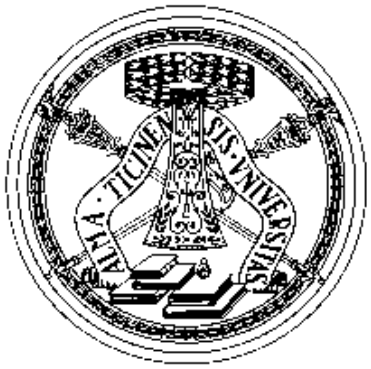
UNIVERSITA' DEGLI STUDI DI PAVIA

Dottorato di Ricerca in Bioingegneria e Bioinformatica

Dipartimento di Ingegneria Industriale e dell'Informazione



A synthetic close-loop regulator in *Escherichia coli* to control the concentration of a signalling molecule: bottom-up design, mathematical modelling and *in vivo* characterization

Advisor:
Prof. Paolo Magni
PhD Program Chair:
Prof. Riccardo Bellazzi

PhD Thesis by
Nicolò Politi

XXVI Ciclo (2010-2013)

No problem is too small
or too trivial
if we can really do something
about it.

R. P. Feynman

Abstract (English)

Many efforts have been recently done in order to introduce the concepts of standardization, abstraction and modularity, typical of engineering, into life sciences. These efforts, together with the technological improvements in the field of molecular biology, gave birth to Synthetic Biology, a novel discipline that aims to create new functionalities in living systems, such as microorganisms, or to rearrange already existing ones. This is carried out by introducing customized DNA sequences into living cells, in order to confer them the ability to address new tasks.
Dealing with living systems, capable of replicating and reacting to environmental stimuli, makes the redefinition and validation of the previously mentioned concepts necessary, especially standardization and modularity. Physical standardization has been achieved by the introduction of standardized assembly techniques, such as the BioBrick™ standard assembly, a procedure to assemble DNA sequences via an easy and iterative procedure, which has become very popular in the last decade. Beside physical standardization, also shared methodologies for the measurement of the activity of biological components are needed; this concept requires further investigations, although some procedures are nowadays shared and commonly used. Moreover, modularity, which allows the predictability of the behaviour of composite systems on the basis of the characterization of their subparts, still needs to be verified, by investigating its validity boundaries.
Keeping these issues in mind, this thesis focuses on the design and assembly of a system with a complex behaviour using a bottom-up approach. Since negative feedback control is widely used in the engineering world and it could bring to complex adaptive circuits in Synthetic Biology, this work focuses on the bottom-up design of a synthetic close-loop controller in *Escherichia coli*. In order to follow a rigorous bottom-up approach, an ordinary differential equation system has been derived for the full circuit and its parameters have been identified from data acquired through *ad hoc* experiments carried out on

subsystems, i.e., smaller functional units of the whole network. Finally, the goodness of the model is based on the comparison between predictions and experimental data obtained from the full system. The thesis describes the whole design process of this circuit, from parts selection to the final circuit testing. In particular, the herein presented genetic circuit aims at implementing the control of the concentration of a signalling molecule in liquid *E. coli* cultures through a negative feedback mechanism. This configuration, called the "close-loop", is expected to maintain the concentration of the target molecule at a given set-point, in contrast to the "open-loop" configuration (i.e., without the negative feedback), where no control is implemented and the molecule concentration is allowed to increase. The DNA sequences needed to implement such a genetic controller were selected and retrieved from the catalogue called Registry of Standard Biological Parts, an open access repository of biological parts in the BioBrick™ format; then, design specifications were defined. The system is based on the *quorum sensing* mechanism, that is exploited by some naturally occurring bacteria to regulate some genetic pathways in a density dependent fashion: the *luxI* gene is responsible for the expression of the LuxI enzyme which produces the signalling molecule 3-oxohexanoyl homoserine lactone ($3OC_6$-HSL); the *aiiA* gene encodes the AiiA enzyme, which is able to degrade $3OC_6$-HSL. When *luxI* is constitutively expressed and *aiiA* is placed under the control of a $3OC_6$-HSL inducible promoter, P_{Lux}, it is possible to obtain a negative feedback in which $3OC_6$-HSL starts to be enzymatically degraded when it reaches a sufficiently high concentration. On the basis of the chosen DNA sequences and the design specifications, an ordinary differential equation system of the genetic controller was derived in order to predict its output (i.e., the concentration of the signalling molecule over time). The parameters were identified on experimental data obtained from *ad hoc* performed tests on the subparts and the model of the complete circuit was used to obtain predictions of the behaviour of the genetic controller. Given the uncertainty affecting model parameters, a Monte Carlo-based strategy has been implemented in order to propagate the uncertainty of the estimates from the subparts models to the output of the complete close-loop model.
A comparison between the open- and the close-loop configuration was performed both *in silico* and *in vivo* to highlight the main differences between the two conditions and the feasibility of a bottom-up approach was verified comparing the predictions with experimental measurements performed on engineered bacteria in a bioreactor with a continuous culture mode.
In Chapter 1, the basic concepts of Synthetic Biology will be illustrated and deeply discussed, introducing the concept of biologic "parts" and the classification based on their function. Then, a review of the state of the art and of the most explanatory works in the field of Synthetic Biology will be shown, with a

particular focus on the mathematical modelling, which is fundamental in the design of complex systems, and on feedback-controlled synthetic circuits.
In Chapter 2, the methods will be illustrated. First, the laboratory protocols used will be reported. An overview on the design of the negative feedback controller will then be given, introducing the mathematical models of the whole system and of its basic subparts. The strategy, where the parameters of the whole system are estimated from *ad hoc* experiments on subparts, will be illustrated and a brief description on data analysis will be given. The Monte Carlo method applied to propagate uncertainty will be deeply described, as well as the strategy used to implement the continuous culture mode used to carry out the most important experiments.
In Chapter 3, results will be presented. Simulations of the complete circuit model with reasonable parameter values will be shown to verify that the desired close-loop behaviour can be achieved, then parameters identification will be described. The Monte Carlo approach results will be illustrated and finally model predictions of the open- and the close-loop configurations will be compared to the experimental data.
In Chapter 4, conclusions will be given, proposing further investigations on the close-loop circuit behaviour and discussing the possible future improvements.
In Appendix A, a study on the stability of three commonly used inducers (i.e., chemicals able to regulate protein expression) will be presented; the stability of these molecules was studied in different experimental conditions and the variables that mostly impact on their half-life were identified through a factorial analysis, using whole-cell biosensors for measurements. The quantification of their decay rate is important to better predict the behaviour of the circuits under investigation in this thesis.

Abstract (Italian)

Recentemente sono stati fatti diversi sforzi al fine di introdurre i concetti di standardizzazione, astrazione e modularità, tipici dell'ingegneria, nell'universo delle scienze della vita. Questi sforzi, insieme ai miglioramenti tecnologici nel campo della biologia molecolare, hanno dato vita alla Biologia Sintetica, una disciplina i cui scopi sono da un lato creare nuove funzionalità, dall'altro riorganizzare quelle già esistenti in sistemi viventi, come ad esempio i microrganismi. Questo avviene introducendo sequenze sintetiche di DNA all'interno delle cellule, al fine di conferire loro la capacità di svolgere nuovi compiti.
Interfacciarsi con sistemi viventi, in grado di replicarsi e reagire a stimoli ambientali, rende necessaria la ridefinizione e la validazione dei concetti precedentemente citati, in particolare quelli di standardizzazione e modularità. La standardizzazione fisica è stata ottenuta con l'introduzione dello standard fisico BioBrick™, una procedura che consente l'assemblaggio di sequenze di DNA mediante un processo semplice e ripetibile, divenuto molto popolare negli ultimi dieci anni. Oltre alla standardizzazione fisica, sono necessarie anche metodologie condivise per la misurazione dell'attività di componenti biologici; questo concetto richiede ulteriori indagini, sebbene alcune procedure siano oggigiorno condivise e comunemente utilizzate. Inoltre, la modularità, che permette di ottenere sistemi compositi con comportamento predicibile, sulla base della caratterizzazione delle loro sottoparti, deve ancora essere verificata, indagando i limiti entro i quali essa può essere ritenuta valida.
Tenendo presente questi problemi, questo lavoro di tesi si concentra sulla progettazione e l'assemblaggio di un sistema con un comportamento complesso mediante un approccio *bottom-up*. Poichè il controllo con feedback negativo è ampiamente sfruttato nel mondo dell'ingegneria e può portare ad ottenere circuiti complessi con comportamento adattativo in Biologia Sintetica, questo lavoro si concentra sulla progettazione *bottom-up* di un controllore sintetico in anello chiuso in *Escherichia coli*. Per seguire un approccio *bottom-up* rigoroso, è stato definito un modello matematico basato su equazioni differenziali

ordinarie per il circuito genetico completo e i suoi parametri sono stati identificati sulla base di esperimenti *ad hoc* svolti sulle sottoparti, cioè piccole unità funzionali dell'intera rete di regolazione. Infine, la bontà del modello è stata valutata sulla base del confronto tra le predizioni e i dati sperimentali ottenuti dai test sull'intero sistema. Questo lavoro di tesi descrive l'intero prcesso di progettazione e realizzazione di questo circuito, dall'individuazione delle parti alla sperimentazione del circuito finale. In particolare, il circuito genetico qui presentato mira a realizzare il controllo della concentrazione di una molecola di segnalazione in colture liquide di *E. coli* attraverso un meccanismo di feedback negativo. Questa configurazione, chiamata "anello chiuso", mira a mantenere la concentrazione della molecola target ad un dato set-point, al contrario della configurazione in "anello aperto" (ovvero senza feedback negativo) dove la molecola target viene accumulata in maniera non controllata. Le sequenze di DNA necessarie per implementare tale controllore genetico sono state scelte e recuperate dal catalogo denominato Registry of Standard Biological Parts, un archivio *open source* dei componenti biologici in formato BioBrick™; successivamente sono state definite le specifiche di progettazione. Il sistema si basa sul meccanismo del *quorum sensing*, che viene sfruttato da alcuni batteri per regolare alcuni pathway genetici in funzione della densità cellulare: il gene *luxI* è responsabile dell'espressione dell'enzima LuxI che produce la molecola segnale 3-ossoesanoil omoserin-lattone ($3OC_6$-HSL); il gene *aiiA* codifica l'enzima AiiA, il quale è capace di degradare il $3OC_6$-HSL. Quando *luxI* è espresso in maniera costitutiva e *aiiA* è posto sotto il controllo del promotore inducibile da $3OC_6$-HSL, cioè P_{Lux}, è possibile ottenere un feedback negativo in cui il $3OC_6$-HSL inizia ad essere degradato enzimaticamente quando raggiunge una concentrazione sufficientemente alta. Sulla base delle sequenze di DNA scelte e le specifiche di progetto, è stato costruito un modello matematico del sistema per prevederne l'output (cioè la concentrazione della molecola segnale nel tempo). I parametri sono stati identificati a partire dai dati sperimentali ottenuti da test *ad hoc* effettuati sulle sottoparti e il modello del circuito completo è stato usato per ottenere previsioni del comportamento del controllore genetico. È stato implementato un approccio basato sul metodo Monte Carlo per propagare l'incertezza delle stime dei parametri delle sottoparti all'output del modello del circuito completo. Un confronto tra la configurazione in anello aperto e quella in anello chiuso è stato eseguito sia *in silico*, che *in vivo*, per evidenziare le principali differenze tra le due condizioni, mentre la fattibilità di un approccio *bottom-up* è stata verificata confrontando le previsioni ottenute con le misure sperimentali provenienti da colture continue di batteri ingegnerizzati.

Nel Capitolo 1 saranno illustrati i concetti di base della Biologia Sintetica, introducendo il concetto di "parte" biologica e la classificazione funzionale delle

parti. Sarà quindi presentata una rassegna dello stato dell'arte e una raccolta degli studi più significativi nel campo della Biologia Sintetica, con una particolare attenzione alla modellizzazione matematica, che è fondamentale nel ciclo di progettazione ispirato dall'ingegneria.
Nel Capitolo 2 verranno illustrati i metodi. In primo luogo saranno presentati i protocolli di laboratorio utilizzati. Saranno poi descritte le fasi della progettazione del controllore con feedback negativo, introducendo i modelli matematici sviluppati per l'intero sistema e per le sue sottoparti. Sarà poi illustrata la strategia adottata per la stima dei parametri dell'intero sistema sulla base di esperimenti *ad hoc* condotti sulle sottoparti e verrà discussa la procedura di analisi dati. Il metodo Monte Carlo, applicato per propagare l'incertezza, verrà descritto in dettaglio, così come la strategia utilizzata per implementare la modalità di coltura continua, utilizzata per effettuare i test più importanti.
Nel Capitolo 3 saranno presentati i risultati. Verranno mostrate le simulazioni riguardanti il modello completo del circuito genetico ottenute con valori ragionevoli dei parametri, per verificare che il comportamento desiderato possa effettivamente essere ottenuto con il circuito progettato, quindi verrà descritta la procedura di identificazione parametrica. Sarà presentato il metodo Monte Carlo, utilizzato nella propagazione dell'incertezza, e infine saranno confrontate le previsioni del modello del circuito in anello aperto e in anello chiuso con i dati sperimentali.
Nel Capitolo 4 saranno fornite le conclusioni, proponendo ulteriori indagini sul comportamento del circuito in anello chiuso e discutendo i possibili miglioramenti futuri.
Nell'Appendice A verrà presentato uno studio preliminare sulla stabilità di tre induttori (composti chimici capaci di regolare l'espressione proteica) comunemente utilizzati; la stabilità di queste molecole è stata studiata in diverse condizioni sperimentali e sono state individuate le variabili che maggiormente impattano sulla loro emivita attraverso un'analisi fattoriale, utilizzando biosensori per effettuare le misure. La quantificazione delle lor emivite è importante al fine di migliorare la predizione del comportamento dei circuiti che sono stati considerati in questo lavoro di tesi.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The last century discoveries and advances about DNA structure and its manipulation [1, 2] represented the basis for the rise of biotechnology and genetic engineering. DNA is a molecule, constituted by nucleotides, which encloses the information needed in the development and functioning of cells and living organisms in general. The way in which the information is decoded and used is explained by the so called *central dogma* of molecular biology [3]: protein-coding DNA sequences (herein called genes) are transcribed into RNA molecules, which are translated into proteins at the ribosome level and finally DNA sequences can be replicated in living cells to propagate the encoded function to the progeny.
In the last decade a new discipline, based on the technological development in the field of molecular biology, has spread all over the world: Synthetic Biology. Several definitions have been given for this discipline; one of these states [4]:

"Synthetic biology aims to design and engineer biologically based parts, novel devices and systems as well as redesigning existing, natural biological systems."

This is an apparently simple but very exhaustive sentence as it covers all of the main features of Synthetic Biology. First of all, it underlines two distinct aspects: on the one hand the construction of new synthetic biological systems, based on unnatural molecules, that might (or might not) be inspired by nature [5]; on the other hand the rearrangement of biological existing entities in novel ways to obtain new functionalities in living systems [5, 6, 7]. In the former case the focus is to deeply understand the mechanisms at the basis of life, while in the latter the goal is the design and implementation of new functions to complete particular tasks (often inspired by electrical circuits or computers). Moreover, these concepts coming from the world of engineering, substantially

differentiate Synthetic Biology, which introduces a rational approach in designing living systems, from genetic engineering and biotechnology [4, 8, 9].
During its short life, Synthetic Biology has seen a first period during which simple functionalities were implemented, combining basic biological parts; nowadays synthetic biologists are facing new challenges as they are moving towards the design and development of more and more complex systems [10]. Yet, dealing with living systems, for example microorganisms, is not trivial as they are made up of biological elements (i.e., molecules) that can interact and interfere with each other affecting their respective behaviour. Cells are able to grow and evolve and these two features might affect the output of a system; particularly, the growth rate of a microbial population might affect the performances of a system, often resulting in its inability to meet the design specifications [11, 12]. Moreover, the capability to evolve could introduce issues on stability [8, 12] (i.e., the predicted behaviour of a biological system is not preserved on a long-time period due to unexpected mutation of one or more of its components). Working in a biological context also means interacting with organisms that are themselves capable of reacting to changes in environment, for example shifts in temperature or pH; therefore they have to be considered as active elements instead of passive *chassis*.
In order to face these problems, a rational approach in assembling and testing new systems is needed. This means that guidelines or at least a set of basic principles are needed in order to build new systems that act in a reliable and predictable way, giving a definition of the boundaries where this is feasible. Despite the differences in the context of application, this discipline could be considered analogous to other engineering fields, for example electrical or computer engineering [4, 13]. The design process of a synthetic biological function should follow the so called engineering cycle: first, specifications are defined, then system is designed according to these specifications (in this phase, parts standardization is recommended); subsequently modelling is required in order to predict the output. Finally, the implementation and the validation steps are performed to verify if specifications are satisfied [4], as shown in Fig. 1.1. The engineering principles of abstraction, standardization, modularity represent the basis of Synthetic Biology [7, 13] and allow to follow this rational approach. Synthetic Biology has focused on several applications, e.g., information processing, chemicals manipulation, materials and structures fabrication, energy and food production, human health and environment improving [4]. Beyond doubt, one of the most illustrative successful examples is the artemisinic acid production, a precursor of an anti-malarial drug called artemisinin, made by engineered yeast [14]. This chemical was extracted from *Artemisia annua* leaves and its cost was very high; thanks to the innovation introduced in [14], today it is possible to produce artemisinin on a large scale

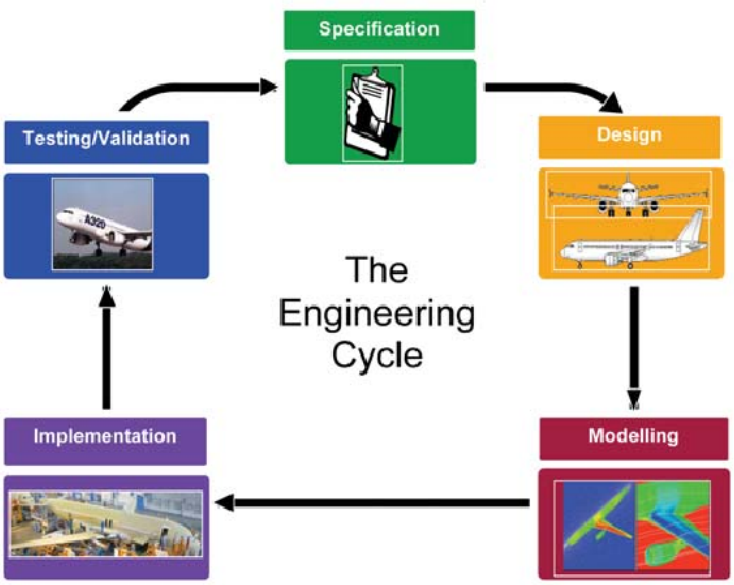


Figure 1.1: **The engineering cycle.** The engineering cycle begins with the definition of the specifications of a system, followed by the choice of the parts and their connections (design); then modelling is performed in order to predict the output and the behaviour; once the system is assembled it is tested and validated to understand if specifications are met and, if not, the system is rationally modified, on the basis of the results [4].

as demonstrated by the Garessio (CN, Italy) facility launched by Sanofi in 2013. Moreover many efforts have been recently done in the field of biofuels [15, 16, 17, 18]. These works demonstrate that Synthetic Biology has a broad range of application.
Synthetic Biology is considered similar to engineering disciplines, as it is based on the key concepts of abstraction, standardization and modularity that will be presented and discussed below.

1.1 Abstraction

Engineering disciplines provide a powerful tool to face complexity: abstraction [13]. Differently from other scientific fields, here this notion needs to be extended in order to be applicable to the biological field [7]. Abstraction means establishing a hierarchy of modules and devices, based on their complexity: users must be able to work at any level of complexity with limited exchange of information between levels, so that the design process can be performed at any level independently from one another [7, 13, 19]. Levels share common interfaces, in order to help communication. This strategy allows for the design of arbitrarily complex systems without the need to take into account unnecessary details [20] as useless information is hidden to manage complexity [8].
A proposed abstraction hierarchy that can be found in the scientific literature (Fig. 1.2) [13] places at the bottom level the DNA (i.e., a sequence of nucleotides); this can be considered as the physical layer in electronic engineering. At the upper level there are parts, intended as atomic functional DNA sequences like, for example, promoters. A combination of parts which performs a slightly complex task is considered a device, and finally, the interconnection of devices that carry out a function with higher complexity is considered a system. An alternative has been proposed in [7], keeping in mind the parallelism with the world of computer engineering: proteins and genes are the bottom level. Their combination results in a set of biochemical reactions which can, in turn, form pathways. These ones can be considered themselves as the basic components of cells, that finally could be grouped in tissues and cultures.

1.2 Basic parts

It is necessary, in order to clarify the concept of "part" in this biological context, to give a brief explanation of what actually is referred to as a part. In this work, a DNA sequence is considered a part and it can be classified on

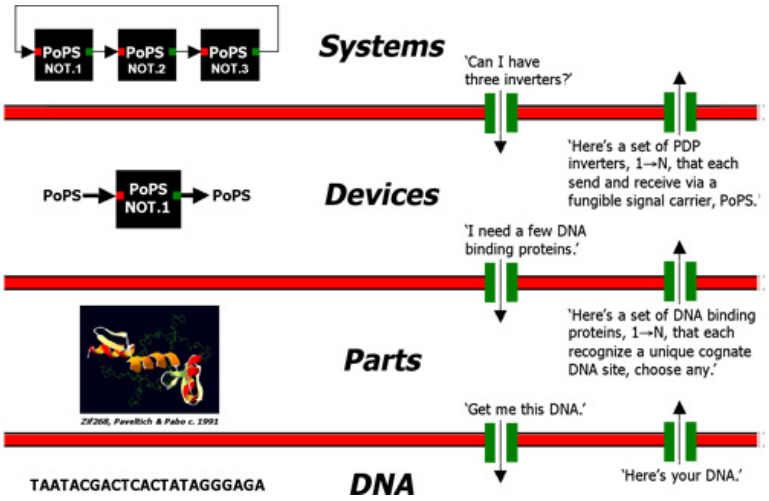


Figure 1.2: **Abstraction hierarchy.** Representation of one possible abstraction hierarchy [13]. From this point of view, the DNA represents the lower level, while, increasing complexity, one can find functional sequences that can be combined to build devices. At the top level there are complex combinations of devices, namely systems.

the basis of its basic function. In fact, depending on their function, parts can be grouped in several categories. The main categories will be here described.

