Elise Ledieu

Mosai*coli*

Modeling report

ETH Zurich iGEM 2015 project

Abstract

Emergence of complex patterns in nature is a fascinating and widely spread phenomenon, which is not fully understood yet. Mosaicoli aims to investigate emergence of complex patterns from a simple rule by engineering a cellular automaton into *E. coli* bacteria. This automaton comprises a grid of colonies on a 3D-printed millifluidic chip. Each colony is either in an ON or OFF state and updates its state by integrating signals from its neighbors according to a genetically pre-programmed logic rule. Complex patterns such as Sierpinski triangles are visualized by fluorescence after several steps of row-wise propagation. Sequential logic computation based on quorum sensing is challenged by leakiness and crosstalk present in biological systems. Mosaicoli overcomes these issues by exploiting multichannel orthogonal communication, riboregulators and integrase-based XOR logic gates. Engineering such a reliable system not only enables a better understanding of emergent patterns, but also provides novel building blocks for biological computers.

Acknowledgments

First of all, I would like to thank my team members: Maximilian Bahls, Daniel Gerngross, Nadine Lobsiger, Claude Lormeau, Stefanie Meier and Sumana Srivatsa who made this experience unique. I would also like to thank our advisors Bartolomeo Angelici, Sabine Oesterle and Lukas Widmer for their support throughout the project. Last but not least, I would like to thank the professors that made this experience possible: Prof. Kobi Benenson, Prof. Sven Panke and Prof. Joerg Stelling.

Outline

1 Introduction

Under the international Genetically Engineered Machine (iGEM) competition framework, ETH Zurich presented its project Mosai*coli*. This project focused on the emergence of complex patterns in bacterial population. It aims at engineering a predictable cellular automaton with colonies of the model organism *E. coli*. From a theoretical point of view, forecasting bacterial behavior is equivalent to building a reliable model. Our modeling strategy was to use mass action kinetics and to retrieve most of parameters from our experimental data. Biologically, different technologies were used: for communication, quorum sensing and for computation at the genetic level, integrases. If quorum sensing is a well-understood phenomena, there is not much available knowledge on integrases. To handle those problems separately, we used a systemic approach to model the bacterial behavior. We decomposed the information flow into submodules, that will be presented on the second section of this report. Then, I will focus on the example of integrase switch modeling. The limits of the quorum sensing module, such as leakiness or cross-talk, were explored and quantified thanks to an exhaustive experimental dataset and its corresponding model.

2 A systemic approach to Mosai*coli*

a Compartementalization

Our set up is based on an information flow. Several functional entities can be identified and compartmentalized into sub-modules. The sensing part corresponds to two blocks of the modeling description. Each block is a sensing system, responding to one particular input (the noise due to the false sensing of the other input is also taken into account). The computing part decomposes into two sub-modules in our model: first, each integrase have to undergo dynamical changes (dimerization, DNA binding) independently from each other; then, the computing operation occurs by coupling both chemical signals. The production module is the last sub module of our single cell model. Another modeling part corresponds to the diffusion of bacteria through the device.

