi}}{\rightleftharpoons}} Phi_2$$

$$DNA_{BP} + Phi_2 \underset{k_{uPhiBP}}{\overset{k_{bPhiBP}}{\rightleftharpoons}} DNA_{BP-Phi2}$$

$$DNA_{BP-Phi2} + Phi_2 \underset{k_{uPhiBP}}{\overset{k_{bPhiBP}}{\rightleftharpoons}} DNA_{BP-Phi4}$$

$$DNA_{BP-Phi4} \xrightarrow{k_r} DNA_{LR-Phi4}$$

$$DNA_{LR-Phi4} \underset{k_{bPhiLR}}{\overset{k_{uPhiLR}}{\rightleftharpoons}} DNA_{LR-Phi2} + Phi_2$$

$$DNA_{LR-Phi2} \underset{k_{bPhiLR}}{\overset{k_{uPhiLR}}{\rightleftharpoons}} DNA_{LR} + Phi_2$$

Figure x. Reactions of PhiC31 – register assembly

Next, we show the diagram related to the actuation of PhiC31 and GP3 at the same time, also we have to consider the reactions that we explain before.

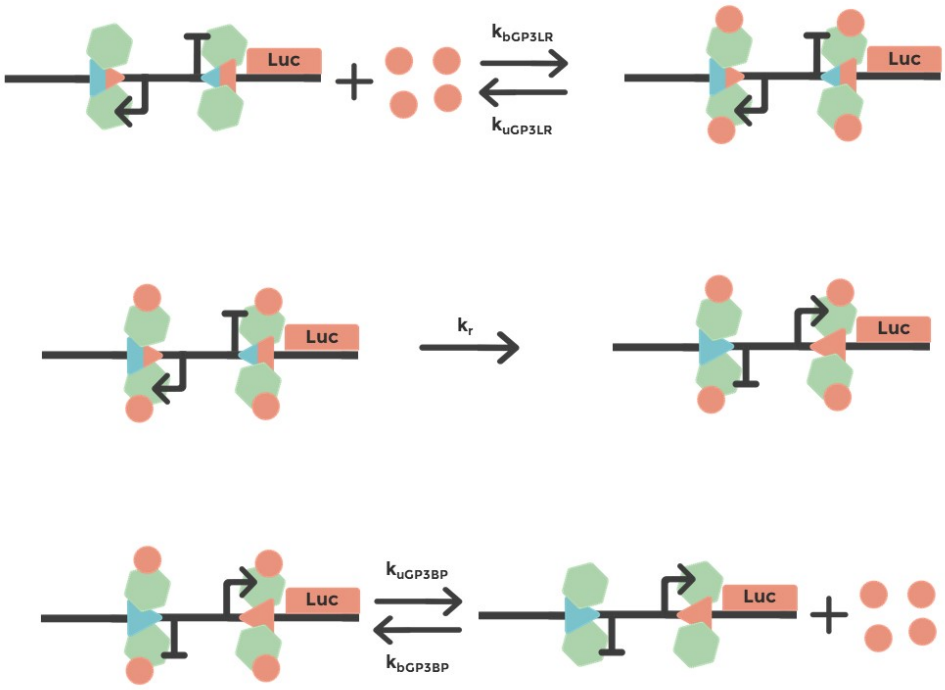


Figure x. Graphic representation of GP3 actuation

Like we can see, a simple model about the GP3 actuation was considered without take into account the interaction between free GP3 and free PhiC31 in intracellular media due to this model provides us with the same information as the complex one. For that we compared the behavior of our model with a very complex model (The mechanism of C31 integrase directionality:

experimental analysis and computational modelling - Alexandra Pokhilko, Jia Zhao et al., 2016).

$$DNA_{LR-Phi4} + 4 \cdot GP3 \underset{k_{uGP3LR}}{\overset{k_{bGP3LR}}{\rightleftharpoons}} DNA_{LR-Phi4-4GP3}$$

$$DNA_{LR-Phi4-4GP3} \xrightarrow{k_r} DNA_{BP-Phi4-4GP3}$$

$$DNA_{BP-Phi4-4GP3} \underset{k_{uGP3LR}}{\overset{k_{bGP3LR}}{\rightleftharpoons}} DNA_{LR-Phi4} + 4 \cdot GP3$$

Figure x. GP3 formal reactions

4.1.2. MODEL AND CONSIDERATIONS

In the literature, there are different approaches to model the recombinase-excisionase action. We started modeling theoretically basic supposed reactions to have a general knowledge as a first contact point. Then, studying the experimental results, we could better understand which reactions can be happening inside ChatterPlant.