1.2.1 Promoters

Promoters are DNA sequences placed upstream of genes, recognized by the RNA polymerase, which is the enzyme that transcribes to mRNA the message encoded in DNA. Promoters are responsible for transcription initiation. The *strength* of a promoter (i.e., the rate of transcription initiation) depends on its nucleotide sequence: a strong promoter is able to recruit more RNA polymerases, hence having a high transcription initiation rate and producing a large amount of the mRNA corresponding to its downstream sequence.
Promoters can have a constant transcriptional output; in this case they are called constitutive promoters. Conversely, due to the presence of operator sites, short sequences able to bind specific elements (transcriptional factors and/ or chemicals), their activity can be regulated (inducible promoters).

1.2.2 Ribosome Binding Sites

Ribosome Binding Sites (RBSs) represent the leading mRNA sequence to which ribosomes bind in order to initiate translation. Thus, DNA sequence of RBSs is placed upstream of the coding sequence and downstream of the corre-

sponding promoter. For this reason, RBSs are transcribed but not translated sequences. The main feature of RBSs is the Shine-Dalgarno sequence which is placed right before the START codon of the coding sequence (Sec. 1.2.3) and is responsible of mRNA binding to ribosomes. The sequence of an RBS is related to the translation initiation: a stronger RBS is a sequence which determines a higher translation initiation rate, hence a higher production rate of the protein of interest. In the last few years several efforts have been done in the processes of forward and reverse engineering of RBSs: a tool called "The RBS Calculator" helps the evaluation of the strength of an RBS and the design of sequences upon defined strength [21].

1.2.3 Coding Sequences

A coding sequence (CDS) is a DNA sequence that, once transcribed into mRNA, encodes the information needed by ribosomes to assemble a polypeptide chain; coding sequences are present downstream of RBSs sequences. A group of three nucleotides is called *codon*; every codon determines the amino acid to be bound to the polypeptide chain, depending on the genetic code of the host organism. The CDS is delimited by a START and a STOP codon. The protein is generally composed by three domains:

- an **N-terminal**, also called the head domain, that contains the START codon and, optionally, a sequence tag, namely a sequence that can be recognized by particular proteins of the host organism (e.g., export or attachment tags);
- a **protein domain** which represents the core of the polypeptide chain. It gives the polypeptide chain its ability to perform a particular task;
- a **C-terminal**, or tail domain, containing a STOP codon and with special optional tags, such as degradation tags.

1.2.4 Terminators

Terminators determine the end of the transcription process of RNA polymerase. They can be ρ-dependent or independent; in the first case a special protein called ρ factor with helicase activity is needed in order to unbind the RNA polymerase from the DNA fragment which is transcribing. For what concerns ρ-independent terminators, they are able to end the transcription process as they form hairpin structures that cause the separation of RNA polymerase from DNA. This latter class of terminators is characterized by sequences which, once transcribed, fold to form structures known as stem-loops that are responsible for the so called intrinsic termination.

1.2.5 Plasmid Vectors

Plasmid vectors are circular double-stranded DNA molecules used to deliver a genetic program into host cells. Generally their features are:

- a **replication origin**, which is a sequence determining the number of copies of the plasmid per cell. In fact, plasmid DNA replication is performed in parallel to chromosomal DNA replication;
- a **selection marker** which is a region that contains an expression cassette for a particular protein which confers the host cell special capabilities, for example the resistance to an antibiotic. This allows for the selection of cells that contain the plasmid vector;
- a **cloning site**, which is a region designed for the purpose to put in the plasmid a user-defined functionality. It is flanked by sequences called restriction sites, which are recognized and specifically cut by enzymes called restriction endonucleases. With molecular biology protocols it is possible to cut and ligate the plasmid vector in order to place desired sequences in the cloning site.

1.3 Standardization

Standardization is one of the key concepts of engineering disciplines; it allows to build interoperable systems or devices that share a common interface. Basic elements that meet established criteria can be interconnected simply and recursively. Moreover, standardization can also refer to the fact that common methodologies (e.g., measurements) are needed in order to share knowledge.
In Synthetic Biology the minimal biological part is a DNA functional unit, which cannot be further divided into smaller components [22]. A standard format for DNA sequences has been proposed, called BioBrick™ Standard [23, 24]; it simplifies the assembly of novel genetic sequences that can be easily shared and re-used. BioBrick™ parts are put in special plasmid backbones which have standardized features: their cloning site is flanked by a prefix sequence containing the EcoRI (E) and XbaI (X) restriction sites and by a suffix sequence containing the SpeI (S) and PstI (P) sites (Fig. 1.3). The four restriction sites must be unique, namely they must be present only in the prefix and the suffix but not in the plasmid vector or in the user defined sequence inserted in the cloning site. Plasmid backbones also have an antibiotic resistance as a selection marker and a replication origin that determines the copy number (Fig. 1.3).

This configuration allows users to digest and ligate the DNA so that parts can be combined and result in a new one that is compliant with the above

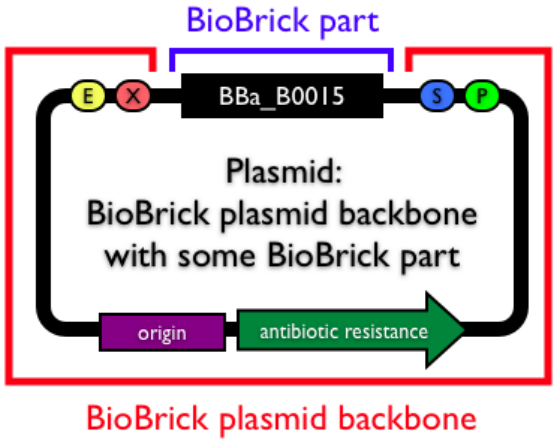


Figure 1.3: **Plasmid backbone.** Structure of a plasmid vector compliant to BioBrick™ Standard [22]. In this schematic representation the main features are highlighted. They are: the replication origin, the resistance marker and the cloning site. This is flanked by the four restriction sites compliant to the BioBrick™ Standard: EcoRI (E) and XbaI (X) in the prefix, SpeI (S) and PstI(P) in the suffix.

mentioned standard (see Fig. 1.4). This way, multiple rounds of insertion on either side of a part are easily possible via a reproducible procedure involving the same four enzymes [8].

A collection of BioBrick™ parts, called Registry of Standard Biological Parts, is held by the Massachusetts Institute of Technology (MIT); parts are catalogued with their names that follow a standard code. Moreover, an on-line documentation is provided [22], and information about their usage and functioning can be easily retrieved.
Although a lot of information is freely and easily accessible on BioBrick™ parts, it is necessary to define basic shared procedures to reliably measure biological parts activity [13]. These might allow standardized quantitative descriptions for biological devices in order to easily establish if parts meet the needed specifications during the design of a new system.
Like in other engineering disciplines, it could be useful to have standardized data-sheets to retrieve information on a device about [25]:

- its function and interfaces (i.e., the input and output);
- the operating context of its characterization, for example the genetic background (the host organism) and the growth conditions (growth medium, temperature and pH);
- its quantitative characterization that could be based on mathematical modelling.

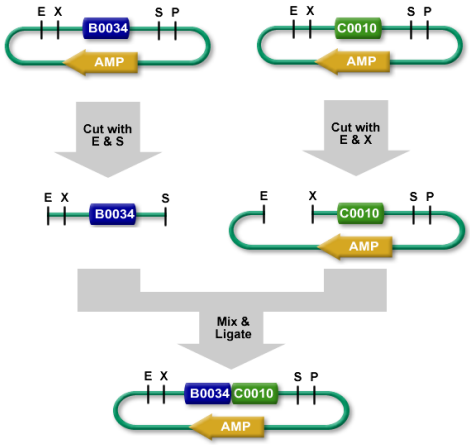


Figure 1.4: **Standard assembly.** Assembly of two BioBrick™ basic parts: BBa_B0034 is digested with EcoRI and SpeI restriction enzymes, while BBa_C0010 with EcoRI and XbaI. Since XbaI and SpeI are isocaudomers, their protruding ends are complementary, so they can be ligated and result in a mixed site that is not recognized by any of the four above mentioned enzymes. At the end of the process the resulting composite part is compliant to the BioBrick™ Standard [22].

A clear example of a standardized measurement procedure is the quantification of promoters activity in terms of Relative Promoter Units (RPU) [26]. Reporting the output of a promoter of interest to that of a standard reference promoter is very useful as it is possible to have reliable measurements across different experimental conditions or instruments. The RPU measurement approach is based on a simple mathematical model that captures the main features of promoter activity. In particular, Polymerases per Second (PoPS), defined as the number of RNA polymerases that pass through a specific point of a DNA strand per time unit, is an interesting parameter which describes the transcription rate of a downstream gene caused by the promoter. By using mathematical modelling and a gene encoding a fluorescent reporter protein downstream of the promoter of interest, PoPS can be indirectly estimated. The mathematical model is reported in Eq. 1.1:

$$\begin{aligned} \frac{d}{dt}[M] &= n \cdot PoPS - \gamma_M \cdot [M] \\ \frac{d}{dt}[I] &= \rho \cdot [M] - (a + \gamma_I) \cdot [I] \\ \frac{d}{dt}[X] &= a \cdot [I] - \gamma_X \cdot [X] \end{aligned} \tag{1.1}$$

where $PoPS$ is the transcription initiations per second per DNA copy while $[M]$ is the mRNA per cell concentration, $[I]$ the immature protein per cell concentration and $[X]$ the mature protein per cell concentration. γ_M and γ_I are the degradation rates of mRNA and immature protein, respectively, and γ_X the degradation rate of mature reporter protein. The degradation terms named

γ take into account both the dilution due to cultures growth and the specific degradation of the species: in the case of mRNA the specific degradation rate cannot be negligible, while, considering the high stability of the reporters used to perform these measurements, the dynamics is dominated by the dilution due to cells duplication. a is the maturation rate of the reporter protein and n is the average copy number of the plasmid bearing the promoter of interest. Finally ρ is the synthesis rate of immature protein per second per mRNA. Assuming the steady state, namely $[\dot{M}] = 0$, $[\dot{I}] = 0$ and $[\dot{X}] = 0$, it is possible to express the $PoPS^{SS}$ (in which SS superscript stands for steady state) in terms of mature reporter protein concentration:

$$PoPS^{SS} = \frac{(a + \gamma_I) \cdot \gamma_M \cdot \gamma_X \cdot [X^{SS}]}{n \cdot \rho \cdot a} \tag{1.2}$$

Finally, the synthesis rate per cell of reporter mature protein is defined as in Eq. 1.3.

$$S_{cell} = a \cdot [I] \tag{1.3}$$

Hence, it is possible to calculate the $PoPS^{SS}$ as in Eq. 1.4.

$$PoPS^{SS} = \frac{(a + \gamma_I) \cdot \gamma_M \cdot S^{SS}_{cell}}{n \cdot \rho \cdot a} \tag{1.4}$$

The RPU value of a promoter Φ is defined as the ratio between its $PoPS^{SS}$ value and the $PoPS^{SS}$ value of the standard reference promoter, which is the BBa_J23101 constitutive promoter from the Registry of Standard Biological Parts [22]. As a result, the activity of a promoter Φ relative to the reference promoter, can be derived as reported in Eq. 1.5.

$$RPU = \frac{\frac{(a_\Phi + \gamma_{I,\Phi}) \cdot \gamma_{M,\Phi} \cdot S^{SS}_{cell,\Phi}}{n_\Phi \cdot \rho_\Phi \cdot a_\Phi}}{\frac{(a_{ref} + \gamma_{I,ref}) \cdot \gamma_{M,ref} \cdot S^{SS}_{cell,ref}}{n_{ref} \cdot \rho_{ref} \cdot a_{ref}}} \tag{1.5}$$

Assuming that:

- the maturation rate of the reporter protein is the same between the two different conditions;
- the average copy number of the plasmid of the reference and the promoter of interest is identical;
- the transcription initiation site and the downstream mRNA are identical, so that the two degradation rates and the two synthesis rates of immature protein per mRNA per second can be considered equal;

- the dilution rate of the protein, equal to the culture growth rate ($\gamma_I = \mu$), is negligible compared to the protein maturation rate a;

it is possible to evaluate the RPU as in Eq. 1.6.

$$RPU = \frac{S^{SS}_{cell,\Phi}}{S^{SS}_{cell,ref}} \tag{1.6}$$

Thus the activity of the promoter can be easily evaluated on the basis of its synthesis rate per cell divided by the synthesis rate per cell of the reference promoter BBa_J23101. This is an easy-to-measure quantity because, given an instrument which allows to monitor the culture growth and the activity of a reporter, it is possible to derive the S^{SS}_{cell} as the time derivative of the reporter signal over time, divided by the culture density expressed, for example, in terms of optical density (OD):

$$S^{SS}_{cell} = \frac{dF}{dt} \cdot \frac{1}{OD} \tag{1.7}$$

Usually, in order to obtain robust measurements, it is convenient to consider the average of this quantity during an appropriate period (for example during the exponential phase of bacterial growth when the above assumptions can be considered to be valid) [26, 25]. Although the RPUs represent a reliable measurement for the transcriptional activity of promoters, a major challenge in their characterization is still their context dependent behaviour; in fact the observable output can vary depending on the upstream or downstream sequence. Several efforts have been done towards this goal, for example by creating a library of insulated promoters that behave in a predictable manner across different genetic contexts [27]. Other parts such as RBSs and terminators are intrinsically context dependent and this strongly affects the measurement of their performances; thus, standardization in measuring biological parameters linked to these elements is a challenging task. Biophysical models have been recently developed to face these problems and give information about the working of these elements. Biophysical models applied to Synthetic Biology try to capture the intermolecular and intramolecular interactions that regulate processes like binding of enzymes to mRNA or mRNA folding, and describe their strengths in terms of free energy. A useful tool developed by Salis et al. [21, 28] allows to predict RBSs' translation initiation rate on the basis of the surrounding sequences or to design the sequence of an RBS with a user-defined output, relying on a thermodynamic model and optimization. Similarly, thermodynamic models, coupled to linear models, were used to predict the behaviour of a library of terminators [29].

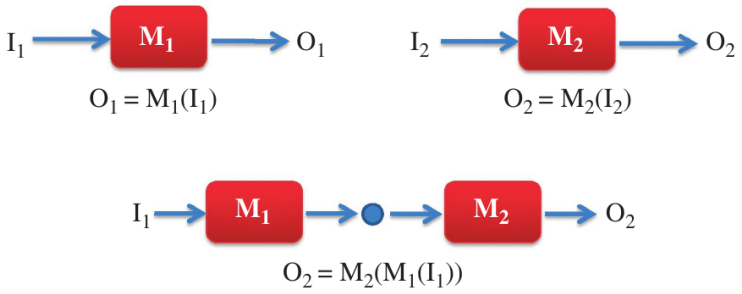


Figure 1.5: **Modularity.** The concept of modularity is herein represented: in a modular framework, given two quantitatively characterized modules (in terms of transfer function), it is possible to quantitatively predict their output when interconnected [31].

1.4 Modularity

Modularity is fundamental in all engineering applications because it allows to build up complex systems, obtained by assembling basic functional modules, with predictable behaviour [30, 31]. After basic functional modules are fully characterized, it is possible to assemble them to obtain more and more complex functions, relying on this property. Once modularity is verified it allows to combine modules guaranteeing the integrity of their properties upon interconnection [32]. That is, if the input-output transfer function of N modules is known, it is possible to interconnect them and to predict the output of their interconnection [31]. In fact, decomposing an interconnected system in smaller modules, means, on the one hand, to reduce complexity and, on the other hand, to design modules and engineering systems more efficiently. In some cases, given a precise characterization, it is possible to predict the output of interconnected systems but in other cases this is not sufficient as the parts may not work as expected when put together [33].

In a biological context, the environment-dependent behaviour of some devices, the interaction with the cell machinery, side effects on their metabolism, intrinsic noise and mutations, make living cells extremely complex systems and these and other elements affect the possibility to modularly program new predictable functionalities. Several works have been performed in order to investigate the degree of modularity of parts and solve the issues related to the interconnection of devices [32, 34, 30, 31]. Also in other disciplines, such as electrical or mechanical engineering, modularity is not always valid but context-dependent effects have been well identified and systems can be predictably composed from characterized basic parts given the working context. For example, temperature is known to affect the impedance of electrical compo-

nents, such as resistors, but designers can take into account such quantitative effect by considering the temperature-impedance characteristics of the component of interest during the design process of circuits. Designers are also able to predict circuit behaviour changes upon interconnection with a downstream load. Considering the latter example, the term "retroactivity" has been introduced in biological engineering to describe an analogous situation to non-zero impedance in electrical circuits, which can lead to interconnection-dependent behaviour changes during genetic network design. In particular, in [35] the issue of retroactivity is deeply discussed in order to understand how this phenomenon can affect the dynamical behaviour of genetic circuits. Retroactivity is due to signals travelling in the opposite direction (i.e., from downstream to upstream) to information [36]; in Fig. 1.6 the two signals that represent retroactivity are shown. Signal s is called retroactivity to the output and is due to the interconnection of the system Σ to a downstream module; signal r is the retroactivity to the input and represents the change in the dynamics of Σ when it is stimulated by an input u. Thus, retroactivity emerges when modules are interconnected and leads to unwanted effects on system output; this occurs, for example, in presence of unwanted repressor binding sites in a downstream circuit that bind a repressor protein involved in an upstream circuit. As demonstrated in [32], the effect of this unpredicted interaction is clearly visible in the activation and de-activation dynamic behaviour: a test circuit, inducible by the chemical Isopropyl-β-D-1-thiogalactopyranoside (IPTG), has been used to trigger gene expression and the dynamics of activation and deactivation in the presence of extra operator sites for the LacI repressor. The behaviour of the circuit was monitored via reporter gene. In both cases a shift in time was observed (a delay or an anticipation) for the response, compared to the circuit without additional operator sites. Hence, this work underlines the importance of studying not only the input-output relation of a system at its steady state but also its dynamics features, in order to better understand its behaviour upon interconnection.

Another phenomenon that impacts on the output of a synthetic genetic regulatory network is the fan-out, a concept which is imported from electrical engineering and is related to retroactivity: it expresses the maximum number of downstream clients that an upstream output transcription factor can regulate without delay or signal attenuation. In fact the interconnection of a set of downstream outputs can affect the output impedance and the working of the circuit. Analogously a genetic network of transcriptional regulators could stumble this problem, affecting the behaviour of the input module and consequently the output of the whole system [37]. In [38] a method to measure the fan-out has been proposed to experimentally characterize the interfaces between modules and quantifying the level of modularity of a gene regulatory

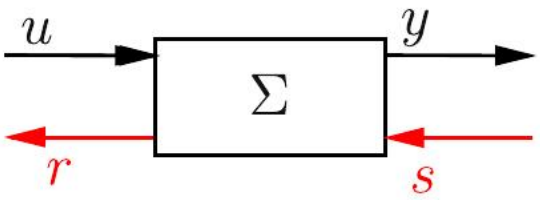


Figure 1.6: **Retroactivity.** Retroactivity is an unwanted effect that is due to the fact that signals could move from downstream to upstream, that is, in the opposite direction in which information is travelling. u represents the input to the system Σ, while y is its output. Red arrows are the two signals linked to retroactivity: r is the retroactivity to the input, that determines changes in Σ when it is stimulated with an input, while s is the retroactivity to the output that emerges upon interconnection with a downstream module [36].

network.

In order to face these problems and reduce the time needed to design and realize new functionalities in living systems it is important to quantitatively characterize basic parts, relating the output of devices to their inputs and to the context of application. The achievement of robust predictions is not trivial as there are several sources of uncertainty: the effects of noise in gene expression, mutation, cell death, undefined and changing extracellular environments, and interactions with cellular context; Synthetic Biology must tackle these hard problems by providing predictable and scalable design for the genetic parts [7, 39].

A powerful tool to manage uncertainty and aid the predictability is mathematical modelling as it enables to describe systems using a limited set of parameters. Nowadays it is possible to define arbitrarily sophisticated models as personal computers have very high computing power, even though there are peculiarities linked to intrinsic noise of biological systems that are not easy to model [40]. Models can be defined prior to physically assembling a system in order to predict its output and verify if there is a configuration able to meet the specifications; they can also be helpful in the engineering cycle to predict how a system could behave upon changes in the working of some of its basic parts, without the need to perform several trial-and-error experiments [41]. Moreover, the comparison between predictions and experimental data could point out inaccuracy in some hypotheses on biological systems and, through this, explain the possible malfunctioning of synthetic devices [4].