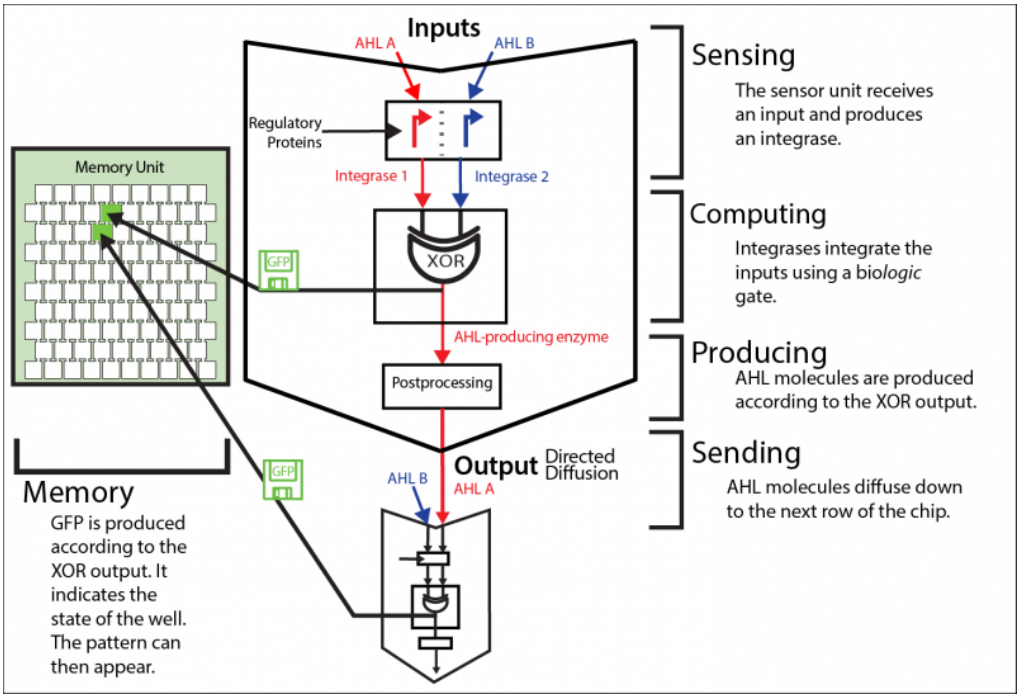

Figure 1 – Information processing in Mosai*coli*

b Computing information in biology: the XOR gate module

We consider a binary exclusive or (XOR) logic gate, with two inputs and one output and we want this function to be computed via biological signals.

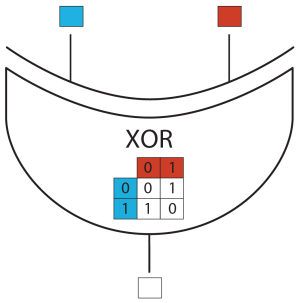

Figure 2 – Truth table of the XOR logic gate

Integrases, such as Bxb1 or ϕC31, are molecules that can flip a DNA sequence. They bind to particular sites on DNA and work as a pair.

The fragment integrases can flip is a terminator. Thus, the terminator can either be on or off.

- T $_{on}$: terminator is on, transcription is blocked.
- T$_{off}$: terminator is off, transcription is active. It corresponds to one flipping of the terminator.

In our design, we are interested in a double flipping. That is to say that two pairs of binding sites surrounds the fragment to be flipped. One pair of binding sites can be bound by DBxb1 and the other one by ϕC31.

The terminator can be flipped once if either DBxb1 or ϕC31 is present. The state T_{off} can be reached via two possible transitions. We further decompose it into two different states: $T_{offBxb1}$(flipping due to presence of Bxb1) and T_{off31}(flipping due to presence of ϕC31).

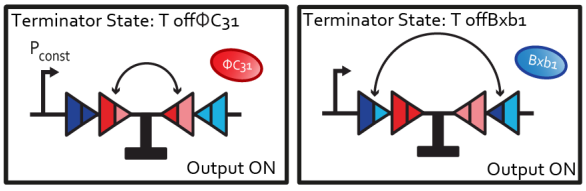

Figure 3 – Decomposition of the off output into two terminator states

Flipping by integrases is irreversible. The initial state, in which the terminator is on, is different from the state after two switches. From this last state, no further evolution of the system is possible. Therefore, we decompose the T_{on} into two different states: $T_{on,i}$(initial state of the system, no flipping) and $T_{on,f}$(final state of the system after two flips).

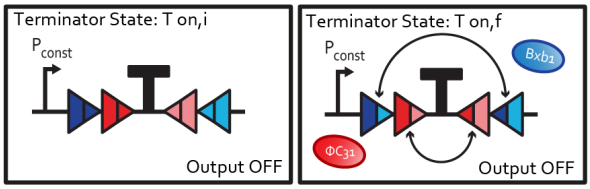

Figure 4 – Decomposition of the on output into two terminator states

Thus, we obtain the truth table of the biologically computed logic gate.