For the integrase-recombinase action:

1. Monomeric PhiC31 doesn't bind to the reporter, only dimers of PhiC31.
2. The recombination performance is irreversible.

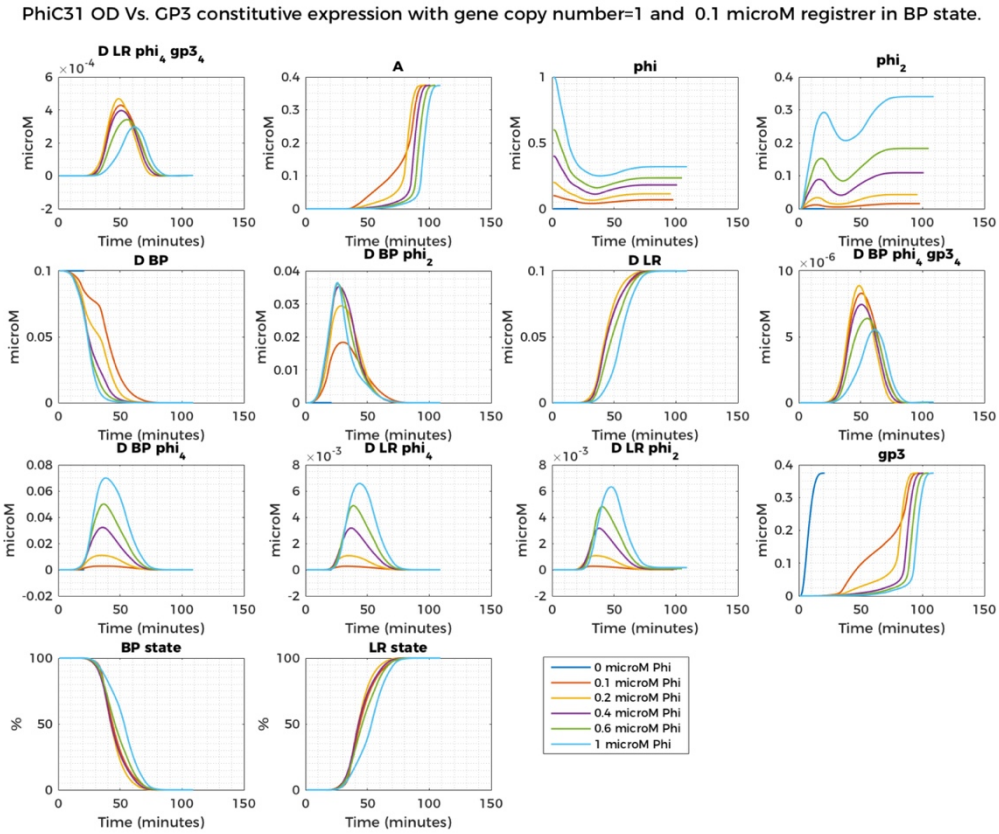
For the integrase-excisionase action:

1. Excisionase doesn't dimerize in the cellular medium.
2. Excisionase doesn't form complexes with PhiC31 in the cellular medium.
3. We consider cooperativity in the excisionase binding to the reporter.
4. The recombination performance is irreversible.

Sensivity analysis:

To continue we are going to show the behavior of our model when we change the proportion between

4.1.3. IN SILICO EXPERIMENTS



Temporal evolution of chemical species.

As integrase-recombinase increases, the LR state increases, while BP state decreases, because of the recombinase performance on the reporter. As a result, starts the expression of the protein A in the reporter since LR is the on state. As can be observed on the evolution of A, for a greater concentration of recombinase, the production has a delay because of the Gp3 that is pretending to reverse de LR state to de BP state. In this case, finally 100% of the register is in LR state.

4.1.4. CONCLUSIONS

- It is crucial to select the concentration of PhiC31 and GP3 and its proportion to obtain a good behavior of our genetic device.
- We have studied the importance of unproductive complex formed by PhiC31 ans GP3.
- In the future, we consider that it is important to study the expression of our target protein with a dexamethasone stimulus.

7. CONCLUSIONS

We have obtained a general vision of our genetic system, understanding it at all and creating strategies to work with it.

We also have optimized our parameter value in the optogenetic circuit, having a model that can predict the protein production depending on the light stimuli.

Taking into account that the experimental data are from plants, it usually is a point in favour because of its difficult extraction and consequently, a difficult optimization and characterization.

With the recombinase model, we have obtained a model of a system that nowadays is un study and we have been able to understand how PhiC31 and GP3 works. Also, we think that it would be useful to future studies in this field.