Different systems were designed and investigated through mathematical modelling, in order to study the modularity in Synthetic Biology. For example in [42] a silencing mechanism called post-transcriptional controller based on a TRANS-CIS interaction that inhibits translation was used to test a pure

bottom-up approach, using a mathematical model to predict the behaviour of the system. The system is composed of a sequence that constitutively produces a Green Fluorescent Protein (GFP) whose mRNA is modified in order to have a CIS regulatory sequence, and a regulated production of a TRANS element that is able to block the translation of GFP. Sub-systems were used to estimate the parameters of the mathematical model and evaluate its output. In this case predictions are not always aligned with the output of the system, highlighting that a pure bottom-up approach is not always effective in designing and assembling new functionalities.
Other works focused on the design and implementation of logic gates in bacteria, for example AND gates. An AND gate is a device whose output is a logical one if and only if all of its inputs are logical ones; otherwise the output is a logical zero. In the field of Synthetic Biology AND gates have been proposed in several works and efforts have been done in order to obtain modular systems. For example in [43] the authors proposed a biological AND gate in *E. coli*, which works with two promoters as inputs that regulate the transcription of two factors that, together, activate an output promoter whose transcription activity depends on the T7 RNA polymerase. The two input promoters are inducible systems while the output of the AND gate was measured by the GFP reporter. A simple model of the steady-state behaviour of the logic gate has been derived in order to make predictions of the behaviour of the whole system once its sub-parts have been characterized. Even though, as demonstrated by the authors, the system is modular in the sense that any inputs or outputs can be placed instead of the ones presented above, there were still some issues concerning the RBSs of the input genes. In fact, in order to obtain an AND gate behaviour, RBSs had to be chosen from a saturation mutagenic library, so that for every input the sub-part has to be re-characterized.
A similar work is presented in [44], where the authors designed another orthogonal AND gate realized in *E. coli*. In this work sub-parts have been characterized in-context, that is, they have been tested in the same experimental conditions as the whole system, focusing on strain, RBSs, medium and temperature. After having estimated the parameters of a model analogous to the one presented in [43] on the sub-parts, a set of candidates has been chosen to realize the desired behaviour. Moreover, a NAND gate, a logic operator which presents a logic one as an output only when all of its inputs are logical zeros, has been realized by connecting a NOT gate downstream of the AND gate. All of these systems have been successfully realized with a forward engineering approach, keeping in mind the context of sub-parts characterization. Alternatively, in order to obtain a predictable complex system upon interconnection of simple modules, the authors proposed to deal with genetically or physically insulated parts, as stated and done in other works (e.g., as in [45]).

More complex systems were realized in [46], where a set of activator-chaperone pairs have been identified from bacteria and used in *E. coli* to realize several AND gates and interconnect them in a layered fashion (Fig. 1.7); some of them were modified in order to meet design specifications. After having characterized AND gates for each activator-chaperone pair with different signals as inputs, the authors realized layered programs reaching a maximum of 4 input AND gate. Again, a model-aided strategy was used to make predictions of the behaviour of the final system. The layered fashion in which the gates are connected could lead to problems in the output of the network: in fact the progression of signals may not be synchronous, leading to faults and incoherent results.
These works demonstrate that modularity is still a major issue in Synthetic Biology and that it is not always possible to design and realize a complex system with a pure bottom-up approach unless by taking into account a large number of variables such as the genetic context and the environment.

1.5 Electronic engineering inspired genetic circuits: state of the art

Several topologies of genetic regulatory networks have been considered and studied as a proof of concept of the possibility of engineering micro-organisms capable of carrying out a specific task. Hereinafter, a set of works will be presented, in which the authors designed and studied different regulatory networks inspired by electric circuit or control theory.

1.5.1 AND gate

One of the basic logic operations in digital electronic is the one performed by the AND gate; this device is connected to a set of inputs and elaborates the information so that its output is a logical one only when all of its inputs are true. Several works focused on the possibility to design and realize a biological AND gate.
In [46] the authors worked on a system based on promoters activated by a complex of a transcriptional factor and its cognate chaperone protein (Fig. 1.7), presenting a set of AND gates with two inputs implemented in *E. coli*. According to this design, the promoter output is high only when the transcription factor and the chaperone are simultaneously expressed.

These chaperone-protein dependent systems were insulated from other organisms, *Salmonella typhimurium*, *Shigella flexneri* and *Pseudomonas aeruginosa*, thus they have to be characterized and, in some cases, modified, in order

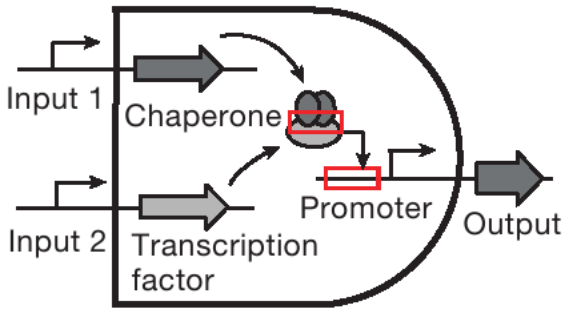


Figure 1.7: **AND gate scheme.** Representation of the generic AND gate realized in [46]; different realizations rely on different chaperone-transcriptional factor pairs able to activate a set of target promoters, selected from several organisms.

to work in *E. coli* and extend their activity range and orthogonality. The AND gates were studied and characterized using inducible promoters as inputs, such as P_{Tet}, P_{BAD} and P_{Lux}, inducible by anhydrotetracycline (aTc), arabinose and 3-oxohexanoyl homoserine lactone ($3OC_6$-HSL), respectively. A mathematical model of the AND gate was derived; its parameters were identified on subsystems and used to guide the assembly of the genetic circuits and predict their outputs. Moreover AND gates were interconnected to generate more complex systems; the mathematical model was extended and used again to successfully predict the output of these networks. Even though results were in accordance with the model prediction for networks with a small number of layers (2-3), the connection of a large number of layers could result in heavy issues due to the pulses that can emerge in incoherent feed-forward loops in signalling networks. This means that signals are not synchronized, thus affecting the correctness of the output.

1.5.2 Toggle-switch

One of the milestones of Synthetic Biology is a device implementing a genetic toggle-switch, realized in engineered *E. coli*. The concept of toggle-switch refers to a particular class of switches allowing to drive electric circuits between two states; thus they belong to the class of bistable circuits. The behaviour of these devices is particularly interesting in Synthetic Biology as, by a genetic network of transcriptional regulators, it can be reproduced in a cell culture. In [47] a model based approach has been adopted in order to obtain a set of toggle-switches; a simple mathematical model was used to investigate the circuits behaviour and the conditions that lead to bi-stability. The genetic network that reproduces the toggle switch in living systems is simple, as it is

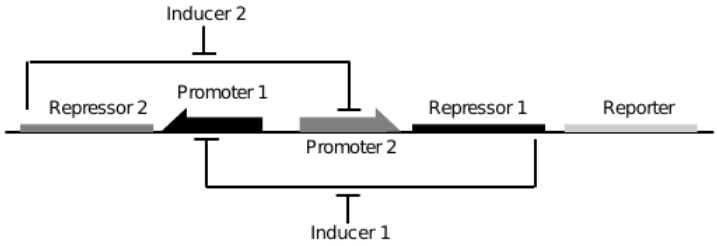


Figure 1.8: **Toggle-switch.** Schematic representation of a biological toggle-switch [47]; the genetic toggle-switch was realized using a promoter-repressor pair, repressing each other. The reporter is co-expressed as an operon with one of the two repressors, in order to monitor the state of the system.

composed by a promoter pair and the respective repressor molecules; a fluorescent reporter is used to monitor the status of the biological device (Fig. 1.8). The circuit can switch between two opposite conditions: when repressor 1 reaches a concentration capable of switching off the transcriptional output of promoter 1, the transcriptional activity of promoter 2 is maximal and the circuit reaches the ON state; conversely, when repressor 2 is capable of blocking the promoter 2, the circuit is in its OFF state.

Switching between the ON and OFF states and viceversa is accomplished by using chemical or physical stimuli, able to inhibit the repression of one specific repressor protein.
A mathematical model of this topology has been derived to define the specifications that have to be met to reproduce the electric toggle-switch behaviour (Eq. 1.8).

$$\begin{aligned} \frac{d}{dt}u &= \frac{\alpha_1}{1+v^\beta} - u \\ \frac{d}{dt}v &= \frac{\alpha_2}{1+u^\gamma} - v \end{aligned} \tag{1.8}$$

The two species u and v represent the concentration of repressors, α_1 and α_2 the promoters maximal transcriptional activities that are modulated by the concentration of the two repressors through the cooperativities β and γ. On the basis of the model, the authors identify the conditions to obtain such a circuit: the strengths of the two promoters have to be balanced and the cooperativities have to be greater than one. In this case it is possible to obtain a network with an unstable equilibrium and two stable equilibria with their respective basins of attraction. On the basis of the parameters it is possible to mathematically derive the conditions that determine the separation between the two basins of attraction. In the practice these conditions are satisfied tuning the transcriptional output of the promoters used in the assembly of

the genetic network; this means that the output of the promoters, in terms of repressor concentration, has to be tuned. This can be achieved varying the strength of the RBSs used to drive the traslation process. Two main versions of the toggle-switch have been designed and assembled: one based on the couple of promoters P_{Lac}, inducible by IPTG, and P_L, a promoter from λ-phage, which can be repressed by the cI thermo-stable protein, which degrades when temperature is raised from 30°C to 42°C; the other version of the genetic toggle-switch is based on P_{Lac} and P_{Tet}, which is a promoter inducible by anhydrotetracycline (aTc). Several RBSs were assembled in the two versions of the circuit and the authors demonstrated the possibility to assemble functional genetic toggle-switches, satisfying the conditions suggested by the mathematical model, even if not all the combinations worked as it was designed. Experimental tests showed the possibility to switch from the OFF to the ON state and go back again to the OFF state; transition from one state to the other can be smooth or present a kind of discontinuity. Also the switching dynamics were taken into account and studied. The importance of this work lies not only in the demonstration of the possibility to build up a digital-like genetic circuit but also on the use of mathematical modelling tools to support the early design steps of the device.

1.5.3 Oscillators

Another milestone of Synthetic Biology is a circuit implementing a ring oscillator in *E. coli* [48]. Like the previous system, it has been designed with the support of a mathematical model in the early design steps, to investigate its feasibility and the choice of components with qualitative behaviour compatible with the expected one. A ring oscillator is a device composed by an odd number of NOT gates; these are connected in cascade, with the output of the last one fed back to the input of the first one, so that the system is able to oscillate autonomously between two logic states, ON and OFF. The *repressilator* is the biological counterpart of a ring oscillator: it is basically built up of a cascade of three inverters and a reporter, a GFP, which gives information about the state of the system (Fig. 1.9). The authors derived a mathematical model of this genetic network in order to find out if the system could actually act as a ring oscillator. The model describes the time course of the cellular concentrations of the three transcriptional regulators chosen to realize the *repressilator*: LacI, TetR and cI, which respectively repress the P_{Lac}, P_{LtetO1} and P_L promoters. The negative feedback loop, essentially caused by three logic inverters, leads to temporal oscillations in mRNA and protein levels. The equations describe the repressor protein concentrations p and their corresponding mRNA concentrations m; as in [48] the indexes i, j denote the three species (i=LacI, TetR,

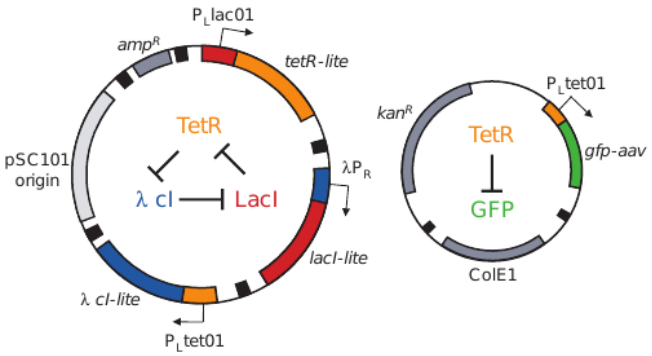


Figure 1.9: **The *repressilator*.** The *repressilator* [48]: the plasmid on the left bears the three repressors, LacI, tetR and cI, expressed by P_L, P_{Lac} and P_{LtetO1} promoters, respectively. The other plasmid is used to measure the state of the genetic circuit, monitoring GFP fluorescence.

cI; j=cI, LacI, TetR, respectively):

$$\begin{aligned} \frac{d}{dt}m_i &= -m_i + \frac{\alpha}{(1+p_j^n)} + \alpha_0 \\ \frac{d}{dt}p_i &= -\beta \cdot (p_i - m_i) \end{aligned} \tag{1.9}$$

In Eq. 1.9 the mRNA concentration is re-scaled by the translation efficiency (ρ) while the protein concentration is re-scaled by the number of repressors needed to half-maximally repress the promoter (K_M); α is the maximum activity of promoter and α_0 its basic activity; time is rescaled by mRNA degradation rate; n is the Hill coefficient and, finally, β is the ratio of protein degradation rate and mRNA degradation rate. The described model has been studied in a symmetric framework, where amounts and parameters (α, β, α_0, and n) of the three repressors are identical, except the specificity towards their corresponding promoter. Depending on the parameter values, the system can converge to a stable equilibrium or to a sustained limit-cycle, leading to oscillations. In order to meet the specifications for the oscillatory behaviour some precautions were taken: repressor proteins were modified in order to have a short half-life through a *ssrA* tag, which is recognized by proteases, while the reporter half-life was reduced to 40 minutes [49]. Despite the rational approach a few problems remain; in fact oscillations with a period of 160 minutes were observed but the population was not synchronized because of noise in gene expression, as revealed by single cell analysis. Another source of noise is cell replication, which introduces another variability source in this genetic network. A different topology was used in [50] to design an oscillatory network, coupling a positive and a negative feedback loop in which a hybrid promoter, $P_{Lac/Ara}$, inducible by arabinose complexed with the activator AraC and repressible by

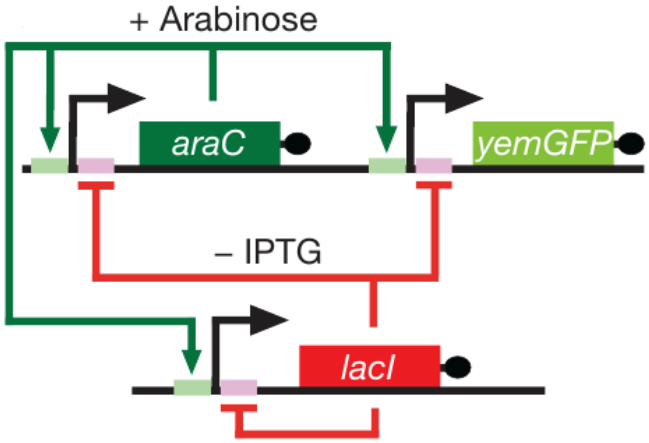


Figure 1.10: **An example of oscillatory genetic circuit.** Positive and negative feedbacks coupled to obtain oscillations [50]; the coupling of both positive and negative feedbacks was obtained using a hybrid promoter, $P_{Lac/Ara}$, which is at the same time induced by AraC and repressed by LacI. It was placed upstream of the coding sequences of *lacI*, *araC* and *yemGFP*.

LacI, was placed upstream of the coding sequences of *lacI*, *araC* and a reporter: the monomeric yeast-enhanced green fluorescent protein (*yemGFP*) (Fig. 1.10). In this way, the genetic network exhibits temporal oscillations of LacI and AraC in presence of arabinose and absence of IPTG. Moreover, the oscillation dynamics could be tuned by adding different amounts of IPTG and/or arabinose, which affect the concentration of active LacI and AraC, respectively. The behaviour of the system was assayed by monitoring the cells behaviour through a microfluidic device which maintains the nutrient flow constant for the duration of the experiments; the behaviour of the circuit resulted to be robust, as a high percentage of the cells showed oscillations (99.5 %) which persisted even in different experimental conditions such as broth type, temperature and inducers concentration. As before, the cells were initially synchronized but they lost synchronization after a limited number of replication cycles, leading to dump oscillations due to the asynchrony of the cells' state. This circuit was also described by a mathematical model. During model revision it was noted that different parameters values could lead to oscillations of the output; moreover, on the basis of the model, the authors were also able to predict that a constantly activated system with a controlled negative feedback loop could also work as an oscillator, in the absence of positive feedback [50]. Thus, a new topology, where LacI inhibits its own production and the production of yemGFP (Fig. 1.11), was designed and tested, revealing that in this case oscillations persist even if they resulted to be less regular and less

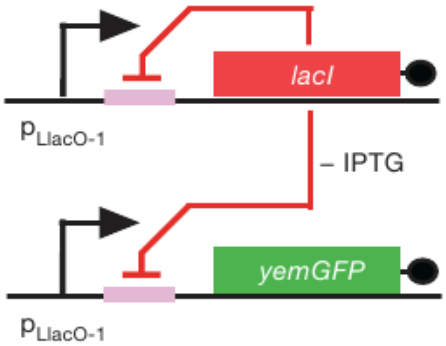


Figure 1.11: **Alternative topology of oscillator.** An alternative topology proposed in [50] for oscillating cell cultures; in this topology LacI inhibits its own production and the production of yemGFP, leading to temporal oscillation of the two species.

tunable by IPTG concentration.

Applying the same design principles as the two previous works, in [51] the authors managed to synchronize the oscillations of a population of *E. coli*. The genetic regulatory network in Fig. 1.12 exploits the genes responsible for the *quorum sensing* mechanism of *Vibrio fischeri*, encoding for the transcriptional factor LuxR and the autoinducer synthetase LuxI which produces the signalling molecule AHL. This molecule activates the transcriptional regulator LuxR which is then able to activate the transcription of the P_{Lux} promoter. Another quorum sensing-linked enzyme, the lactonase AiiA, is able to degrade AHL. LuxR is consitutively expressed, while *luxI*, *aiiA* and *yemGFP* CDSs are placed downstream of P_{Lux} promoters. The dynamic production and degradation of AHL, upon expression of LuxI and AiiA, respectively, leads to temporal oscillations of autoinducer and, as a consequence, of yemGFP. Synchronization is due to the nature of the regulatory signal, which is diffusible throughout the population and thus it is not affected by cell-to-cell variability as in the other oscillator designs. A peculiar microfluidic system has been used to study the behaviour of this genetic circuit: it is composed by a small chamber, where cells are trapped, and a channel able to feed the chamber with fresh medium flowing at user-defined velocity. In this way, cells and also the signalling molecule can be washed out and this has to be taken into account as the flow velocity is responsible for the AHL concentration and, hence, of the oscillations period. Because AHL is washed away and enzimatically degraded by AiiA, when cell density is low the P_{Lux} promoter is not able to start transcription; however, when the cell density reaches a critical value the P_{Lux} promoter is activated and the three genes are transcribed. For this reason it is possible to see a transient fluorescent signal, which then attenuates because of the short reporter

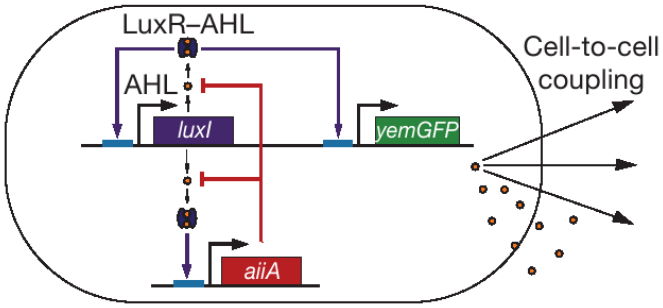


Figure 1.12: **Synchronized oscillator.** Schematic representation of the regulatory network that leads to synchronized oscillations in *E. coli* [51], exploiting the *quorum sensing* mechanism of the marine bacterium *V. fischeri* to synchronize the cells behaviour. *luxI*, *aiiA* and *yemGFP* are placed downstream of the P_{Lux} promoter; in this configuration the production and degradation of AHL lead to the oscillation of yemGFP signal.

half-life and because the AiiA enzyme degrades the AHL signalling molecule so that P_{Lux} promoter transcriptional activity is no more induced. Here the AHL has a double role: it is responsible for the transcriptional activity of the three genes and also is the way through which cells communicate and synchronize their state. A mathematical model was used to quantitatively describe the behaviour of this oscillator, finding that a broad range of parameters could lead to oscillations; moreover, through simulation, it was demonstrated that if communication between cells is avoided oscillations cannot be synchronized anymore.

1.5.4 Pulse generator

Another interesting circuit that was considered in Synthetic Biology is the pulse generator, which is a system capable of transiently switch on a pattern generating a pulse with peculiar features, like, for example, its duration. The pulse generator has been designed and assembled in [52] by using AHL as input signal, which triggers the pulse generation. AHL can be either provided exogenously or produced by a population of sender cells containing a LuxI expression system. In order to realize such a genetic circuit in a general way, a population of sender cells and a population of receiver cells were designed (Fig. 1.13): senders harbour an inducible circuit that produces AHL upon aTc induction, while receivers, i.e., the pulse generator cells, harbour a plasmid for the constitutive production of LuxR transcriptional factor which, in the presence of sufficiently high AHL concentrations, activates the P_{Lux} promoter. This promoter has been assembled upstream of the coding sequence of the cI

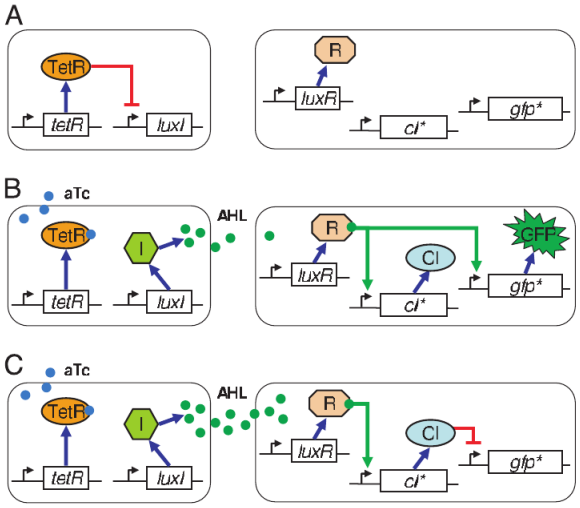


Figure 1.13: **Pulse generator.** Working of the pulse generator [52] when two different cell populations are present: a sender population that generates the signal represented by AHL which is in turn sensed by the receiver population that implements the pulse generator.

repressor, while a modified P_{Lux} promoter, with an additional cI operator site, drives the transcription of an *ssrA*-tagged green fluorescent protein (GFP). This way, when AHL induces the transcription of the P_{Lux} promoter, cI and GFP are produced; however, when the cI repressor reaches a critical intracellular concentration, it binds the operator site placed downstream of the modified P_{Lux} sequence, thus blocking the transcription of GFP. As a result, after AHL induction, fluorescence increases, reaching a maximum value and then, after the cI repressor blocks GFP production, it turns off again.