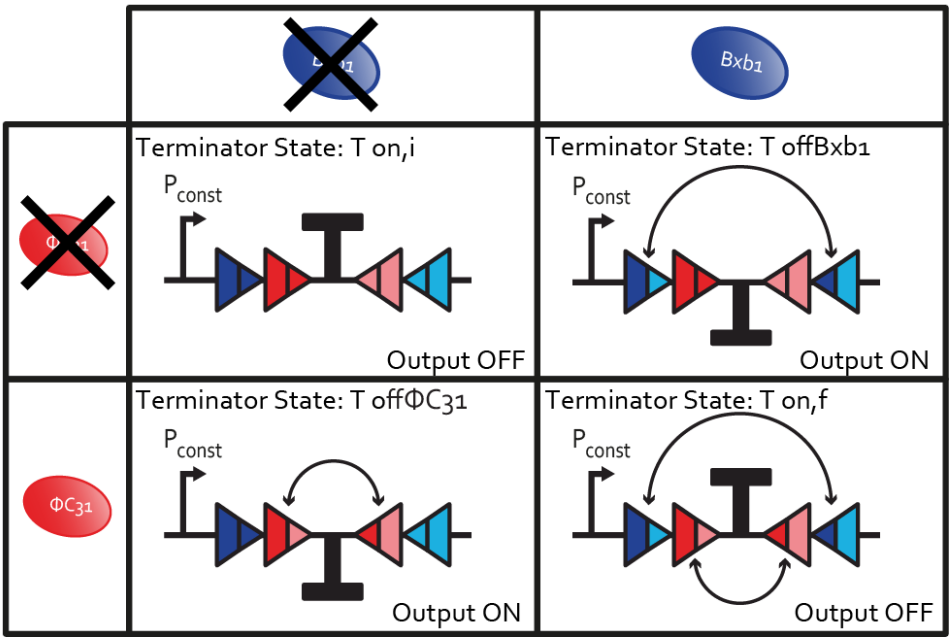

Figure 5 – Truth table of the XOR biologic gate: Summary of the model, coupled with the biological explanation

3 The Bxb1 sub-module: Integrase processing

In our design, integrases compute the output of the logic gates. Integrases allow flipping one fragment of DNA. Therefore, they are of central importance to our design. However, their characterization in literature is incomplete. In particular, quantitative insight into dimerization rates and DNA-binding rates is lacking. Such data is necessary to set up a mathematical model to describe the overall Mosaicoli process. We decided to estimate the missing parameters from published from published experiments based on a model that we developed ourselves.

a Definitions

There are two chemical species involved:

- Bxb1: Serine integrase that can fold into two conformations - Bxb1a and Bxb1b. We chose to use a common connotation for both conformations - Bxb1.
- DBxb1: Dimerized form of Bxb1. We chose to use a common connotation for both homodimers, DBxb1a and DBxb1b.

Each dimer of integrases can specifically bind to a DNA binding site. As the flipping is irreversible, these DNA binding sites can be in three possible states:

- SI_{Bxb1}: inactive DNA binding site. No dimer is bound to this site, which has never been flipped.
- SA_{Bxb1}: active DNA binding site. A dimer is bound to this site.
- SF_{Bxb1}: flipped DNA binding site. This site has been irreversibly flipped.

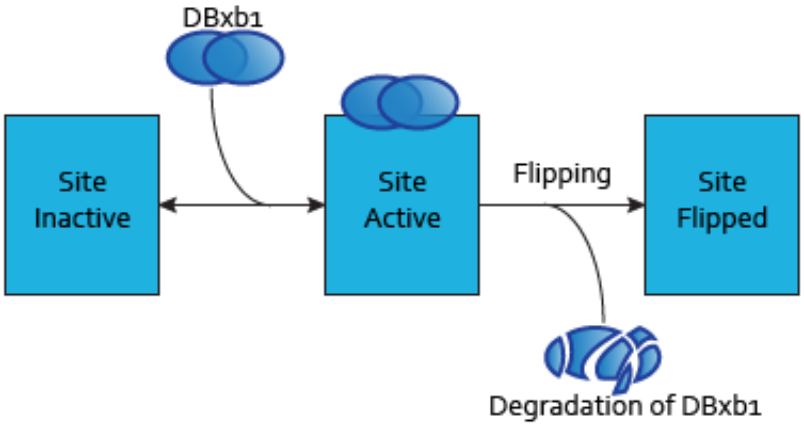

Figure 6 – The three different states of DBxb1-DNA binding sites

b Reactions

$$Bxb1 + Bxb1 \leftrightarrow DBxb1$$
$$DBxb1 + SI_{Bxb1} \leftrightarrow SA_{Bxb1}$$
$$Bxb1 \rightarrow$$
$$DBxb1 \rightarrow$$

Applying mass action kinetic laws, we obtain the following set of differential equations:

$$\frac{d[Bxb1]}{dt} = -2k_{DBxb1}[Bxb1]^2 + 2k_{-DBxb1}[DBxb1] - d_{Bxb1}[Bxb1]$$
$$\frac{d[DBxb1]}{dt} = -k_{SABxb1}[DBxb1][SI_{Bxb1}] + k_{-SABxb1}[SA_{Bxb1}] + k_{DBxb1}[Bxb1]^2 - k_{-DBxb1}[DBxb1] - d_{DBxb1}[DBxb1]$$
$$\frac{d[SA_{Bxb1}]}{dt} = k_{SABxb1}[DBxb1][SI_{Bxb1}] - k_{-SABxb1}[SA_{Bxb1}]$$

where

- k_{DBxb1}: Dimerization rate of Bxb1
- k_{-DBxb1}: Dissociation rate of DBxb1
- k_{SABxb1}: Rate of formation of SA_{Bxb1} fom DBxb1 and SI_{Bxb1}
- $k_{-SABxb1}$: Dissociation rate of SA_{Bxb1}
- d_{Bxb1}: Degradation rate of Bxb1
- d_{DBxb1}: Degradation rate of DBxb1