The authors first focused on the pulse generator cells in liquid cultures, maintained by dilutions at optical densities of 0.1 - 0.3: a small library of circuits including RBSs of different strengths upstream of cI and two different cI repressors (with different affinities towards the operator sites) was constructed to test the possibility to obtain the desired behaviour and to tune the pulse characteristics; it was found that not all the combinations lead to a pulse generator-like genetic network. One of the functional candidates was tested and several features of the circuit were discovered and investigated. First of all, the pulse amplitude, i.e., the maximum fluorescence produced by cells, depends on the concentration of AHL, reaching a saturation value around 50 nM. Then the authors studied the behaviour of the system when AHL is added at different rates: in this case they discovered that the faster the induction, the faster the shut down. This was probably due to the delay introduced in the system by the transcription and translation processes regarding the *cI* repres-

sor. It was also found out that after the first pulse a refractory period occurs, thus obtaining a fluorescence intensity similar to the first pulse after about 140 minutes.
The pulse generator was also studied in terms of spatio-temporal behaviour, spotting a sender on a solid medium and observing the fluorescence of surrounding pulse generator cells. As expected, the nearest cells acted as a pulse generator, while the most distant cells did not induce. Obviously, the signal propagates with finite velocity, so that the pulse of the farther cells is delayed if compared to the one of the nearest cells.
Moreover, a mathematical model with 23 species has been derived; its 44 parameter values were mainly obtained from the literature or given reasonable values. Focusing on the transcription, translation, and decay of cI and GFP, 1000 sets of randomly chosen parameters were taken into account to find the combinations that best matched experimental observations [52]. Once the best combination was identified, the model was used to further investigate the circuit behaviour: the model well described the behaviour of the system, although it was not used for a pure bottom-up approach. However, it could be useful when assembling new genetic networks similar to or relying on the pulse generator.

1.6 The importance of negative feedback regulation and examples of application

Arising from the control theory, in the last few years the assembly of regulatory networks capable of reacting to changes in environment, controlling interesting pathways, has become more and more important. In fact there are a lot of examples of pathways that need to be dynamically regulated, in order to maximize the yield of the production of the molecule of interest. Up to the present, several works focused on the possibility to rationally design and predict the level of expression for a particular gene of interest as a function of the promoters strength [26, 30], the RBSs translation initiation rate [21, 28] and the plasmid copy number [53]. In this framework, computational methodologies have also been developed, with the aim of studying gene deletion strategies to maximize the fluxes of interest [54]. All of these strategies aim at determining an optimal flux through a pathway but they are poorly flexible as there is not any sensor or feedback which could modify the flux dynamically regarding, for example, cell density or the amount of a specific accumulated molecule. On the other hand dynamical control could lead to better yields as it can prevent the accumulation of toxic intermediates modulating fluxes but it is certainly

much more difficult to achieve. Hence there is a trade-off between the easiness of statically balancing a flux and dynamically controlling it, raising the yield [55]. From these arguments derives the need to rationally design and realize genetic regulatory networks that are able to sense the presence of toxic by-products that can affect the yield of a process; for example, a robust negative feedback controller can be useful in order to flexibly regulate the expression of a gene of interest regarding some important parameters.
Negative feedback could be a solution to face the problem of dynamically regulating the output of a genetic circuit or reducing noise in gene expression. It has been demonstrated in [56] that, with different topologies, a genetic regulatory network is able to reduce noise, shifting its frequency from low to high values; this is very important as downstream modules can easily filter-out this component of noise that cannot impact the system output anymore. Results were confirmed through mathematical modelling of the three different topologies that were used as proof of concept. Another important proof of concept was realized in *Saccharomyces cerevisae* where a testbed circuit, controlling the expression of a target gene fused with a GFP as a reporter, has been modelled and tested. This circuit implements a feedback regulation [57] that has been studied in a particular microfluidic set-up, controlling its outcome as a function of galactose and glucose concentration.
Several other works in Synthetic Biology have focused on the implementation of negative feedback loops. For example, a genetic network that implements the dynamic control of the production of a target protein has been realized, using a regulatory system which is capable of sensing the formation of inclusion bodies, insoluble aggregates of recombinant proteins that need downstream processing to become functional again [58] and produce stress on cells. The regulation introduced by the authors can face all of those problems that affect heterologous protein over-expression; in fact a lot of variables are involved in this process (e.g., host strain, medium composition, process conditions) and could impact on outcome and also the cells state. An IPTG inducible device drives the production of the protein of interest while the accumulation of the protein is able to induce a stress sensitive promoter, P_{IbpAB}, that is placed upstream of a repressor coding sequence, able to down-regulate the transcription of the target product (Fig. 1.14). The sensing system, composed by the P_{IbpAB} promoter and its RBS was engineered in order to span dynamic range of target protein expression. Following a model-based fine tuning approach, the behaviour of the stress-inducible promoter was first investigated using RFP protein as output and focusing on the different dynamic range obtained varying the stress sensitive promoter-RBS pair driving RFP expression. After this, the whole genetic network was deeply examined. In order to have an easy to measure quantity, the target protein of this work was the green fluorescent

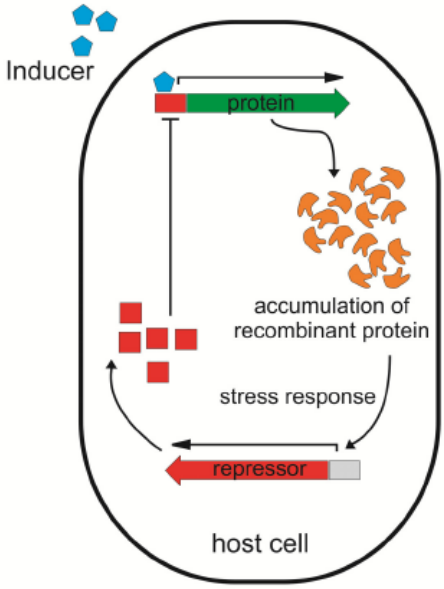


Figure 1.14: **Negative feedback for protein production.** Schematization of a negative feedback control for the production of a recombinant protein [58]. The production of a recombinant protein, using a regulated promoter, increases the stress of the cell; a stress sensitive promoter is used to drive the transcription of the repressor protein that down-regulates the transcriptional activity of the promoter placed upstream of the coding sequence of the recombinant protein.

protein (GFP), while the input promoter was engineered in order to be repressed by the TetR repressor. It was observed that the GFP yield without the negative feedback was higher, even if in this case a great percentage (over 40 %) was insoluble due to the formation of inclusion bodies. In the case of negative feedback regulation the total amount of GFP per cell was clearly lower because of the regulatory mechanism, but in general the soluble fraction was higher. A system of delayed differential equations was used to capture the behaviour of the genetic circuit: it takes into account mathematical description of the input device, the recombinant protein production and the stress-induced expression of the repressor. Its parameters were partially estimated from experimental data, whereas some of them were retrieved from previous works: the predictions were quite in agreement with experimental data, even though some phenomena such as the depletion of the cell resources for the production of TetR repressor were not modelled. However, despite some discrepancies, this work demonstrated the possibility to increase the yield of a common pathway used in industrial production of recombinant proteins, through a negative feedback loop.

Similarly, in order to obtain a better yield in fatty acids ethyl ester (FAEE) for biodiesel production from sugar, a dynamic sensor-regulator system (DSRS) was proposed in [15]. This mechanism allows for optimal expression of the genes involved in this particular metabolic pathway, preventing from the accumulation of toxic by-products that can slow down cultures growth or determining adaptive responses that lead to suboptimal yield and productivity. Considering fatty acids as the intermediate to control, the biosensors used in this study exploit the fatty acid sensing protein FadR (a regulator protein which regulates the fatty-acid beta oxidation, whose binding to the target sequence is inhibited by long chain fatty acyl-CoA compounds) and its cognate regulator. The sensor module was engineered focusing on the promoter: the P_{fadBA} DNA binding sequence of natural *E. coli* promoter, which is a target for FadR, was used to engineer the P_L promoter from λ-phage and P_R promoter from phage T7. When fatty acids are absent, the output of these promoters is expected to be low as the FadR regulator binds to the FadR recognition sites of the promoter sequence; otherwise, when fatty acids are present they are activated to acyl-CoA by acyl-CoA synthase (encoded by *fadD*) and FadR binds to them, releasing the promoter. In this way, the authors managed to obtain a 60-fold fluorescence change as a function of oleic acid, much higher than that of native promoters. As reported in Fig. 1.15 the fatty-acid sensitive promoter was placed upstream of the ethanol pathway genes, *pdc* and *adhB*, and upstream of the *fadD* and *atfA*, respectively responsible for the production of acyl-CoA synthesis from fatty acids and condensation of intermediates

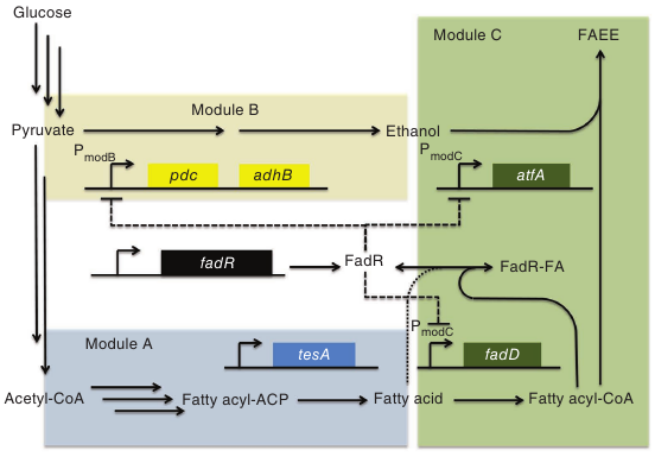


Figure 1.15: **FAEE pathway.** Representation of the complex pathway for the production of FAEE in *E. coli* [15]; the authors introduced a negative feedback based on the sensing of the fatty acids concentration through the mechanism of the FadR regulator protein. Module A is fundamental for the production of fatty acids; module B contains the *pdc* and *adhB* genes for the conversion of sugars into ethanol; the products of modules A and B are processed by the protein expressed by module C, an acyl-CoA synthase FadD, and a wax-ester synthase, AtfA, to form FAEE.

to the final product, the FAEE. Ideally, this way, the production of FAEE can be tuned dynamically by intermediates concentration, avoiding their accumulation or their utilization to the detriment of cells basic functions, such as membrane biosynthesis. Different DSRSs were implemented, obtaining good results as the yield of the process reached 28 % of the theoretical maximum, in contrast to the 9.4 % of the state of the art strain, which does not implement dynamic feedback control [15]. These results were confirmed also comparing the productivity of the DSRS strains to that of cultures implementing static control. Moreover, a mathematical model of 26 rate constants and 19 molecular species has been used to simulate the system and to compare the DSRS and the static control; in accordance with the experimental data, it was found out that the DSRS topology can improve the yield across a broad range of promoter strengths.

A system that controls the growth of cell cultures through a negative feedback regulation has been implemented, again exploiting the *quorum sensing* mechanism of *V. fischeri* [59].

The genetic circuit is represented in Fig. 1.16: the authors placed the *luxR* and *luxI* genes under the transcriptional control of the P_{Lac} promoter, while the *ccdB* gene, encoding a toxic protein, fused with *lacZ*α, was placed downstream

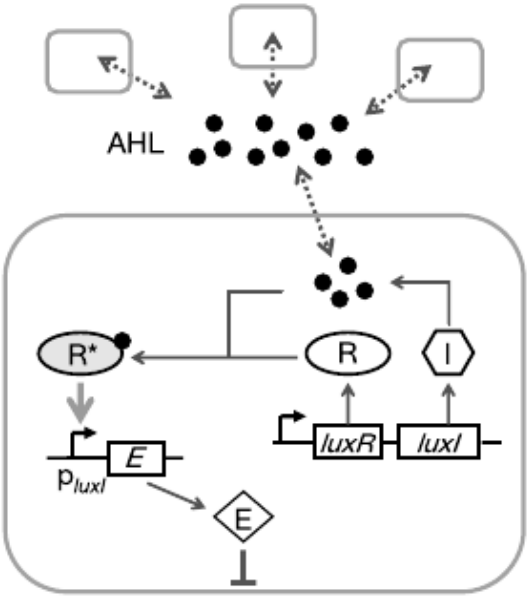


Figure 1.16: **Controlling population dynamics.** The system that controls population growth through cell-cell communication [59]; cells induced with IPTG produce the AHL, whose concentration is proportional to the number of cells. As the culture grows, the AHL reaches a critical concentration and the ccdB toxic protein is expressed and causes cells death. R represents the LuxR transcription activator, I the LuxI AHL synthetase, E the ccdB toxic protein.

of the P_{Lux} promoter. A mathematical model (Eq. 1.10) that captures the main features of the system has been proposed.

$$\begin{aligned} \frac{d}{dt}N &= k \cdot N \cdot (1 - \frac{N}{N_m}) - d \cdot E \cdot N \\ \frac{d}{dt}E &= k_E \cdot A - d_E \cdot E \\ \frac{d}{dt}A &= v_A \cdot N - d_A \cdot A \end{aligned} \tag{1.10}$$

N is the number of cells, E the per cell concentration of the killing protein and A the AHL concentration which is supposed to be the same inside and outside the cells. d is the killing rate of cells due to the intracellular concentration of E, k_E is the production rate of E, d_E its degradation; finally v_A is the rate constant for AHL production, proportional to N, while d_A is the AHL degradation. Parameters were identified from experimental data. At low cell densities, the AHL concentration is not sufficient to activate the killing module. When cells grow, AHL reaches a sufficient concentration that is capable of triggering the production of the toxic protein. The pH-dependent stability of AHL was exploited to reach the desired steady state of cell density: by using pH as a control signal, AHL reaches different concentration levels (higher for lower pH values) and thus cells can reach different stationary phase levels. However, the genetic tuning of the circuit was not tested, even if mathematical models may aid such process.

Taken together, the provided examples have shown that complex engineering-inspired circuits can be successfully realized and mathematical modelling is a powerful tool to aid their design. The central issue in the engineering of biological circuit is the bottom-up design: all the cited examples relied on trial-and-error design approaches or model-guided semiquantitative design steps, but a rigorous bottom-up process has not been validated yet on a large scale. Only a few examples of successful prediction of interconnected networks, composed of characterized modules, have been reported. Since negative feedback control is widely used in the engineering world and, similarly, as described in the last examples, it could bring to complex adaptive circuits in Synthetic Biology, this thesis focuses on the bottom-up design of a synthetic close-loop controller in *E. coli*.
The aim of this network is to regulate the concentration of a signalling molecule (3-oxohexanoyl homoserine lactone - $3OC_6$-HSL, also called HSL or AHL) in liquid bacterial cultures. In order to follow a pure bottom-up approach, a mathematical model has been derived for the full circuit and its parameters have been identified from *ad-hoc* acquired experimental data on subsystems (smaller functional units of the whole network). Finally, the goodness of the model is based on the comparison between predictions and experimental data obtained from the full system. The thesis describes the whole design process

of this circuit, from parts selection to the final circuit testing.

Chapter 2

Materials and methods

In this chapter the laboratory materials, instruments and experimental set-up used to study the genetic controller will be presented. Then, the design process adopted to engineer the circuit will be described. Finally, the mathematical modelling techniques will be illustrated, focusing on the characterization of the circuit subparts and subsystems and on the prediction of the full system behaviour.

2.1 Materials

2.1.1 Media

LB broth was used to grow cultures propagating the plasmids of interest; it was prepared according to the protocol reported in [60]:

- 10 g/L Bacto-tryptone;
- 5 g/L Yeast Extract;
- 10 g/L NaCl.

When preparing LB agar plates 15 g/L agar was added.
M9 medium [60] was used to grow bacteria in all the quantitative experiments because of its low autofluorescence; it was prepared following this protocol:

- 739 mL/L of autoclaved deionized water;
- 100 μL/L $CaCl_2$ 1 M;
- 200 mL/L of 5 X M9 salts (M6030, Sigma Aldrich);

- 20 mL/L casamino acids 10%;
- 34 mL/L thiamine hydrochloride 10 mg/mL;
- 5 mL/L glycerol 80%;
- 2 mL/L $MgSO_4$ 1 M.

In promoters, enzymes and final system tests, M9 was pH-adjusted to 6 by adding a proper volume of hydrochloric acid (Sigma Aldrich) to grant a higher $3OC_6$-HSL stability (see App. A); in $3OC_6$-HSL quantification experiments, the pH was not adjusted and it was near 7.

2.1.2 Inducers

In this work three inducers were used; these chemical compounds regulate the transcriptional activity of the promoters adopted in the realization of the regulatory network presented and its subsystems.

- **Anhydrotetracycline (aTc)** (631310, Clontech) is a tetracycline analogue, able to bind the TetR repressor and up-regulate the transcriptional activity of the P_{TetR} promoter, which is not repressed anymore by TetR. Stock solution is 2 mg/mL concentrated and is stored at -20°C.
- **Isopropyl-β-D-1-thiogalactopyranoside (IPTG)** (I1284, Sigma Aldrich) is a non-metabolizable galactose analogue which is able to bind the LacI repressor and triggers the activity of the P_{Lac} promoter, which is not repressed anymore by LacI. IPTG is given at 200 mM concentration and stored at -20°C.
- **3-oxohexanoyl homoserine lactone ($3OC_6$-HSL)** (K3007, Sigma Aldrich) is a chemical compound found in some marine bacteria such as *V. fischeri*. This molecule was exploited in the experiments regarding the activation of the P_{Lux} promoter and the $3OC_6$-HSL concentration measurement. It is purchased as a powder stock, which is dissolved in deionized water to prepare a 2 mM solution and stored at -20°C.

2.2 Cloning

Each part used in this work is compliant to the BioBrick™ Standard [23, 24]. Consequently, every junction found between parts after assembly, the so called scar, has the TACTAG sequence if the downstream part is a CDS or, otherwise, the TACTAGAG sequence. TOP10 *E. coli* competent cells

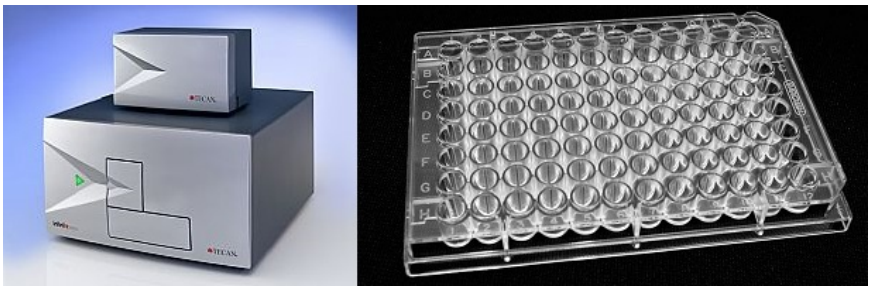


Figure 2.1: **Tecan Infinite F200.** Photograph of the Tecan Infinite F200 used in this work to dynamically measure optical density and fluorescence in liquid cell cultures. It is composed by a central unit, the bigger one, and an injection unit; the former contains the filters and the circuitry that performs measurements and is designed to accomodate the 96-well microplates. The latter is composed by the circuitry and the fluidics necessary to automatically dispense liquid in the plates. On the right, a typical 96-well microplate is also shown.

(Invitrogen) were used to propagate the intermediate parts while final parts were transformed in home-made competent MG1655-Z1 cells [60]. Machery-Nagel kits were used for DNA purification. Cells were selected with the proper antibiotic: if transformed with pSB4C5 plasmid, then selection was made in chloramphenicol at 12.5 μg/mL, while 20 μg/mL of kanamycin were used to select cells transformed with pSB3K3 plasmid; if both plasmids were present, the selection was done in 12.5 μg/mL chloramphenicol together with 20 μg/mL kanamycin. Once transformed, they were long-term stocked in 20% glycerol at -80°C; glycerol stocks were prepared with 250 μL 80% glycerol and 750 μL cells grown in LB with the proper antibiotic. All plasmids were assembled using basic or composite parts from the Registry [22] 2009, 2010 or 2011 Distribution (Tab. 2.1).

2.3 Instruments

2.3.1 Tecan Infinite F200

The Tecan Infinite F200 (Fig. 2.1) is a microplate reader that is used in several applications, thanks to its flexibility and modularity. Different kinds of microplates can be used for this instrument, ranging from 6 to 384 wells. The machine is able to measure absorbance, fluorescence and luminescence through a system including lamp, detectors and filter-containing slides. A mechanical system that dispenses liquid in the wells is also available.

The i-control™ software is used to program the Infinite F200. Users can program kinetic cycles to measure the quantities of interest at a specific sampling

Table 2.1: **List of parts and plasmids used to design the genetic devices.** BioBrick™ codes are reported according to the Registry of Standard Biological Parts. RBS strength is reported as described in the individual Registry pages of RBSs (measured via a specific reporter system), even though their efficiency may significantly change in different applications.

BioBrick™	**Description**
BBa_J23101	constitutive promoter used as a standard reference
BBa_R0040	TetR-repressible promoter
BBa_R0062	LuxR-3OC$_6$-HSL inducible promoter
BBa_B0030	medium-strong RBS
BBa_B0031	weak RBS
BBa_B0032	medium-weak RBS
BBa_B0034	strong RBS
BBa_C0060	AiiA lactonase from *Bacillus* sp. 240B1 [61]
BBa_C0061	*luxI* CDS, 3OC$_6$-HSL synthetase
BBa_E1010	RFP coding sequence
BBa_B0015	double terminator
pSB1A2	high copy number plasmid (pUC19-derived pMB1 replication origin) [24]
pSB3K3	medium copy number plasmid (pMR101-derived p15A replication origin) [62]
pSB4C5	low copy number plasmid (pSC101 replication origin) [24]

time. It is also possible to set the inner temperature of the instrument and shaking movement of the microplate. In this way, several hour-long incubation experimental tests can be performed, monitoring the quantities of interest. In this work the instrument was used to measure mainly two quantities in liquid cultures in 96-well microplates:

- the optical density at 600 nm (OD_{600}) which is a quantity proportional to the per-well cell number. It is defined as:

$$OD_{600} = -\log_{10} \frac{I}{I_0}$$

  where I_0 is the intensity of the incident radiation while I is the intensity of the transmitted radiation;

- the red fluorescence of cultures (excitation wavelength: 535 nm; emission wavelength: 620 nm).

In every experimental measurement performed with the Infinite F200 reader experimental controls are used to normalize the two quantities of interest: in particular, the optical density of sterile both is used to normalize the growth of cultures while the fluorescence of a culture that does not produce the fluorescent reporter of interest is used as blank for the fluorescence signal.

2.3.2 LAMBDA MINIFOR fermenter-bioreactor

The LAMBDA MINIFOR is a laboratory-scale fermenter-bioreactor that can be used to carry out batch, fed-batch or continuous fermentation experiments in a temperature/pH/oxygen-controlled environment (Fig. 2.2). In particular, the continuous operation mode is used to maintain a liquid culture at a fixed density and, in this case, the set-up is also called chemostat (Sec. 2.8). It is composed by:

- a 0.4 L glass vessel which can be autoclaved-sterilized; it has a set of glass necks through which probes can be immersed in the contained liquid and detect some parameters that need to be controlled in a typical fermentation process (i.e., pH, temperature and pO_2). These glass necks can also be used to dispense acids or bases when controlling the pH or to sample the culture contained in the vessel, through a sampling port, without the risk of contaminating the inner environment;

- a screw cap that tightly closes the vessel and insulates it from the external environment by a silicone membrane. A stirring axis is linked to it so that it is possible to assemble mixing discs on it;

Figure 2.2: **The LAMBDA MINIFOR fermenter-bioreactor.** Photograph of the fermenter-bioreactor used in this work to maintain liquid cultures at a constant OD_{600}.

- a motor unit that moves the stirring axis in order to mix liquid cultures;
- a pO_2 probe to monitor the concentration of dissolved oxygen in liquid cultures;
- a pH-temperature probe which detects both the pH and temperature of the liquid culture. In order to have reliable measurements the pH probe is calibrated with gauging solutions;
- peristaltic pumps used to inject solutions, e.g., to adjust the liquid cultures pH, or to extract the content of the vessel. Peristaltic pumps can be either programmed by the user with a fixed flow rate (e.g., when running experiments in chemostatic conditions) or they can be controlled by the central operating unit which elaborates the parameters of interest and calculates the rate of the pumps (e.g., for base addition in pH-controlled experiments);
- a central operating unit that allows the user to set parameters such as temperature, pH and stirring frequency. It also contains the heating module and is responsible for the regulation of the above mentioned parameters as it implements control algorithms that automatically set the pump rates (if required) and the heating power.

2.4 Circuit design

The circuit design will be herein described; the choice of promoters, genes and strain used to implement the negative feedback controller will be discussed. The overview of the designed system is shown in Fig. 2.3 and will be extensively discussed in the following sections.

2.4.1 Parts selection

The genetic parts used in this study come from the MIT Registry of Standard Biological Parts. The system, depicted in Fig. 2.3, controls the concentration of the 3-oxohexanoyl homoserine lactone ($3OC_6$-HSL), a small signalling molecule that freely diffuses through the *E. coli* cell membranes. The control is actuated via *luxI* and *aiiA* genes. The former encodes for an autoinducer synthetase from marine bacterium *V. fischeri*, which synthesizes the autoinducer molecule $3OC_6$-HSL, while the latter encodes for a lactonase enzyme from the *Bacillus* sp.240B1 that is able to degrade $3OC_6$-HSL intracellularly [61]. Two regulated promoters are used to modulate the transcription of these two genes: the P_{TetR} promoter, inducible by anhydrotetracycline (aTc), drives

the *luxI* transcription and the P_{Lux} promoter, whose transcriptional output is inducible by $3OC_6$-HSL, modulates *aiiA* transcription. The P_{Lux} promoter requires the LuxR transcriptional activator to be turned on by $3OC_6$-HSL: LuxR is able to form a complex with $3OC_6$-HSL (two molecules of LuxR bind two molecules of $3OC_6$-HSL, giving a heterotetramer) and finally this complex binds the lux box DNA sequence, contained in P_{Lux}, to up-regulate its transcriptional activity. The circuit contains a constitutive expression cassette for the *luxR* gene, driven by the P_{Lac} promoter. On the other hand, the P_{TetR} promoter requires the TetR repressor to be an effective inducible system: TetR (a homodimer) represses P_{TetR}, but the aTc inducer can bind TetR, thus liberating the promoter and triggering its activity. For this reason, the circuit also requires a constitutive expression of TetR to be actually inducible by aTc. Although the *tetR* gene is not explicitly present in the circuit, it is contained in the strain genome, as it will be described in Sec. 2.4.2. The intracellular concentrations of LuxI and AiiA are function of the upstream promoters strength, promoter induction level and RBS strength. For this reason, RBSs must be chosen to ensure that the enzymes are properly translated from mRNA into proteins.

Taken together, the described elements implement a negative feedback: $3OC_6$-HSL is produced by LuxI; when $3OC_6$-HSL reaches a critical concentration, it triggers AiiA production which degrades it, thus self-regulating its concentration. Fig. 2.3 illustrates the whole designed circuit and the interactions among its elements. Each part has a code, which corresponds to the BioBrick™ code in the Registry of Standard Biological Parts. Since the RBSs of *luxI* and *aiiA* require a proper tuning, their code is not included and it will be selected from a list of candidates, compatible with the correct circuit behaviour. The copy number of the described elements also needs to be tuned for the same reasons, by selecting suitable plasmid vectors according to their replication origins, which determine specific DNA copies per cell. The basic subsystems of the circuit presented in Fig. 2.3 have been assembled or extracted from the Registry collection to construct the final circuit and the required subsystems to characterize promoters and enzymes. These subsystems are:

- P_{TetR} promoter driving RFP transcription in pSB4C5 plasmid vector;
- P_{TetR} promoter driving RFP transcription in pSB3K3 plasmid vector;
- P_{Lux} promoter driving RFP transcription in pSB3K3 plasmid vector;
- P_{TetR} promoter driving *luxI* transcription in pSB4C5 plasmid vector;
- P_{TetR} promoter driving *aiiA* transcription in pSB3K3 plasmid vector;

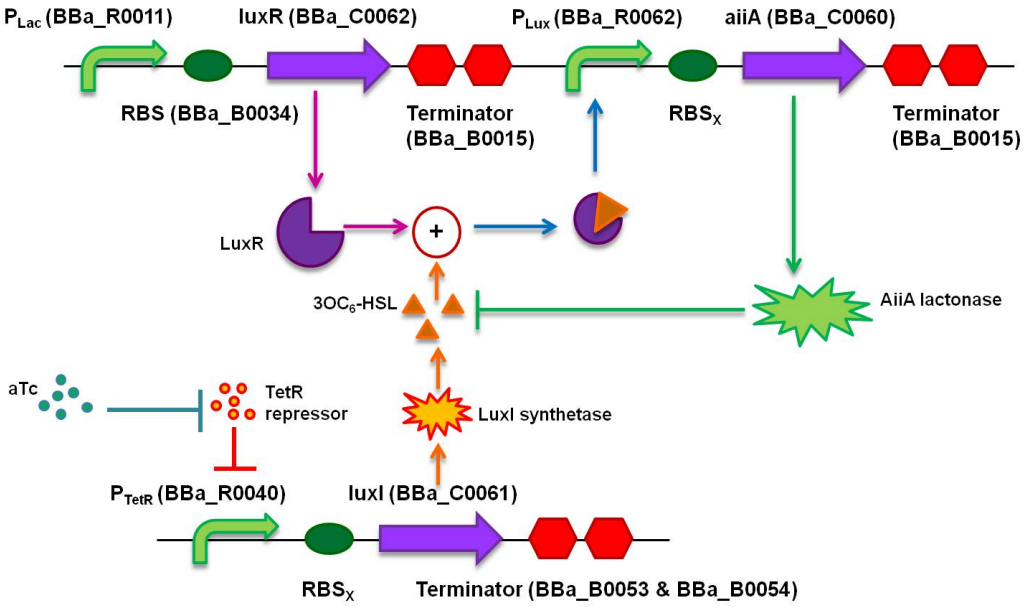


Figure 2.3: **Schematic representation of the negative feedback for the control of $3OC_6$-HSL in *E. coli*.** Design of the synthetic genetic controller, realized in *E. coli* MG1655-Z1. The TetR repressor is constitutively produced by the *tetR* gene, placed in the MG1655-Z1 genome. The P_{TetR} promoter, inducible by aTc, is placed upstream of the *luxI* coding sequence, tuning its transcription rate. Once translated, the LuxI enzyme produces the $3OC_6$-HSL which, together with the LuxR transcription factor constitutively expressed by the P_{Lac} promoter, up-regulates the transcriptional output of the P_{Lux} promoter. This, in turn, is placed upstream of the *aiiA* coding sequence, which encodes for a lactonase, an enzyme able to degrade $3OC_6$-HSL. This mechanism, properly tuned, is expected to regulate the concentration of $3OC_6$-HSL in *E. coli* cultures.

- BBa_ J23101 reference promoter with RFP in pSB3K3 plasmid vector.

2.4.2 Strain selection

The MG1655 strain of *E. coli* [63] has a fully-sequenced genome, available in biological databanks, and it is widely used in many research studies. A particular engineered strain, called MG1655-Z1 and derived from MG1655, was used in this work [64, 65]. It has a Z1 cassette integrated in its genome in the λ-att locus. The Z1 cassette contains a constitutive over-expression cassette for TetR and another one for LacI, which down-regulate the P_{TetR} and P_{Lac} transcriptional activities, respectively. In this way it is possible to tune the P_{TetR} and P_{Lac} promoters by adding a proper concentration of aTc and IPTG, respectively. The designed close-loop controller circuit has a P_{TetR} promoter which can be effectively tuned to produce LuxI. The synthetic circuit also contains a LuxR expression cassette driven by P_{Lac}, which is inhibited by LacI in the MG1655-Z1 strain. In this case, the basic activity of P_{Lac} (i.e., its activity in the repressed state) is exploited to produce LuxR at low, but sufficient levels to activate P_{Lux} in presence of $3OC_6$-HSL.

2.5 Data processing for ${S_{cell}}^{SS}$ computation

OD_{600} and RFP time series acquired with the Infinite F200 reader need to be processed before evaluating the measures of interest, such as S_{cell}. The commonly used protocol, programmed via the i-control™ software (Tecan), is a kinetic cycle with this features: linear shaking (3-mm amplitude, 15 s), wait (5 s), OD_{600} measurement, fluorescence measurement (excitation at 535 nm, emission at 620 nm, gain=50, to detect RFP signal), repeat cycle every 5 min. For each experiment, control wells were used to normalize the quantities of interest: sterile broth was used as a blank for absorbance, while a non-fluorescent culture was used to normalize the measured fluorescence. Raw absorbance and fluorescence data were always subtracted the respective blank. After data pre-processing, in order to compute the synthesis rate per cell (S_{cell}), the numerical derivative of fluorescence was divided by the respective OD_{600} time series. The signal was averaged during the exponential phase of cultures growth, thus obtaining the synthesis rate per cell at the steady state, namely ${S_{cell}}^{SS}$. It was expressed as red fluorescent arbitrary units per cell per minute (A.U.$_r$ cell^{-1} min^{-1}).

2.6 Mathematical modelling and parameter identification

A mathematical model of the whole regulatory network herein presented has been derived; it is based on Ordinary Differential Equations (ODEs) and takes into account the dynamics of inducible promoters, enzymes and the signalling molecule $3OC_6$-HSL. According to the bottom-up approach adopted in this work, the model was used to predict the behaviour of the whole designed system, with parameter values identified from subsystems. This situation is analogous to the retrievement of static/dynamic characteristics of the components of interest from a data-book and the prediction of the interconnected system behaviour from the knowledge of individual components quantitative behaviour. Here, *ad-hoc* subsystems were designed, characterized and used to identify all the full model parameters. Parameters were identified using the *lsqnonlin* routine in MATLAB R2011b (MathWorks, Natick, MA).
The mathematical model of the whole genetic regulatory network will be introduced, giving explanation of the parameters; then subsystems and their relative models will be presented explaining the protocols and measurement strategies adopted.

2.6.1 Model of the full synthetic close-loop regulator

An ODE model of the complete system presented in Fig. 2.3 has been derived. It is useful to underline the hypotheses that lie behind the mathematical formalization of the behaviour of the genetic circuit:

- as LuxR is constitutively produced under the control of the P_{Lac} promoter, the model only takes into account the $3OC_6$-HSL concentration to describe the transcription activation of the P_{Lux} promoter;
- all the molecules are expressed as intracellular concentrations except $3OC_6$-HSL, which can freely diffuse through cell membranes and, for this reason, the inner and the outer concentrations are assumed to be the same;
- protein production, triggered by a given chemical inducer concentration, is modelled as described in Sec. 1.3 with the exception that transcription is not explicitly modelled, while translation is described by a differential equation with protein synthesis rate per cell (P); P is also a state variable assumed to have a first-order dynamics, with time constant $\frac{1}{r_{p_X}}$. It empirically summarizes the processes from transcription activation to the synthesis of a functional protein. $\frac{dP}{dt}$ is expressed as a synthesis rate per cell per min, A.U. cell^{-1} min^{-1};

- OD_{600} is proportional to the cell concentration in liquid cultures and is used to describe the cellular growth;
- growth of bacterial cultures was modelled as a logistic function described by two parameters: the growth rate and the saturation value;
- the relation between the promoter output and inducer concentration was modelled as a Hill function, described by a maximum output, a basal activity, a dissociation constant and a Hill coefficient;
- the enzymes activities were modelled as Hill functions of the intracellular enzymes concentration. They are described by three parameters: a maximum activity, a half-saturation constant and a Hill coefficient. In this case no basal activity was introduced when no enzyme is present;
- first-order degradation for enzymes, due to a degradation tag in the protein sequence, is taken into account as well as the linear spontaneous degradation of $3OC_6$-HSL.

The controller is described by the following equation system (Eq. 2.1) when it is considered in batch cultures:

$$
\begin{aligned}
&\tfrac{d}{dt}OD_{600} = \mu \cdot OD_{600} \cdot \left(1 - \tfrac{OD_{600}}{OD_{600,max}}\right) \\
&\tfrac{d}{dt}PTetR = -r_{P_{TetR}} \cdot PTetR + r_{P_{TetR}} \cdot \alpha_{P_{TetR}} \cdot \left(\delta_{P_{TetR}} + \frac{1-\delta_{P_{TetR}}}{1+\left(\frac{k_{P_{TetR}}}{[aTc]}\right)^{\eta_{P_{TetR}}}}\right) \\
&\tfrac{d}{dt}PLux = -r_{P_{Lux}} \cdot PLux + r_{P_{Lux}} \cdot \alpha_{P_{Lux}} \cdot \left(\delta_{P_{Lux}} + \frac{1-\delta_{P_{Lux}}}{1+\left(\frac{k_{P_{Lux}}}{[HSL]}\right)^{\eta_{P_{Lux}}}}\right) \\
&\tfrac{d}{dt}[LuxI] = PTetR - (\mu + \gamma_{LVA}) \cdot [LuxI] \\
&\tfrac{d}{dt}[AiiA] = PLux - (\mu + \gamma_{LVA}) \cdot [AiiA] \\
&\frac{d}{dt}[HSL] = OD_{600} \cdot \left(\frac{k_{LuxI,max}}{1+\left(\frac{k_{M,LuxI}}{[LuxI]}\right)^{\eta_{LuxI}}}\right) + \\
&\qquad - \left(OD_{600} \cdot \frac{k_{AiiA,max}}{1+\left(\frac{k_{M,AiiA}}{[AiiA]}\right)^{\eta_{AiiA}}} + \gamma_{HSL}\right) \cdot [HSL]
\end{aligned}
\tag{2.1}
$$

The species modelled are the cell growth, OD_{600}, the promoters synthesis rate, $PTetR$ and $PLux$, the enzymes per cell concentrations, $[LuxI]$ and $[AiiA]$, and the $3OC_6$-HSL concentration, $[HSL]$. In Eq. 2.1 $OD_{600,max}$ is a parameter that represents the saturation value for cultures growth, while cell

growth rate is μ. Promoters Hill functions are described by α_{p_X}, δ_{p_X}, k_{p_X} and η_{p_X}, while r_{p_X} describes the first-order dynamics of protein synthesis, following promoter induction. $k_{X,max}$, $k_{M,X}$ and η_X are the Hill parameters for enzymes activity. Their intracellular level is modulated by the synthesis rates indirectly triggered by promoter induction and by the degradation due to LVA tag, represented by γ_{LVA} [1], which was always set to the one estimated in [49]. Moreover, enzymes are diluted because of cell division at the same rate of cells growth, μ. Finally, the spontaneous degradation of $3OC_6$-HSL was described by γ_{HSL}. Parameter description and measurement units are summarized in Tab. 2.2.

When considering the continuous culture mode (i.e., the chemostat, described in Sec. 2.8), equations are slightly different as $3OC_6$-HSL is washed away at the same rate of bacterial growth. The relative equation system is shown in Eq. 2.2:

$$
\begin{aligned}
&\tfrac{d}{dt}OD_{600} = 0 \\
&\tfrac{d}{dt}PTetR = -r_{P_{TetR}} \cdot PTetR + r_{P_{TetR}} \cdot \alpha_{P_{TetR}} \cdot \left(\delta_{P_{TetR}} + \frac{1-\delta_{P_{TetR}}}{1+\left(\frac{k_{P_{TetR}}}{[aTc]}\right)^{\eta_{P_{TetR}}}} \right) \\
&\tfrac{d}{dt}PLux = -r_{P_{Lux}} \cdot PLux + r_{P_{Lux}} \cdot \alpha_{P_{Lux}} \cdot \left(\delta_{P_{Lux}} + \frac{1-\delta_{P_{Lux}}}{1+\left(\frac{k_{P_{Lux}}}{[HSL]}\right)^{\eta_{P_{Lux}}}} \right) \\
&\tfrac{d}{dt}[LuxI] = PTetR - (\mu + \gamma_{LVA}) \cdot [LuxI] \\
&\tfrac{d}{dt}[AiiA] = PLux - (\mu + \gamma_{LVA}) \cdot [AiiA] \\
&\frac{d}{dt}[HSL] = OD_{600} \cdot \left(\frac{k_{LuxI,max}}{1+\left(\frac{k_{M,LuxI}}{[LuxI]}\right)^{\eta_{LuxI}}} \right) + \\
&\qquad - \left(OD_{600} \cdot \frac{k_{AiiA,max}}{1+\left(\frac{k_{M,AiiA}}{[AiiA]}\right)^{\eta_{AiiA}}} + \gamma_{HSL} + \mu \right) \cdot [HSL]
\end{aligned}
\tag{2.2}
$$

In this case, the OD_{600} was maintained at the constant level $OD_{600}(t_0)$, while the $3OC_6$-HSL decay accounts also for the dilution rate μ, which represents $3OC_6$-HSL wash due to culture dilution. The mathematical model was used first of all to simulate the behaviour of the system and to verify the possibility to obtain a controller varying the values of the parameters in a reasonable range. Once the parameters from subsystems were estimated, the identified model was used to check whether simulations match the actual behaviour of

[1] LVA tag is a sequence of amino acids recognized by specific intracellular proteases allowing a faster protein degradation

Table 2.2: **Parameters of the model of the genetic controller.** List of the model parameters with their respective description and measurement units.

Parameters		
Name	**Description**	**Measurement Unit**
μ	Growth rate of liquid culture	min^{-1}
$OD_{600,max}$	Saturation value for liquid cultures optical density	cell
$r_{P_{TetR}}$	Activation rate of P_{TetR} promoter	min^{-1}
$\alpha_{P_{TetR}}$	Maximum output of P_{TetR} promoter	A.U. cell^{-1} min^{-1}
$\delta_{P_{TetR}}$	Percentage leakage of P_{TetR} promoter	-
$k_{P_{TetR}}$	Half saturation constant of P_{TetR} promoter	ng/mL
$\eta_{P_{TetR}}$	Cooperativity of P_{TetR} promoter	-
$r_{P_{Lux}}$	Activation rate of P_{Lux} promoter	min^{-1}
$\alpha_{P_{Lux}}$	Maximum output of P_{Lux} promoter	A.U. cell^{-1} min^{-1}
$\delta_{P_{Lux}}$	Percentage leakage of P_{Lux} promoter	-
$k_{P_{Lux}}$	Half saturation constant of P_{Lux} promoter	nM
$\eta_{P_{Lux}}$	Cooperativity of P_{Lux} promoter	-
γ_{LVA}	Degradation rate due to LVA tag	min^{-1}
γ_{HSL}	Spontaneous $3OC_6$-HSL degradation rate	min^{-1}
$k_{LuxI,max}$	Maximum LuxI a expressed asctivity	nM cell^{-1} min^{-1}
$k_{M,LuxI}$	Half saturation constant of LuxI activity	A.U. cell^{-1}
η_{LuxI}	Cooperativity for LuxI enzyme activity	-
$k_{AiiA,max}$	Maximum AiiA activity	cell^{-1} min^{-1}
$k_{M,AiiA}$	Half saturation constant of AiiA activity	A.U. cell^{-1}
η_{AiiA}	Cooperativity for AiiA enzyme activity	-

the genetic controller in liquid *E. coli* cultures. Simulations were carried out with MATLAB R2011b (MathWorks, Natick, MA) using the *ode15s* routine. As uncertainty affects the parameter estimates of each of the above mentioned subsystems, simulations of the behaviour of the circuit were carried out using the Monte Carlo method in order to propagate the uncertainty of the parameters to the output of the device. The Monte Carlo method used to perform this task is described in Sec. 2.9.