Even if degradation rates were not determined specifically for the serine integrases and their dimerized form, degradation rates of proteins in *E. coli* are available. We assume that the degradation rates of dimerized forms are two times higher than the degradation rates of monomers: $d_{DBxb1} = 2*d_{Bxb1}$. To characterize integrases behavior, we focused on estimating the parameters for dimerization and DNA-binding.

c Characterization of the integrase DNA-binding reaction

The parameter fitting is based on data from Bonnet's paper [Bonnet et al., 2013]. Their experimental setup is different from the one used in Mosai*coli*. However, they experimentally retrieve a transfer function between their input aTc and Bxb1 switching rate. To use this experimental data, we have to model their induction mechanism using aTc. We use a simplified version of this phenomenon proposed by the 2013 UCSF iGEM team (Source : `http://2013.igem.org/Team:UCSF/Modeling`). They modeled the induction with aTc as a leaky Hill function.

$$\frac{d[Bxb1]}{dt} = k_{mRNA_{Bxb1}} * (A_L + B_L * \frac{[aTc]^n}{[aTc]^n + K_L^n}) - 2k_{DBxb1} * [Bxb1]^2 + 2k_{-DBxb1} * [DBxb1] - d_{Bxb1} * [Bxb1]$$
$$\frac{d[DBxb1]}{dt} = k_{DBxb1} * [Bxb1]^2 - k_{-DBxb1} * [DBxb1] - k_{SABxb1} * [DBxb1] * [SI_{Bxb1}] + k_{-SABxb1} * [SA_{Bxb1}] - d_{DBxb1} * [DBxb1]$$
$$\frac{d[SA_{Bxb1}]}{dt} = k_{SABxb1} * [DBxb1] * [SI_{Bxb1}] - k_{-SABxb1} * [SA_{Bxb1}]$$

where

- A_L: Basal expression level of the tet promoter

- B_L: Maximal expression level of the tet promoter
- n: Hill exponent
- K_L: Half-maximal effective concentration of aTc
- $k_{mRNA_{Bxb1}}$: Translation rate of Bxb1 (assumed)

The following set of assumptions is made:

- The back-reaction from DBxb1 binding to the inactive site state is considered to be negligible compared to the flipping rate. That is to say that once a site is active, it can only be flipped. Thus, an active site would only be a transitional state in our whole cell model. As the flipping is not modeled in our integrase subsystem, active sites are not transformed into flipped sites at the end of the information pipeline. Thus, we consider that we can express the switching rate given active site concentration.
- As switching needs two active sites to be effective, the switching rate is approximated to: $\left(\frac{SA_{Bxb1}}{S_{TOT}}\right)^2$. This approximation is understated by probabilistic considerations.
- Given the normalization of the paper, basal rate of production of Bxb1 is not taken into account. Thus, we consider that $A_L = 0$
- The activation by aTc is assumed to be dominant over degradation and dimerization of Bxb1. It is supposed to be valid on the range of aTc concentration considered. $\frac{k_{mRNA_{Bxb1}} * \left(A_L + B_L * \frac{[aTc]^n}{[aTc]^n + K_L^n}\right)}{d_{Bxb1} K'_{DBxb1}} >> 1$ with $K'_{DBxb1} = \frac{k_{-DBxb1} + d_{DBxb1}}{k_{DBxb1}}$