2.6.2 Promoters characterization

Promoters are responsible for the transcription of genes and they can be constitutive or regulated (Sec. 1.2.1). In this work two inducible promoters were used: the P_{TetR} promoter, inducible in *E. coli* MG1655-Z1 by aTc, as previously described, and P_{Lux} promoter, whose transcriptional activity is

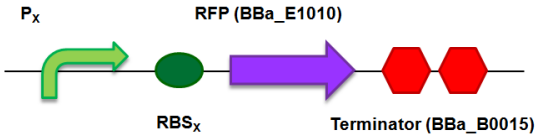


Figure 2.4: **Subsystem for promoters parameters estimation.** Schematic representation of the genetic circuit used to estimate the promoters parameters: the promoter of interest was assembled upstream of the coding sequence of the RFP reporter.

modulated by $3OC_6$-HSL concentration.

Promoters were ligated upstream of the RFP reporter in order to obtain an observable output (Fig. 2.4). While the measurable output was an ${S_{cell}}^{SS}$, thus measured in A.U.$_r$ cell^{-1} min^{-1}, the corresponding promoter activity is expressed as A.U cell^{-1} min^{-1}. The model in Eq. 2.3 was used to capture a generic regulated promoter behaviour.

$$\begin{aligned} &\frac{d}{dt}P = -r_P\left(P - \alpha_P \cdot \left(\delta_P + \frac{1-\delta_P}{1+\left(\frac{k_P}{[inducer]}\right)^{\eta_P}}\right)\right) \\ &\frac{d}{dt}I = P - (\mu + a)\cdot I \\ &S_{cell} = a \cdot I \end{aligned} \tag{2.3}$$

In this model (Eq. 2.3), similarly to the equations described in Sec. 1.3, I is the immature (non-fluorescent) protein concentration, which matures into a fluorescent form with maturation rate a.

Steady-state characterization

Considering the model in Eq. 2.3 at the steady state, it is possible to obtain the mathematical relations used to identify the static parameters relative to the promoters. In fact:

$$\begin{aligned} P^{SS} &= \alpha_P \cdot \left(\delta_P + \frac{1-\delta_P}{1+\left(\frac{k_P}{[inducer]}\right)^{\eta_P}}\right) \\ I^{SS} &= \frac{\alpha_P \cdot \left(\delta_P + \frac{1-\delta_P}{1+\left(\frac{k_P}{[inducer]}\right)^{\eta_P}}\right)}{(\mu+a)} \\ S_{cell}^{SS} &= \frac{a}{(\mu+a)} \cdot \alpha_P \cdot \left(\delta_P + \frac{1-\delta_P}{1+\left(\frac{k_P}{[inducer]}\right)^{\eta_P}}\right) \end{aligned} \tag{2.4}$$

During the experimental tests regarding promoters static transfer function, the measured quantity is proportional to the RFP synthesis rate per cell, ${S_{cell}}^{SS}$, (given an inducer concentration), which is computed as described in

Sec. 2.5. In order to obtain reliable measures across different experiments, this quantity is divided by the ${S_{cell}}^{SS}$ of a reference culture, in this study the BBa_J23101 promoter expressing RFP (with the BBa_B0034 RBS) in pSB3K3 plasmid in MG1655-Z1.

The protocol used for the identification of the above mentioned parameter is the following: long-term glycerol stocks were streaked on LB agar plates supplemented with the proper antibiotic in order to isolate single colonies, considered as biological replicates. Plates were grown overnight at 37°C, then three single colonies for each construct were inoculated overnight in 0.5 mL selective M9 pH 6 in 2-mL tubes, at 37°C, 220 rpm. Then, cultures were 1:100 diluted in 1 mL fresh selective broth in 15-mL tubes, grown for 1 hour and subsequently induced with 10 μL of aTc or 3OC_6-HSL (100 x stock) at different concentrations; for each concentration used at least three replicates were considered. After 1-hour growth in the same conditions, 200 μL of each culture were transferred into a 96-well microplate and assayed with the protocol described in Sec. 2.5.

Dynamic characterization

The model shown in Eq. 2.3 was used to identify the r_{p_X} parameters, which describe the dynamic behaviour of promoters given the Hill parameters previously identified via steady state characterization. Similarly to the protocol of steady state characterization, long-term glycerol stocks were streaked on selective LB agar plates and grown overnight at 37°C; then, single colonies were inoculated in 0.5 mL M9 pH 6 medium supplemented with the proper antibiotic in 2-mL tubes at 37°C, 220 rpm. After overnight growth, cultures were 1:100 diluted in 200 μL of fresh broth in a 96-well microplate, which was incubated in the F200 reader at 37°C, monitoring absorbance. When cells reached an optical density of 0.03-0.05, cultures were induced with 2 μL of inducer (to yield a final concentration of 100 ng/mL of aTc or 1 μM of 3OC_6-HSL) and the microplate was incubated again, monitoring fluorescence and optical density as previously described. Data were processed as described in Sec. 2.5 to obtain the synthesis rate per cell (S_{cell}) time series. The r_{p_X} parameter was identified from such data (at least in triplicate), given the maturation rate a and the parameters involved in the static induction curve, previously estimated.

Measurement of the RFP maturation rate

The RFP maturation rate was estimated from cultures bearing plasmid with regulated or constitutive production of RFP. Long-term glycerol stocks were streaked on selective LB agar plates with the proper antibiotic; single colonies were inoculated and grown overnight at 37°C in M9 pH 6 in 2-mL

tubes, 220 rpm; then cultures were 1:100 diluted in 200 μL of fresh broth in a 96-well microplate. They were incubated at 37°C, monitoring growth and fluorescence as described in Sec. 2.5. Inducers (aTc and $3OC_6$-HSL) were properly added at the beginning of the experiment. After a 3-hour growth, bacteria are in exponential phase and gentamycin (1 mg/mL) was added, continuing to monitor cultures growth and fluorescence. Gentamycin stops the translation processes of cells by irreversibly binding the 30S ribosomal subunit, hence their life cycle is stopped. In this way, the amount of immature protein that has been accumulated in the culture at the time of gentamycin supplementation begins to mature in its functional fluorescent form and no more immature protein is synthesized, so the observed dynamics of fluorescence represents the RFP maturation. Data were normalized by subtracting the initial value and expressing the fluorescence as the percentage of the maximum in order to facilitate the fitting procedure. A first-order exponential model with a delay (Eq. 2.5) was used to estimate the maturation rate of RFP; in this model, A is an arbitrary gain, a is the maturation rate expressed as min^{-1} and τ is the delay expressed in min.

$$Normalized\ A.U._r = A \cdot \left(1 - e^{-a \cdot (t-\tau)}\right) \tag{2.5}$$

2.6.3 Enzyme characterization

After having evaluated the parameters of promoters, the enzymes parameters were considered. In particular, the production of the signalling molecule $3OC_6$-HSL by the LuxI enzyme and its degradation by AiiA were evaluated from *ad hoc* assembled subsystems. The device producing $3OC_6$-HSL upon aTc induction is represented in Fig. 2.5; in this system, the P_{TetR} promoter is assembled upstream of the *luxI* gene, driving its transcription. In this case there are no terminators downstream of the CDS: the double terminator (BBa_B0053 and BBa_B0054) flanking the cloning site of the pSB4C5 plasmid was exploited to stop transcription. For the same reason, when using the pSB3K3 plasmid, the *his* terminator downstream of its cloning site was exploited.

Single colonies were picked after overnight growth at 37°C on streaked selective LB agar plates and inoculated in 0.5 mL M9 pH 6 at 37°C, 220 rpm overnight in 2-mL tubes. Then cultures were 1:100 diluted in 5 mL fresh broth in 15-mL tubes and grown at 37°C, 220 rpm. After one-hour growth they were induced with different aTc concentrations (0, 0.2, 0.5, 1, 2, 4, 8, 100 ng/mL). After one additional hour (starting point of the measurement) cultures began to be monitored, sampling at this time point and after 1, 2 and 4 hours. The growth of the cultures was monitored in terms of OD_{600}. Samples were centrifuged for 1 min at 13000 rpm, then the cell-free supernatants were stocked

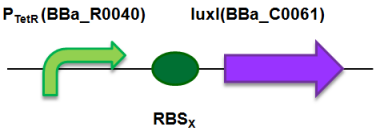


Figure 2.5: **Subsystem for LuxI parameters estimation.** Schematic representation of the genetic circuit used to estimate the LuxI enzyme parameters: the P_{TetR} promoter is assembled upstream of the *luxI* coding sequence. In this case, the transcriptional terminator downstream of the pSB4C5 or pSB3K3 cloning site (not shown in figure) is used to stop transcription.

at -20°C and used to quantify the $3OC_6$-HSL concentrations through the use of a biosensor (Sec. 2.7). As the P_{TetR} transcriptional activity is assumed to be at its steady state after 1-hour induction, the model used to estimate the activity of LuxI does not take into account the activation dynamics of this promoter. Moreover, because only the exponential phase is considered, the OD_{600} equation does not include a logistic function, but it is simply constituted by an exponential growth. The resulting model is reported in Eq. 2.6.

$$\begin{aligned} \frac{d}{dt}OD_{600} &= \mu \cdot OD_{600} \\ \frac{d}{dt}[HSL] &= k_{LuxI}(aTc) \cdot OD_{600} - \gamma_{HSL} \cdot [HSL] \end{aligned} \tag{2.6}$$

In Eq. 2.6, μ is the culture growth rate estimated from data points and γ_{HSL} is the spontaneous degradation rate of the $3OC_6$-HSL (it is reported in the equation but actually it was set to zero as $3OC_6$-HSL half-life in these conditions is much longer than the time needed in this experimental set-up, see App. A). The $k_{LuxI}(aTc)$ parameter is referred to the production rate of $3OC_6$-HSL per OD_{600} unit and is considered a function of aTc concentration as the P_{TetR} output modulates the per cell concentration of the enzyme upon aTc induction. This parameter, estimated from this experimental set-up, was then used to identify the parameters of the equation that links the activity of LuxI to the per cell enzyme concentration at the steady state:

$$k_{LuxI} = \frac{k_{LuxI,max}}{1 + \left(\frac{k_{M,LuxI}}{[LuxI]^{SS}}\right)^{\eta_{LuxI}}} \tag{2.7}$$

The RFP level produced by a promoter in the same conditions as the enzyme expression subsystem (i.e., same strain, plasmid and RBS) was used to estimate the intracellular level of LuxI, in the hypothesis that proteins produced in the same conditions as RFP follow the same trend as the fluorescent protein, except for a possible scale factor. For this reason, the intracellular concentration of LuxI is expressed in A.U. cell^{-1}. Given the previously estimated activation parameters and aTc concentration, it is thus possible to

estimate the per cell concentration of enzyme at the steady state of promoter transcription. In particular, the LuxI protein at the steady state for an aTc concentration is given by Eq. 2.8.

$$PTetR(aTc)^{SS} = \alpha_{P_{TetR}} \cdot \left(\delta_{P_{TetR}} + \frac{1 - \delta_{P_{TetR}}}{1 + \left(\frac{k_{P_{TetR}}}{[aTc]} \right)^{\eta_{P_{TetR}}}} \right) \tag{2.8}$$

The concentration of LuxI enzyme is described in Eq. 2.9.

$$\frac{d}{dt}[LuxI] = PTetR - (\mu + \gamma_{LVA}) \cdot [LuxI] \tag{2.9}$$

Considering Eq. 2.9 at the steady state, it is possible to evaluate $[\overline{LuxI}]$ (Eq. 2.10).

$$[LuxI](aTc)^{SS} = \frac{PTetR(aTc)^{SS}}{(\mu + \gamma_{LVA})} \tag{2.10}$$

The activity of the AiiA enzyme was characterized with a similar procedure. The transcription of the *aiiA* gene was driven by the P_{TetR} promoter, as depicted by the schematic representation in Fig. 2.6. Single colonies were inoculated from streaked selective LB agar plates in 0.5 mL M9 pH 6 medium and grown overnight at 37°C, 220 rpm in 2-mL tubes. They were then 1:100 diluted in 5 mL fresh broth in 15-mL tubes and grown again at 37°C, 220 rpm. After one-hour growth cultures were induced with different aTc concentrations (0, 0.2, 0.5, 1, 2, 4, 8, 100 ng/mL), waiting for cells to grow for another hour. Then $3OC_6$-HSL was added; its initial concentration was 100 nM. At this time cultures began to be monitored in terms of OD_{600} and sampled (after 0, 1, 2 and 4 hours). Samples were centrifuged 1 min at 13000 rpm and cell-free supernatants stocked at -20°C; then, the biosensor (Sec. 2.7) was again used to quantify the remaining $3OC_6$-HSL. Data were used to identify the parameters of the model in Eq. 2.11, which links the activity of the AiiA enzyme to the decay of $3OC_6$-HSL.

$$\begin{aligned} \tfrac{d}{dt} OD_{600} &= \mu \cdot OD_{600} \\ \tfrac{d}{dt}[HSL] &= -\left(k_{AiiA}(aTc) \cdot OD_{600} + \gamma_{HSL}\right) \cdot [HSL] \end{aligned} \tag{2.11}$$

In Eq. 2.11, the first equation represents cultures growth while the second one describes the decay of the signalling molecule due to the action of AiiA enzyme. In particular, the AiiA per cell level is expressed as a function of aTc due to the regulation of its transcription by the P_{TetR} promoter in the used subsystem of Fig. 2.6. Again, the estimated activities were used to identify the parameters of the equation linking the intracellular concentration of AiiA to

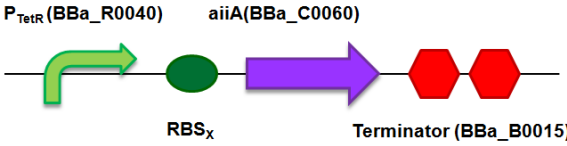


Figure 2.6: **Subsystem for AiiA parameters estimation.** Schematic representation of the genetic circuit used to estimate the AiiA enzyme parameters: the P_{TetR} promoter is assembled upstream of the *aiiA* coding sequence.

the activity of AiiA (Eq. 2.12). These parameters were identified considering the steady state of AiiA concentration.

$$k_{AiiA} = \frac{k_{AiiA,max}}{1 + \left(\frac{k_{M,AiiA}}{[AiiA]^{SS}}\right)^{\eta_{AiiA}}} \tag{2.12}$$

The intracellular concentration of AiiA enzyme as a function of aTc concentration was computed as in Eq. 2.13.

$$[AiiA](aTc)^{SS} = \frac{PTetR(aTc)^{SS}}{(\mu + \gamma_{LVA})} \tag{2.13}$$

After being estimated, the previously described parameters were used to simulate the behaviour of the whole circuit, assuming that the identified parameters do not change when parts are used in a different context.

2.7 3OC$_6$-HSL quantification

3OC$_6$-HSL is able to diffuse through cell membranes and activate the transcription of the P_{Lux} promoter, in presence of LuxR. For these reasons, it is possible to measure 3OC$_6$-HSL production and degradation in the cultures of interest by taking supernatant samples and analyzing them, assuming that intracellular and extracellular concentrations are identical due to diffusion. A biosensor is used to carry out this task. A recombinant *E. coli* culture bearing the P_{Lux} promoter upstream of the RFP gene was used as biosensor. The circuit also includes a constitutive expression cassette for LuxR. Because the RFP synthesis rate per cell is function of 3OC$_6$-HSL concentration, it is possible to evaluate the inducer concentrations of real samples via the protocol described below.

0.5 mL of selective M9 were inoculated with 1 μL of the biosensor glycerol stock and incubated for 16 h at 37°C, 220 rpm. Bacteria were 50-fold diluted in fresh selective M9 and grown under the same conditions as before. After

2-3 h from dilution, 190-μL aliquots were transferred into a 96-well microplate. Cells at this point are at an OD_{600} of about 0.03 (exponential growth phase). Wells were induced with 10 μL of properly diluted samples. Standard calibration curves were obtained by inducing the wells with 10 μL of properly diluted inducer amounts; 10 μL of properly diluted supernatants (in order to fall into the linear range of the biosensor) were assayed. The microplate was incubated in the Infinite F200 reader and it was assayed with the following kinetic cycle, programmed via the i-control™ software (Tecan): linear shaking (3-mm amplitude, 15 s), wait (5 s), OD_{600} measurement, fluorescence measurement (excitation at 535 nm, emission at 620 nm, gain=50, to detect RFP signal), repeat cycle every 5 min. Each calibration curve dilution sample was prepared in the same growth medium of the samples. A non-fluorescent culture and sterile medium were always included in each experiment to estimate the background fluorescence and absorbance. Calibration curve points and unknown samples points were assayed in duplicate.

The acquired time series were processed as previously described (Sec. 2.5) to obtain a value proportional to the average RFP protein synthesis rate per cell (${S_{cell}}^{SS}$), which is related to the inducer concentration in the wells.
The obtained standard calibration curve was fitted with the Hill function in Eq. 2.14.

$$S_{cell}^{SS} = \alpha_{P_{Lux}} \cdot \left(\delta_{P_{Lux}} + \frac{1 - \delta_{P_{Lux}}}{1 + \left(\frac{k_{P_{Lux}}}{[HSL]} \right)^{\eta_{P_{Lux}}}} \right) \tag{2.14}$$

The inducer concentration of the unknown samples was computed from their S_{cell} value via the following equation (Eq. 2.15).

$$[HSL] = k_{P_{Lux}} \cdot \left(\frac{\alpha_{P_{Lux}} \cdot \delta_{P_{Lux}} - S_{cell}^{SS}}{S_{cell}^{SS} - \alpha_{P_{Lux}}} \right)^{\frac{1}{\eta_{P_{Lux}}}} \tag{2.15}$$

When required, the resulting value was multiplied by the applied dilution factor of the sample.

2.8 Chemostat experimental set-up

Chemostat is a term that refers to a generic bioreactor in which the volume of liquid cell cultures is maintained constant by simultaneously adding fresh medium and removing cultured broth.
In this work the chemostat was used to maintain the cell density at a fixed value, in order to work in an experimental context in which cells remain indefinitely in their exponential growth phase, during which their growth rate is

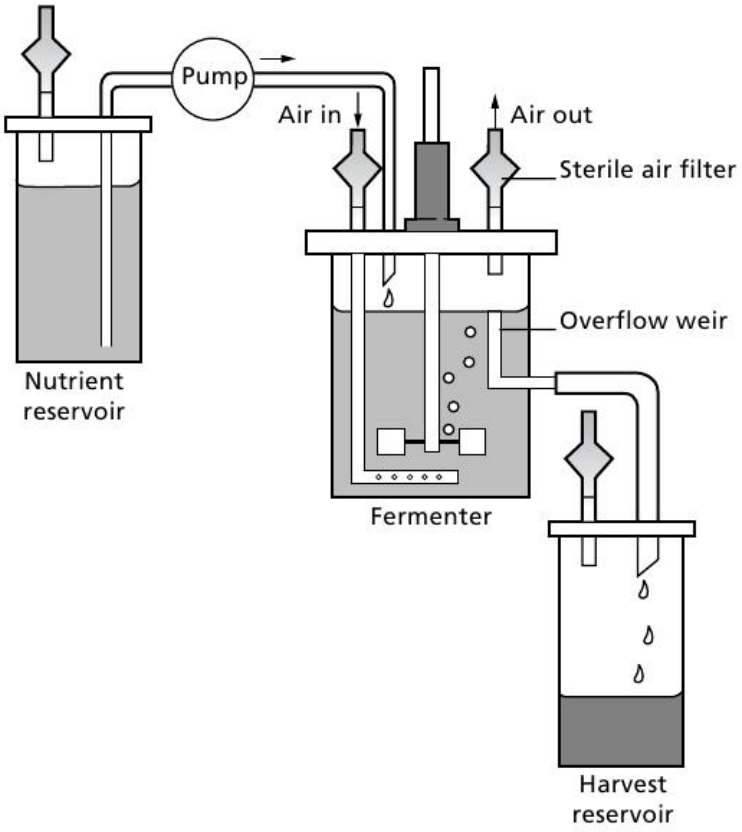


Figure 2.7: **Typical chemostat configuration.** A typical configuration of a chemostat: the fresh medium is withdrawn from a reservoir and put into the reactor at a desired flow rate, while an overflow weir maintains the volume constant. Stirring is performed. Other devices for the control of parameters such as oxygenation or pH could be present [66].

constant and maximum, and protein expression is not affected as in the saturation phase. Moreover, this was done as the parameters of the model proposed in Sec. 2.6.1 were identified during the exponential growth.
Fig. 2.7 shows a typical configuration for a continuous culture apparatus: stirring is performed and cultures are continuously fed with fresh medium. At the same time, an overflow weir or a levelling device washes away cultured medium in order to maintain the volume inside the reactor constant.