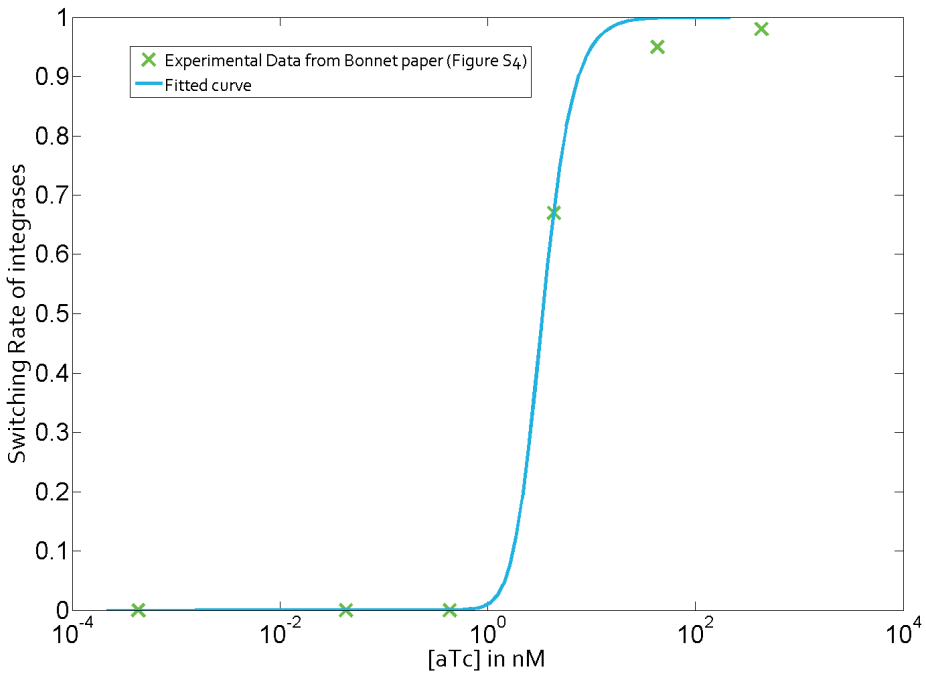

Figure 7 – Parameter fitting of the dissociation rate constant of K_{SABxb1}

We consider the system at steady-state. After derivation, the following explicit equation can be retrieved:

$$[SR]_{qss} = \left(\frac{B_L * [aTc]^n}{\lambda_1 K_L^n + (B_L + \lambda_1)[aTc]^n} \right)^2$$

where

$$\lambda_1 = \frac{2 * d_{Bxb1} * K_{SABxb1}}{k_{mRNA_{Bxb1}}}; K_{SABxb1} = \frac{k_{-SABxb1}}{k_{SABxb1}}$$

Using the Least Absolute Residual method, we determined the lumped parameter $_1$. Here is the value with its 95% confidence bounds:

$$\lambda_1 = 1.82e-07(1.649e-07, 1.992e-07)$$

We assume that:

- d_{Bxb1} corresponds to the order of magnitude of 10^{-2} min^{-1}, as most of the protein in *E. coli*. [Maurizi, 1992]
- $k_{mRNABxb1}$ is of the order of magnitude 10^{-1} min^{-1} $mRNA^{-1}$. We estimated to be a low value because the starting codon of Bxb1 is GTG (and not ATG) and this parameter also takes into account folding time.

Thus, K_{SABxb1}'s order of magnitude is 10^{-6} nM. The interpretation of this dissociation constant, K_{SABxb1}, is that the DNA binding reaction is really specific, as it can be expected for integrases. By assuming that k_{SABxb1}, the rate of formation of SA_{Bxb1}, is not rate limiting and fixing it to 1, we find that $k_{-SABxb1}$'s order of magnitude is 10^{-6} nM.

d Characterization of the dimerization reaction

The fitted data are once again extracted from Bonnet's paper [Bonnet et al., 2013]. As in the previous section, the system is simplified on the sensing level (Source : `http://2013.igem.org/Team:UCSF/Modeling`).

The following set of assumptions is, then, made:

- The back-reaction from DBxb1 binding to the inactive site state is considered to be negligible compared to the flipping rate. That is to say that once a site is active, it can only be flipped. Thus, an active site would only be a transitional state in our whole cell model. As the flipping is not modeled in our integrase subsystem, active sites are not transformed into flipped sites at the end of the information pipeline. Thus, we consider that we can express the switching rate given active site concentration.
- As switching needs two active sites to be effective, the switching rate is approximated to: $\left(\frac{SA_{Bxb1}}{S_{TOT}}\right)^2$. This approximation is understated by probabilistic considerations.
- Given the normalization of the paper, basal rate of production of Bxb1 is not taken into account. Thus, we consider that $A_L = 0$
- We neglect the degradation rate of DBxb1, d_{DBxb1}, with respect to the dimerization reaction.