In order to maintain the cell density inside the fermenter constant, cultures need to be diluted at the same rate at which they are growing; given that batch cultures growth in exponential phase can be mathematically described by Eq. 2.16:

$$\frac{d}{dt}OD_{600} = \mu \cdot OD_{600} \tag{2.16}$$

the dilution rate D of Eq. 2.17 must equal μ.

$$\frac{d}{dt}OD_{600} = \mu \cdot OD_{600} - D \cdot OD_{600} \tag{2.17}$$

The dilution rate is given by $D = \frac{k}{V}$, where k is the inflow expressed as mL min^{-1}, and V is the volume of the culture, expressed in mL [66]. Thus, when D equals the growth rate μ, which is dependent from the OD_{600} but in the exponential phase is maximum, one could obtain indefinitely constant culture density.
In the experimental set-up used in this work, the LAMBDA MINIFOR fermenter-bioreactor (Sec. 2.3.2) was used to realize the chemostat: two peristaltic pumps were used to put in fresh medium (IN pump) and to remove the cultured broth from the 0.4 L vessel (OUT pump) through silicone tubes. The IN pump needed to be calibrated: different velocities were used in order to estimate the characteristic that relates the revolutions per minute of the pump to the actual flow rate. The OUT pump was connected to the sampling port that was used as a levelling device in order to maintain constant the culture volume in the reactor. Oxygenation was monitored but not controlled; however, as typical experiment volumes were between 25 - 30 mL, it was always maximal. Temperature was set to 37° and stirring set to 10 Hz. Finally, pH was measured through the pH probe (not controlled) to check the maintenance of pH 6.
The protocol used for continuous culture experiments is here described: glycerol stocks were streaked on selective LB agar plates, then a colony was picked and inoculated in 5 mL M9 pH 6 in a 15-mL tube and grown overnight at 37°C, 220 rpm shaking. Then, the culture was 1:100 diluted in 35 mL fresh M9 pH 6 in the 0.4 L vessel and grown at 37°C, stirring 10 Hz until the desired optical density was reached. During this period OD_{600} was measured by sampling each hour and measuring absorbance with the Infinite F200 reader. Data were subtracted their blank and used to estimate the cultures growth rate μ; this was fundamental in order to compute the flow rate of the IN pump and set its revolution per minutes in relation to the previously estimated characteristic. However, growth rate was highly reproducible among different experiments with both open- and close-loop circuits. When cells reached the critical OD_{600} the chemostat mode started: the OUT pump was always set to 200 rpm while the IN pump to the value given by $k = \mu \cdot V$. At this time point cultures were induced with 100 ng/mL aTc, in order to tune on the production of $3OC_6$-HSL in both the open- and close-loop genetic circuits; cells were subjected to a 100 ng/mL constant aTc concentration by adding a proper amount of this inducer in the input reservoir. Cultures were then sampled each half an hour in order

to monitor the optical density and properly tune the rpm of the IN pump, if OD_{600} trend was significantly increasing or decreasing (due to possible errors in tubes positioning). Moreover, an aliquot ($\sim$ 100 μL) of the sample was centrifuged (13000 rpm, 1 min) and the cell free supernatant stored at -20° for the subsequent $3OC_6$-HSL quantification (Sec. 2.7).

2.9 Monte Carlo method for uncertainty propagation

The model that describes the negative feedback for the control of $3OC_6$-HSL is composed by 6 equations and 20 parameters. To verify the possibility to bottom-up design a genetic circuit that performs a non-trivial function, *ad hoc* assembled subsystems were considered for the parameters identification. In this way, each subsystem is described by a mathematical model with a reduced number of equations compared to the full circuit model, and parameters could be identified from experiments on such subsystems.
The uncertainty affecting the parameters needs to be propagated from the simplest subsystems to the whole genetic controller, giving the measure of the confidence intervals of the observable output, i.e., the concentration of $3OC_6$-HSL. Thus, it is important not only to estimate the parameters on the basis of the experimental tests but also to give an estimate of their distributions.
Eq. 2.18 describes the output $y(t)$ of a generic system, where g is the model output prediction and p represents the vector of the M parameters that have to be identified.

$$y\left(t\right) = g\left(t, p\right) \tag{2.18}$$

The observations, z_k, can be described by Eq. 2.19, in which the index k ($1 \leq k \leq N$) represents the *k-th* sample.

$$z_k = y\left(t_k\right) + \nu_k = g\left(t_k, p\right) + \nu_k \tag{2.19}$$

In Eq. 2.19, the ν_k term is an additive error that affects the z_k samples and it is usually modelled as a random variable with zero-mean; its variance could be constant, σ_ν^2, or dependent on the measurement itself, $\sigma_{\nu,k}^2$, considering, for example, a constant CV model. Eq. 2.19 can be rewritten in matrix notation as in Eq. 2.20, given that $z = [z_1, ..., z_N]^T$, $g = [g\left(t_1, p\right), ..., g\left(t_N, p\right)]^T$ and $\nu = [\nu_1, ..., \nu_N]^T$.

$$z = g\left(p\right) + \nu \tag{2.20}$$

From these hypotheses, it could be stated that $E[\nu] = 0$ and $\Sigma_\nu = E[\nu\nu^T]$. Moreover, the matrix Σ_ν can be rewritten as in Eq. 2.21, where B is an N-by-N matrix.

$$\Sigma_\nu = \sigma_\nu^2 B \tag{2.21}$$

Considering a constant variance model, it could be assumed, as it was done in this study, $B = I$, namely the N-by-N identity matrix. Under this assumptions, the covariance matrix Σ_ν is known, except for a scale factor σ_ν^2. As the noise affecting observations was hypothesized to be additive, the residuals r (Eq. 2.22) can be considered a realization of the random variable ν, with the same statistical properties.

$$r = z - g\left(p\right) \tag{2.22}$$

The goal of parameter estimation is to minimize the sum of squared residuals (Eq. 2.23).

$$J\left(p\right) = \parallel z - g\left(p\right) \parallel^2 \tag{2.23}$$

Thus, the parameters vector $\hat{p}$ is given by Eq. 2.24.

$$\hat{p} = \underset{p}{\arg\min}\, J\left(p\right) \tag{2.24}$$

If the minimization problem is linear (i.e., $g\left(p\right)$ linearly depends on p), then $\hat{p}$ can be evaluated analytically; otherwise, several algorithms have been proposed in order to approximate the solution (e.g., the Gauss-Newton algorithm). Moreover, it is of interest to give some statistical details about the dispersion of the parameters in terms of standard deviation or percentage coefficient of variation (CV). Unless hypothesizing a normal distribution of the parameter estimation error, it is not possible to evaluate the confidence intervals or to obtain the parameters distribution; consequently, it would not be possible to propagate the uncertainty from the simplest subsystems to the output of the model here proposed. Under the assumption that the parameter estimation error is normally distributed, the covariance matrix can be approximated by:

$$\Sigma_p = (S^T \Sigma_\nu^{-1} S)^{-1} \tag{2.25}$$

where S is the Jacobian matrix evaluated by the minimization routine. The diagonal elements of the covariance matrix are the parameters variances.
Moreover, the unknown σ_ν^2 needed to define Σ_ν, can be computed from the sum of squared residuals evaluated for $\hat{p}$ (Eq. 2.26).

$$\hat{\sigma}_\nu^2 = \frac{J\left(\hat{p}\right)}{N - M} \tag{2.26}$$

There are several techniques specifically designed to obtain the parameters distributions; the one implemented here relies on Monte Carlo simulations. This is a useful methodology, based on random sampling, used in several applications such as physics or mathematics, when it is not possible to obtain a closed-form expression for the solution of a problem. The general Monte Carlo algorithm applied to parameter estimation consists in performing a large number of parameters estimations on synthetic data sets. In particular, the first step of the algorithm is the generation of several synthetic data sets; these have to maintain the same features as the original observations. For this purpose, noise sampled from a suitable distribution is added to the model predictions evaluated in $p = \hat{p}$, i.e., $\hat{y} = g(t, \hat{p})$; assuming that noise is normally distributed, with zero mean and variance $\hat{\sigma}_\nu^2$, the synthetic data sets are generated according to Eq. 2.27:

$$y_{noisy}^i = \hat{y}^i + h^i \tag{2.27}$$

where $1 \leq i \leq W$ refers to the *i-th* synthetic data set generated, while h is an additive noise independently sampled from the gaussian distribution with zero mean and variance $\hat{\sigma}_\nu^2$.
Then, as a second step, parameters are iteratively estimated on each of the W synthetic data sets generated and according to Eq. 2.22, residuals are evaluated as:

$$r^i = y_{noisy}^i - g(t, p^i) \tag{2.28}$$

In this way, a large number of estimates can be done in order to obtain the parameters distributions. It is also possible to give a measure of the dispersion of the parameters, evaluating the standard deviations, the CVs, or exploiting the percentiles. When the model depends on previously estimated parameters, during the identification procedure, they are extracted from the already estimated distributions.
In this study, after having *a posteriori* estimated the variance of the data, according to Eq. 2.26, a normal distribution was assumed for ν, with zero mean and variance $\hat{\sigma}_\nu^2$. Noise was added to model nominal predictions, generating 10^4 synthetic data sets that were subsequently used for parameters estimation. The uncertainty was propagated from the simplest systems (i.e., the ones that do not require "upstream" parameters to be identified) to the more complex ones (i.e., the ones that rely on the estimated values of other parameters, which are characterized by a distribution) and then to the whole model of the negative feedback network herein proposed.
Not all the parameters were considered to be affected by uncertainty; in fact, the cultures growth rate μ, the saturation growth value $OD_{600,max}$, the spontaneous $3OC_6$-HSL degradation γ_{HSL} and the protein degradation rate due

to the LVA tag, γ_{LVA}, were set to fixed values. The identification of the RFP maturation rate, a, is the starting point for uncertainty propagation as it appears in the model of the promoters activation (Eq. 2.3). Once the RFP maturation rate distribution was estimated, it was used to propagate the variability to the steady state characterization of the promoters. In particular, once obtained 10^4 synthetic data sets generated by adding noise to the model predictions, given the nominal values for the parameters, 10^4 runs of identification for α_{p_X}, δ_{p_X}, k_{p_X} and η_{p_X} were performed, sampling a from its own distribution. Subsequently, applying the same strategy, the distributions of the poles characterizing the promoters dynamics were estimated, generating 10^4 synthetic data sets by adding noise to the model predictions evaluated for the nominal parameters, and sampling from the distributions of α_{p_X}, δ_{p_X}, k_{p_X}, η_{p_X} and a at each identification step. Regarding enzyme activities, for each experiment performed on cultures expressing LuxI or AiiA, the predictions of the model, estimated on the basis of the experimental observations, were used to generate 10^4 synthetic data sets by adding gaussian noise to the model predictions and to obtain the distributions of $k_{LuxI}(aTc)$ and $k_{AiiA}(aTc)$, given an aTc concentration. The distributions of the per cell concentrations of the enzymes at the steady state, given an aTc concentration, were obtained by sampling from the distributions of steady state promoters parameters, as the P_{TetR} promoter was considered at its steady state as hypothesized in Sec. 2.6.3. Finally, sampling from the distributions of the per cell enzymes concentrations and the enzymes activities at the same aTc concentration, it was possible to generate the synthetic data sets used to estimate the parameters describing the enzymes activities in terms of the Hill functions shown in Sec. 2.6.3 (see Eq. 2.7 and Eq. 2.12).
Once the distributions of all the parameters involved in the full model of the feedback circuit were obtained, the 95% confidence interval of the simulated output could be evaluated, randomly sampling from the parameters distributions previously computed.

Chapter 3

Results

This chapter is dedicated to the results obtained for the negative feedback controller. First, simulations are performed to study the behaviour setting the parameters to plausible values, focusing on the possibility of realizing such a genetic circuit. Then, parameter identification from subsystems will be illustrated. As data are affected by noise, parameter values uncertainty was propagated to the observable output by Monte Carlo method, thus having the possibility to obtain its confidence intervals. A discussion on the experimental set-up used to test the performances of the whole close-loop circuit will follow and, finally, the simulations and experimental results will be shown. The full circuit was tested in growing cultures (batch culture mode) and in chemostatic conditions (continuous culture mode, Sec. 2.8). The latter condition is used to maintain a defined cell density over time, thus preventing the cultures to enter the saturation growth phase, in which cells may exhibit an unpredictable behaviour.

3.1 Simulation results

The mathematical model presented in Sec. 2.6.1 has been derived relying on the *a priori* known molecular interactions and transcription regulations of the two involved promoters, considering that P_{TetR} drives the production of the target molecule and P_{Lux} acts as a sensor, driving the transcription of *aiiA* in relation to the $3OC_6$-HSL concentration. Initially, plausible values were assigned to the parameters in order to establish if, in principle, the designed genetic circuit could work as a negative feedback controller in the batch culture mode (Fig. 3.1). Different parameter combinations have been tested and compared with the open-loop condition, a situation in which no regulation is present in the synthetic genetic program and the target molecule is continu-

Table 3.1: **Parameter values used in simulations.** Parameters are reported with their values used in simulations.

Parameter	Value
μ	0.0126
$OD_{600,max}$	1
$r_{P_{TetR}}$	0.05
$\alpha_{P_{TetR}}$	1
$\delta_{P_{TetR}}$	0.01
$k_{P_{TetR}}$	2.5
$\eta_{P_{TetR}}$	5
$r_{P_{Lux}}$	0.1
$\alpha_{P_{Lux}}$	2
$\delta_{P_{Lux}}$	0.01
$k_{P_{Lux}}$	150
$\eta_{P_{Lux}}$	1
γ_{LVA}	0.0173
γ_{HSL}	e-4
$k_{LuxI,max}$	8
$k_{M,LuxI}$	12
η_{LuxI}	2
$k_{AiiA,max}$	0.5
$k_{M,AiiA}$	25
η_{AiiA}	3

ously produced without being degraded by AiiA. If not differently reported, the parameters used for simulations are reported in Tab. 3.1; most of the parameters were set to reasonable values on the basis of previous experiments (data not shown) while some of them were set to values found in the literature.

It can be observed from Fig. 3.1 that $\alpha_{P_{TetR}}$ and $k_{LuxI,max}$ impact on the dynamics and the steady state value of $3OC_6$-HSL concentration; if they increase, the $3OC_6$-HSL at the steady state increases, reaching its maximum more and more quickly. $\alpha_{P_{Lux}}$ is related to the controller which is faster and more effective as the value of this parameter increases. The same behaviour can be observed for $k_{AiiA,max}$; when this parameter was set to zero, the open-loop condition was reproduced, as no regulation occurs.

Analogously, several parameter sets were used to simulate the continuous culture condition and highlight the differences between the non regulated and the regulated system. The difference between the open-loop and the close-loop

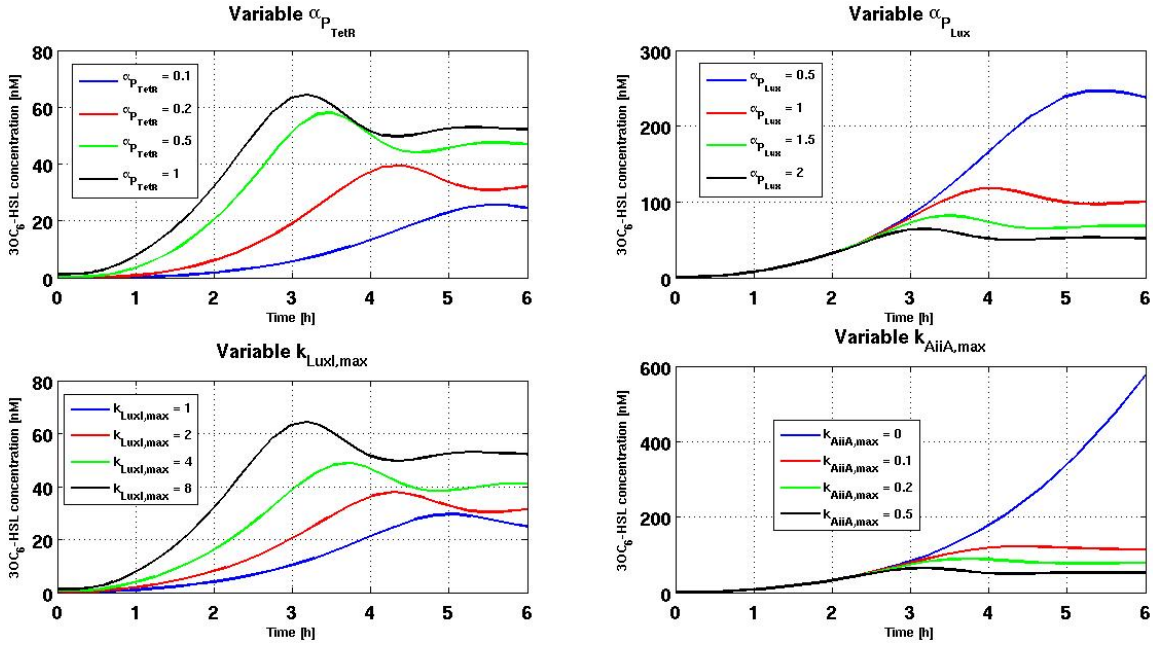


Figure 3.1: **Simulations for different parameter values in batch culture mode.** Effects of parameters variation on the steady state concentration of $3OC_6$-HSL: the parameters considered are $\alpha_{P_{TetR}}$, $\alpha_{P_{Lux}}$, $k_{LuxI,max}$ and $k_{AiiA,max}$. The starting OD_{600} was set to 0.02. When $k_{AiiA,max}$ is set to zero the behaviour is analogous to an open-loop circuit, i.e., a device that only produces the signalling molecule without degrading it.

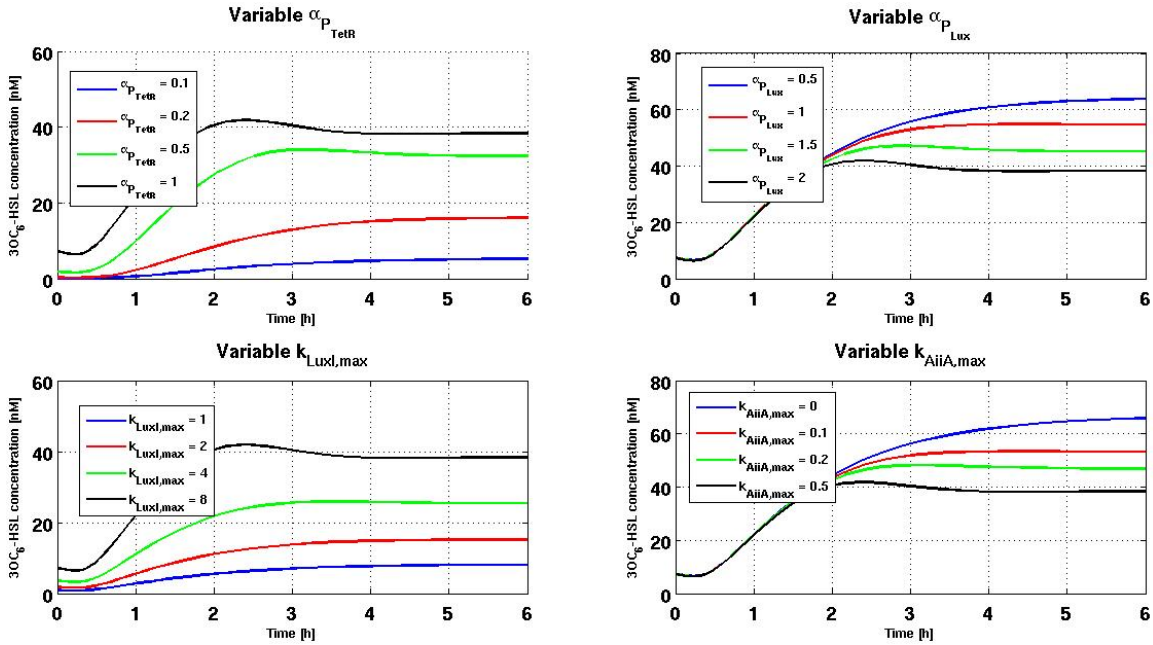


Figure 3.2: **Simulations for different parameter sets in continuous liquid cultures.** Effects of parameters variation on the steady state concentration of $3OC_6$-HSL: the parameters considered are $\alpha_{P_{TetR}}$, $\alpha_{P_{Lux}}$, $k_{LuxI,max}$ and $k_{AiiA,max}$. The constant OD_{600} was set to 0.12. When $k_{AiiA,max}$ is set to zero the behaviour is analogous to an open-loop circuit only producing the signalling molecule. Unlike growing liquid cultures, in this case the signalling molecule is diluted at a constant rate equal to the cells growth rate.

is visible at the steady state also in the chemostat mode (Fig. 3.2); it is of interest to underline that the continuous culture mode is characterized by the fact that the $3OC_6$-HSL is not only degraded by the AiiA enzyme but also diluted by a factor equal to cell growth rate (Sec. 2.8).

Again $\alpha_{P_{TetR}}$ and $\alpha_{P_{Lux}}$ impact on the steady state in an opposite direction, as increasing the former means increasing the $3OC_6$-HSL steady state concentration while increasing the latter means degrading the signalling molecule faster. The parameters related to the enzymes behave in a similar manner as the higher $k_{LuxI,max}$, the higher the steady state, while the higher $k_{AiiA,max}$, the lower the concentration at the steady state. In both cases, i.e., batch culture and chemostat, when introducing a degrading action, which is biologically carried out by the AiiA enzyme, the concentration of $3OC_6$-HSL reaches a steady state which can vary depending on the parameters values; conversely, when the degradation is not present, in the case of growing cultures the steady state concentration of the signalling molecule grows continuously as a function of the increasing optical density, while in the case of continuous cultures it reaches a

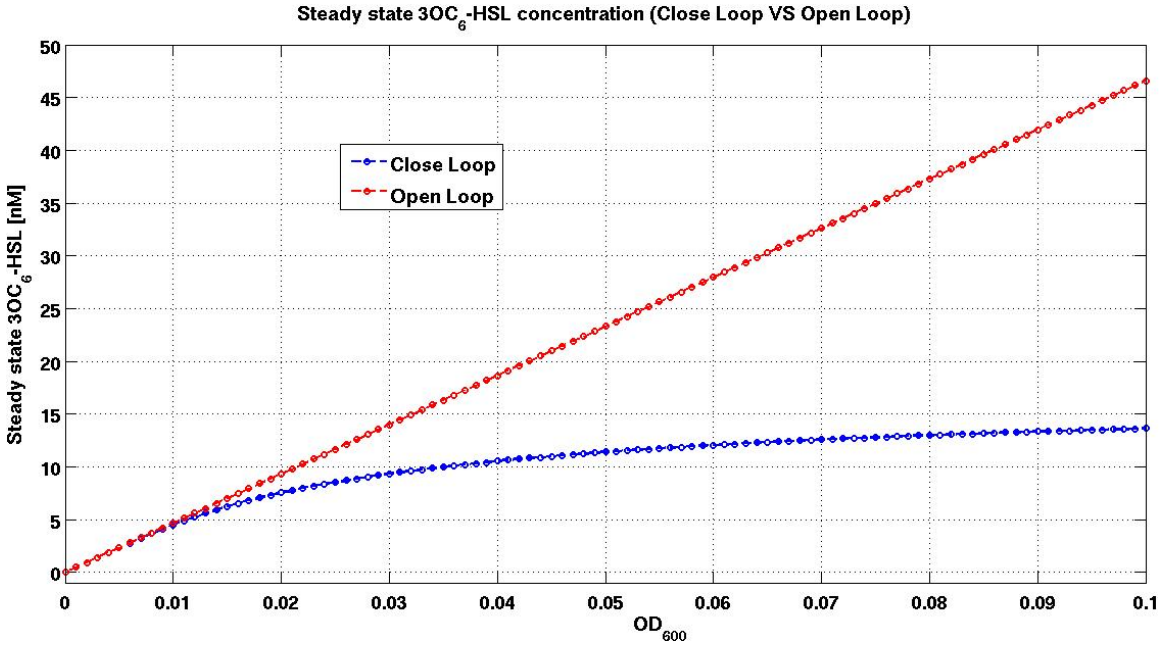


Figure 3.3: **Steady state concentration as a function of optical density in continuous cultures.** In the case of the non-regulated genetic circuit the steady state concentration of $3OC_6$-HSL is predicted to be linearly dependent on the optical density, while, when cells implement the negative feedback, the action of AiiA enzyme is clearly visible as the steady state value grows significantly slower when optical density is increased.

steady state higher than in the presence of the negative feedback.
Simulations also show the steady state concentration as a function of cell density in chemostatic conditions; the model was simulated for the open- and close-loop (Fig. 3.3). When no negative feedback is present, the steady state concentration grows linearly with the optical density, while in the case of the regulatory genetic network the concentration at the steady state also grows with optical density but with a significantly lower slope, reaching stationariety at sufficiently high density values.