We consider the system at steady-state. After derivation, the following explicit equation can be retrieved:

$$[SR]_{qss} = \left(\frac{(B_L * [aTc]^n)^2}{((B_L * [aTc]^n)^2 + \lambda_2 * \lambda_1 (K_L^n + [aTc]^n)^2)}\right)^2$$

where

$$\lambda_1 = \frac{2 * d_{Bxb1} * K_{SABxb1}}{k_{mRNA_{Bxb1}}}; \lambda_2 = \frac{d_{Bxb1} * K_{SABxb1}}{2 * k_{mRNA_{Bxb1}}} K_{DBxb1} = \frac{k_{-DBxb1}}{k_{DBxb1}}; K_{SABxb1} = \frac{k_{-SABxb1}}{k_{SABxb1}}$$

As the value of $_1$ was derived in the previous characterization step, we use the Least Absolute Residual method to determine the lumped parameter $_2$. Here is the value with its 95% confidence bounds:

$$\lambda_2 = 8.211e-07(7.421e-07, 9.001e-07))$$

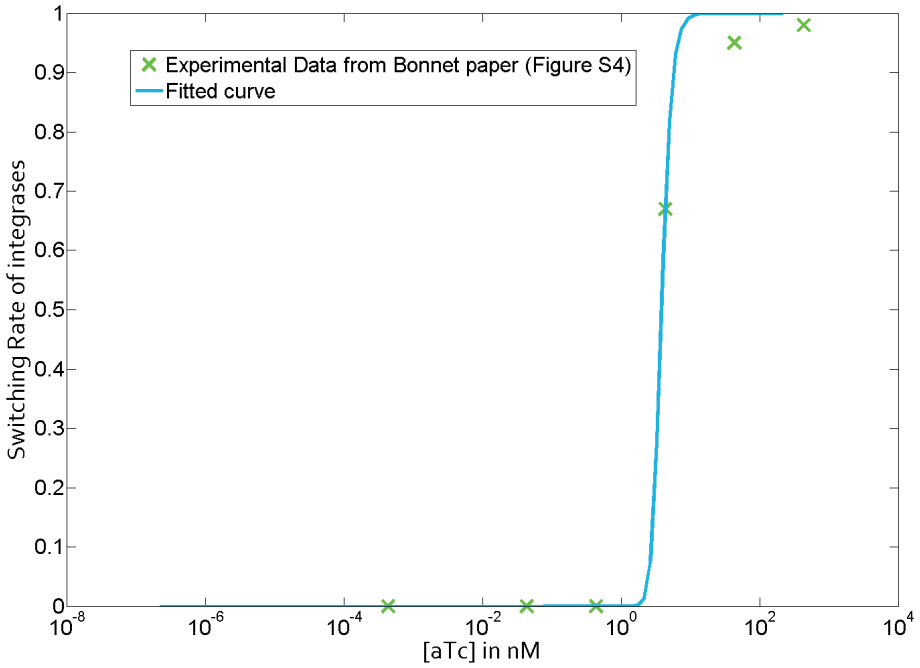

Figure 8 – Parameter fitting of the dimerization rate constant of K_{DBxb1}

We assume that:

- d_{Bxb1} corresponds to the order of magnitude of $10^{-2} min^{-1}$, as most of the protein in *E. coli* [Maurizi, 1992].
- $k_{mRNA_{Bxb1}}$ is of the order of magnitude of $10^{-1} min^{-1}$. We estimated it to be a low value because the starting codon of Bxb1 is GTG (and not ATG) and this parameter also takes into account the folding time.

Thus, K_{DBxb1}'s order of magnitude is $10^{-6} nM$. The interpretation of this dissociation rate constant is that the dimerization reaction is really specific, as it can be expected for integrases.

By assuming that k_{DBxb1}, the rate of formation of D_{Bxb1}, is not rate limiting and fixing it to 1, we find that k_{-DBxb1}'s order of magnitude is 10^{-6} nM.

e Results

The figures show the predicted transfer function for the ratios of active sites of Bxb1 (SA_{Bxb1}) to total number of sites ($SBxb1_{TOT}$) and active sites of ϕC31 (SA_{31}) to total number of sites ($S\phi C31_{TOT}$) as a function of their respective integrase concentrations based on the fitted parameters. With the fitted parameters, we found the Km value to be 10 nM.

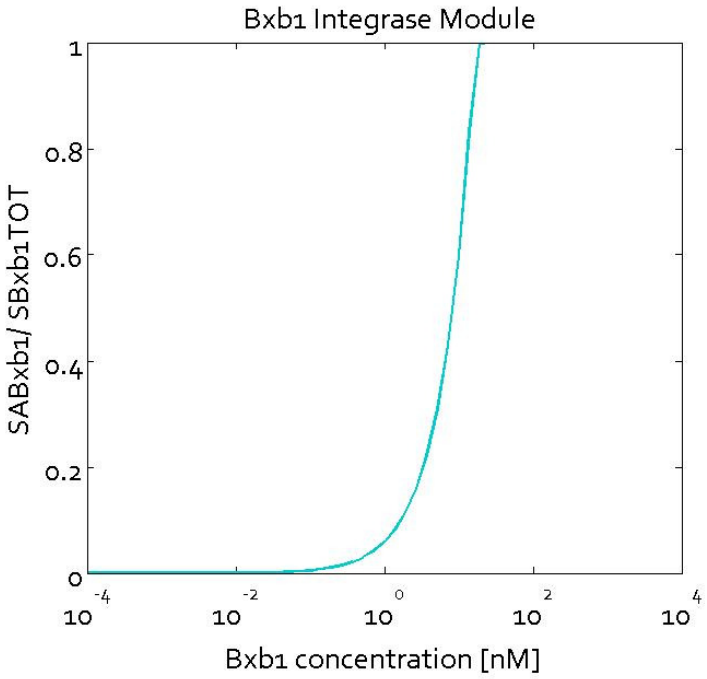

Figure 9 – Predicted transfer function for Bxb1 module.

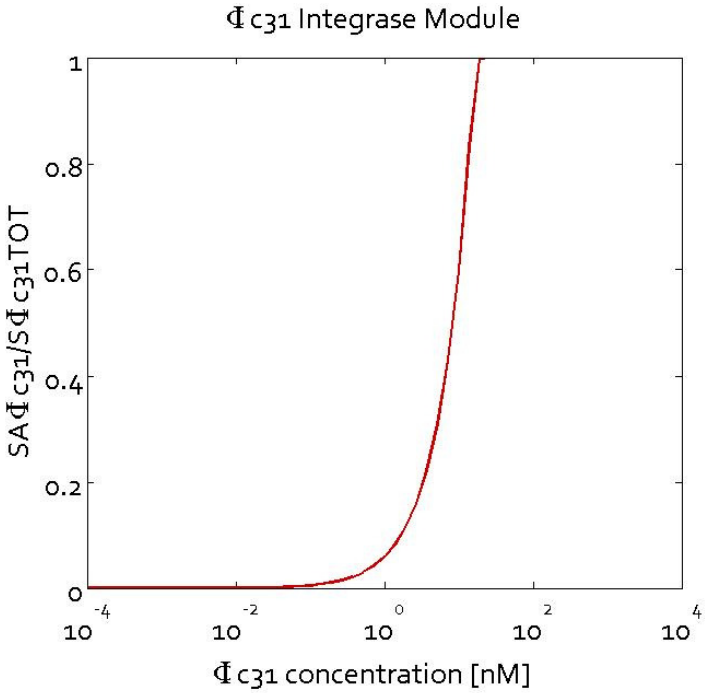

Figure 10 – Predicted transfer function for ϕC31 module.

4 The limits of the quorum sensing module

For our Mosiacoli project, we were looking for molecular systems that allow orthogonal cell-to-cell communication in order to implement connected XOR logic gates. We decided to exploit the quorum sensing systems LuxI/LuxR, LasI/LasR, and RhlI/RhlR in order to achieve the required orthogonal cell-to-cell communication. We developed a model for these cellular information processing. Even though the corresponding inducer molecules are commercially available and the systems often used, in particular in iGEM projects, potential crosstalk activity between the different systems may be a severe problem.

a Leakiness

The leakiness of the quorum sensing promoters is a major issue in our system. As the signal propagates row-wise, error diffusion could lead to a totally different pattern. The goal is then to control the leakiness. This issue was particularly observed and addressed in the case of the Lux promoter during our experiments set. This leakiness is dependent on LuxR concentration in the cell.

Leakiness was modeled as an offset in the classical Hill function.

$$rFluo = a + b\frac{[AHL]^n}{K_m^n + [AHL]^n}$$

where

- rFluo is the relative fluorescence (absolute measured fluorescence value over OD),
- a the basal expression rate (Leakiness),
- b the maximum fold expression rate,
- n the Hill coefficient,
- K_m the activation concentration of AHL.

Given this offset and the maximal expression, the signal over noise ratio can be derived. This ratio, which can then be compared amongst all curves, characterizes the impact of leakiness on the behavior of a system. That allows us to quantify the impact of a riboregulating construct associated to a promoter. The leakier a construct in its native form is, the more impact the riboregulator will have, and the more likely it is for the riboregulator to increase the signal over noise ratio. Our final constructs (Promoters with a riboregulating system) have the following parameters:

Tab 1 – Leakiness depending on promoters

Promoter used	**Signal over noise ratio**
P_{Lux} without riboregulating system	23
P_{Lux} with riboregulating system	79
P_{Las} without riboregulating system	84
P_{Las} with riboregulating system	55

b Cross-talk

We investigated the existence of cross-talk between three quorum sensing systems (Lux, Las and Rhl). Each quorum sensing system is based on three components: a signaling molecule, a regulatory protein and a promoter. Cross-talk implies non-orthogonality of the communication systems. It corresponds to the fact that LasAHL can activate the Lux promoter, even if it is not its native communicating pathway. There are 27 combinations possible and only 3 native combinations between signaling molecule, regulatory protein and promoter.

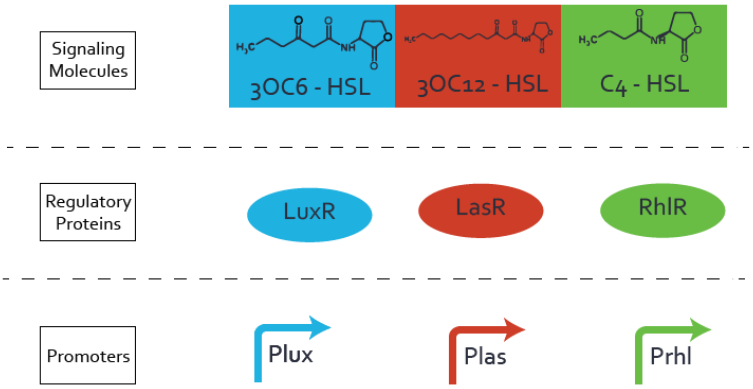

Figure 11 – 27 possible combinations of the quorum sensing module

From the exhaustive experimental data, the central role of regulatory proteins was identified, allowing the characterization of two-levels of cross-talk. Each experimental data set was fitted to an Hill function using the Least Absolute Residual method.

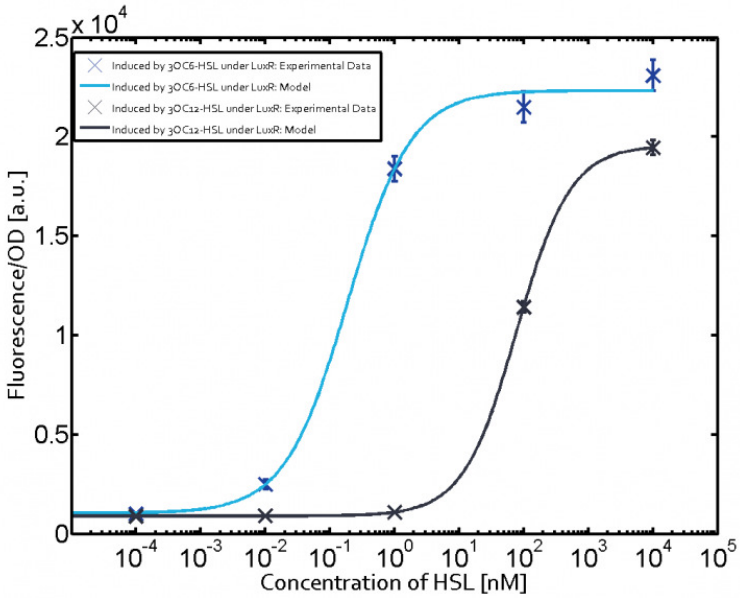

Figure 12 – Here is the promoter of interest is Plux. Two different combinations are presented: the native Lux system (Plux, LuxR and LuxAHL) and the Lux promoter-regulator system activated by LasAHL. The second combination indeed turns on the transcription, even if it is not the native system. Depending on the concentration of the signaling molecules, it can lead to confusion of inputs.

5 Conclusion

As a modeling team, we predicted the whole system's behavior with a comprehensive fully derived model. We have a thorough model for the whole Mosaicoli system based on a detailed study of every module of this system, including quorum sensing, integrases, and diffusion. We derived the formulae from mass action kinetics and state precisely which approximations we do and why we think we are justified to do so. All our steady state simulations and some of our dynamic simulation results fit to experimental data from our own experiments or found in literature. My contribution on the modeling work was mostly on the integrases' modeling and on the fitting of parameters coming from our experiments. I was also pro-active on the Human Practice part of the project, for which I took part into Science Slams and interviewed some complexity specialists.

Bibliography

[Bonnet et al., 2013] Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013). Amplifying genetic logic gates. *Science*, 340(6132):599–603.

[Maurizi, 1992] Maurizi, M. (1992). Proteases and protein degradation inescherichia coli. *Experientia*, 48(2):181.