According to the shown results, in principle the circuit can work as a negative feedback regulator. It reaches a steady state that should be independent of cell density for a wide range of OD_{600} values, while the open-loop configuration shows a linear dependence of steady state $3OC_6$-HSL with cell density. Theoretically, given the circuit architecture, the steady state could be tuned by increasing or decreasing the produced LuxI and/or AiiA enzymes via circuit modifications, e.g., by changing their RBSs or plasmid copy number (see Fig. 3.1 and Fig. 3.2). Another modification that could affect the $3OC_6$-HSL value is the expression level of LuxR, which binds $3OC_6$-HSL and activates

Table 3.2: **RBSs efficiencies** RBSs strengths as declared in the Registry of Standard Biological Parts; BBa_B0034 is the reference RBS and has efficiency 1.

RBS (BioBrick Code)	Declared efficiency
BBa_B0031	0.07
BBa_B0032	0.3
BBa_B0030	0.6
BBa_B0034	1

P_{Lux}: this modification is expected to change the switch point of the Hill function describing P_{Lux} activation, thus changing the AiiA production profile upon induction by $3OC_6$-HSL (simulations not shown). Before proceeding with subsystems characterization, parameter identification and full circuit testing, it is important to choose parts that provide non-zero activity of the two enzymes. In fact, it is well known that promoters with low activity can lead to insufficient levels of mRNA and thus of protein; even with a high level of mRNA, low-strength RBSs can lead to insufficient amount of translated protein; similar effects leading to insufficient enzyme levels can occur if the circuit is present at a too low copy number. Given the P_{Lux} and P_{TetR} promoters, the selection of suitable RBSs and plasmids was performed.

3.2 Ribosome Binding Sites and plasmid selection

RBSs are important for mRNA stability and can strongly impact on the process of production of a protein. Several efforts have recently been done in order to predict the translational output of a RBS (in terms of transcriptional initiation rate) given the downstream sequence [28, 21]. In fact the strength of an RBS is strongly dependent on the downstream sequence, thus giving conflicting results when two different coding sequences are considered. As regards the RBSs used in this work, their efficiencies have been previously evaluated via a specific fluorescent reporter system (see Tab. 3.2) and documented in the Registry [22]. However, there is no guarantee that the efficiency is the same when used with a different gene downstream. For this reason, the four RBSs were tested with *luxI* and *aiiA* to check if translational efficiency was sufficient to detect enzyme activity.

The RBS with the highest efficiency is BBa_B0034 while the weakest is BBa_B0031; BBa_B0032 and BBa_B0030 have a medium strength (Tab. 3.2).

The P_{TetR} promoter was used to regulate the transcription, maximally induced with 100 ng/mL aTc in *E. coli* MG1655-Z1. The low copy plasmid pSB4C5 was used for these tests. The results were quite different from the ones reported in Tab. 3.2. Results showed that no significant AiiA activity could be detected in low copy plasmid with any RBS at full induction of P_{TetR} (data not shown). Conversely, LuxI could be effectively tuned as a function of the RBS, yielding a detectable activity with the BBa_B0030 and BBa_B0034 RBSs (Fig. 3.4). The *aiiA* gene with the four RBSs was also tested in the pSB3K3 medium copy plasmid, again under the control of P_{TetR}. AiiA gave significantly detectable activity in two conditions, with BBa_B0034 and BBa_B0031 RBSs (Fig. 3.5). The *luxI* gene was also tested in pSB3K3, only with the BBa_B0030 RBS, yielding similar activity compared to the low copy condition (data not shown). Finally, AiiA was tested via the same P_{TetR} inducible system in the high copy plasmid pSB1A2 with the four RBSs, yielding similar results in terms of absolute activities compared to the medium copy plasmid (data not shown). For these reasons, in the final circuit, the *luxI* gene under the control of P_{TetR} was placed in a low copy plasmid (pSB4C5) and used with the BBa_B0030 RBS, while the *aiiA* gene under the control of P_{Lux} was placed in a medium copy plasmid (pSB3K3) and used with the BBa_B0034 RBS (Fig. 3.6). The non regulated counterpart is represented in Fig. 3.7; here, no enzymatic degradation occurs as AiiA is not produced by cells. The relative activities of the enzymes are very different for AiiA and LuxI. Moreover, they are not in accordance with the RBS efficiencies predicted with the RBS Calculator tool [21], given the downstream sequence and assuming that enzyme activity is proportional to enzyme concentration (data not shown).

3.3 Parameters identification from subsystems

Promoters and enzymes parameters were estimated studying the subsystems presented in Sec. 2.6.2 and Sec. 2.6.3. Parameters were thus identified on simple functional modules in order to guarantee the identifiability of the models on the basis of the experimental set-up and observations. Moreover, subsystems were studied during the exponential phase of bacterial growth, thus the predictions made by the model are consistent only assuming that the cells are growing exponentially.

3.3.1 Steady-state characterization of promoters

The two promoters of interest for the realization and investigation of the regulatory network proposed are P_{TetR} and P_{Lux}. Both are characterized by a static and a dynamic behaviour: the first one regards the input-output charac-

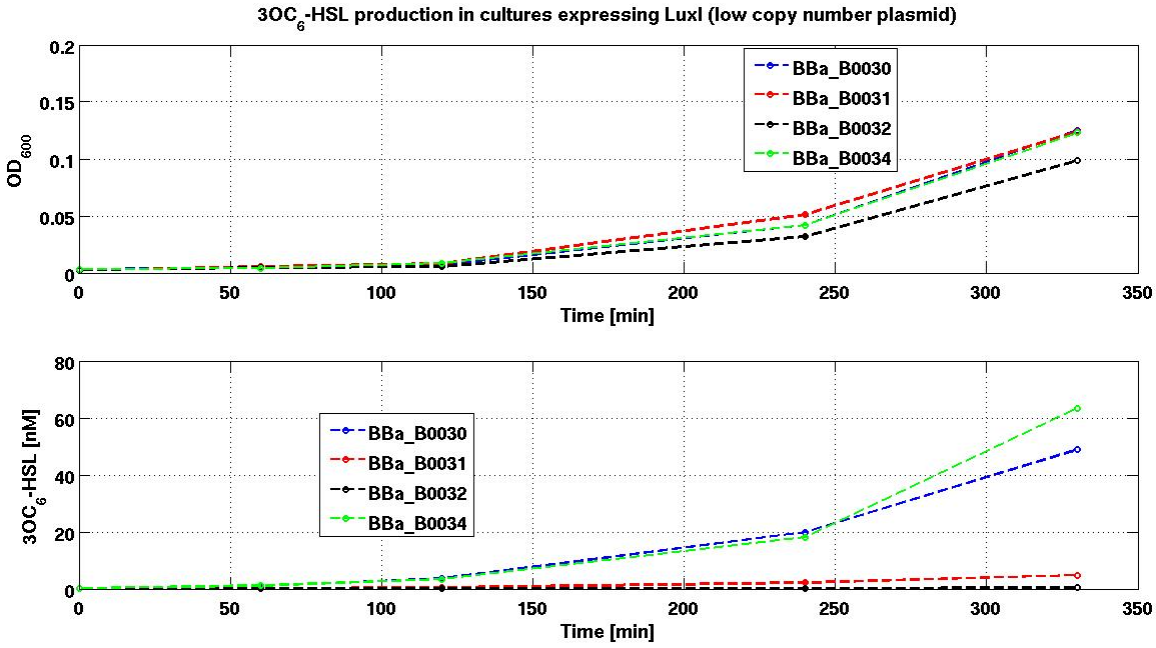


Figure 3.4: **Preliminary screening for LuxI $3OC_6$-HSL production at full induction (100 ng/mL aTc).** Production of $3OC_6$-HSL varying the RBS strength for LuxI in low copy number plasmid. Curves are identified by the code of the RBS used for the genetic construct.

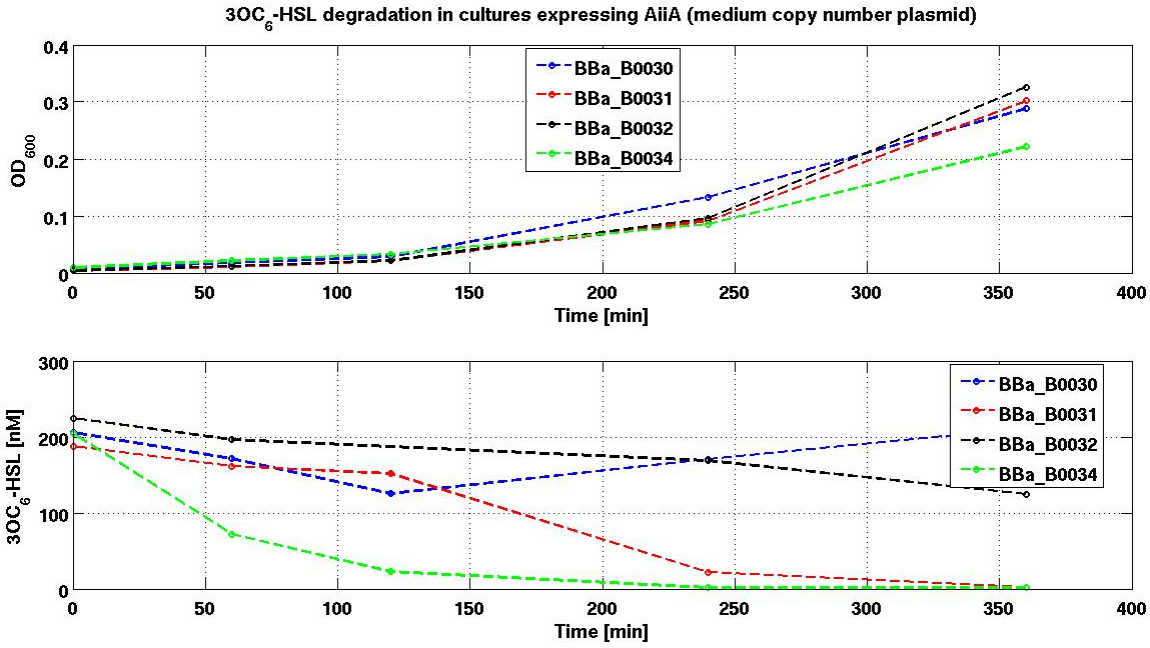


Figure 3.5: **Preliminary screening for AiiA $3OC_6$-HSL degradation at full induction (100 ng/mL aTc).** Degradation of $3OC_6$-HSL varying the RBS strength for AiiA in medium copy number plasmid. Curves are identified by the code of the RBS used for the genetic construct.

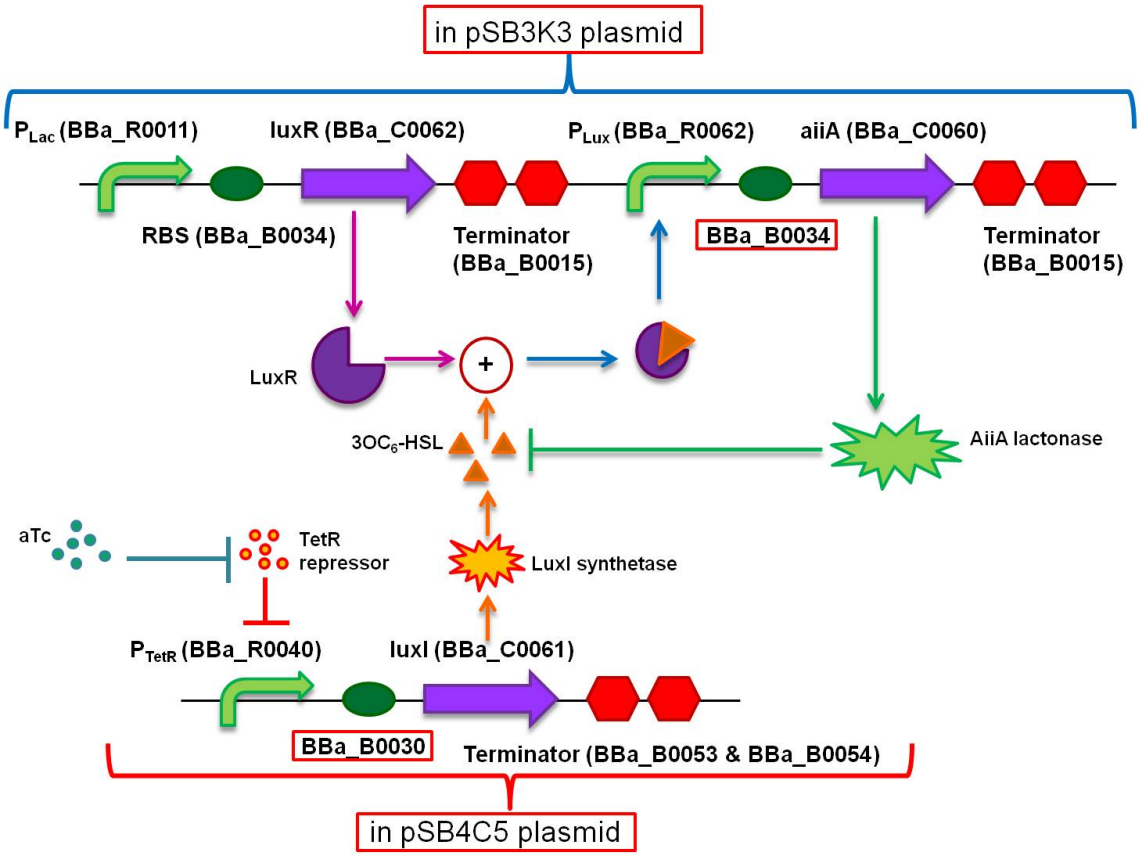


Figure 3.6: **Schematic representation of the feedback negative controller with RBSs.** The scheme represents the version of the genetic circuit that was investigated in this work: the BBa_B0030 was chosen as the RBS upstream of the *luxI* coding sequence and tested in the low copy number plasmid pSB4C5; BBa_B0034 was placed upstream of the *aiiA* gene and tested in the medium copy number plasmid pSB3K3.

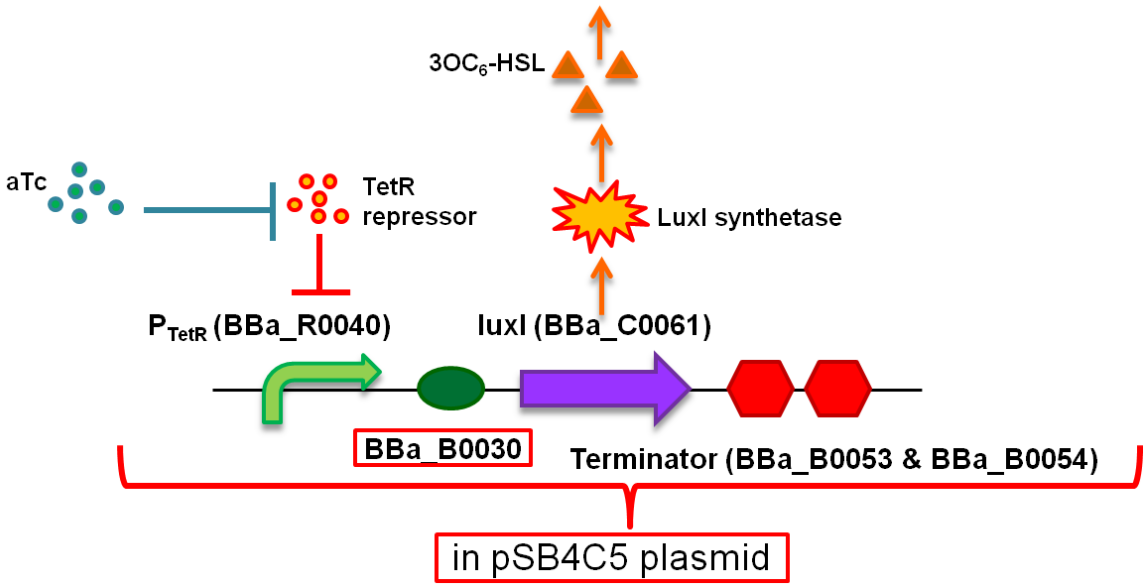


Figure 3.7: **Schematic representation of the open-loop configuration.** The scheme represents the open-loop configuration, in which the BBa_B0030 RBS was placed upstream of the *luxI* coding sequence. In this case no AiiA expression occurs, thus lacking a negative feedback.

teristic function while the second one the activation of the transcription upon induction. The parameters of interest for the static transfer function are α_P, δ_P, η_P and k_P, while the r_P is linked to the dynamic of transcription activation (see Eq. 2.3). When studying promoters it was essential to estimate the maturation rate of the reporter used, the Red Fluorescent Protein (RFP).

RFP maturation rate

The experimental set-up for RFP maturation rate estimation was previously described (Sec. 2.6.2). In order to estimate this parameter, a constitutive promoter, BBa_J23101, in both the pSB3K3 and pSB4C5 plasmids, and the two above mentioned promoters, P_{TetR} and P_{Lux} were studied. For each of these devices three independent replicates were considered and the maturation rate was jointly estimated from the data collected and processed from each run. The four conditions were studied together as, despite the differences between the expression system (i.e., promoter, RBS and plasmid copy number), the reporter is the same for all of them, so it is reasonable to consider the same maturation rate. An independent value for the static gain A was assigned to each of the twelve time series; the same was done for the delay τ, while the maturation rate a was the same for all the tested conditions.

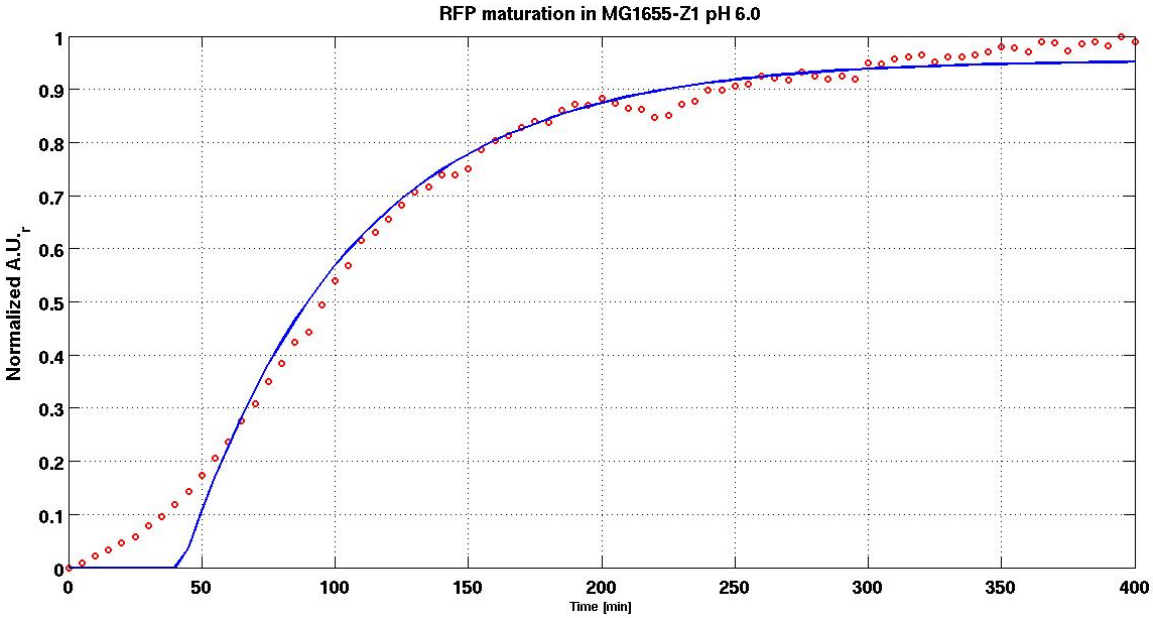


Figure 3.8: **RFP maturation rate.** In the above figure an example of fitting of the maturation data is depicted; particularly it refers to one of the 3 replicates of the P_{TetR} promoter induced with 100 ng/mL aTc. Individual A as well as τ parameters were estimated, while a, the maturation rate, was simultaneously estimated for all the conditions. t=0 corresponds to the time point of gentamycin addition.

A representative maturation profile is shown in Fig. 3.8. The estimated value was 0.0157 min^{-1} with a coefficient of variation (CV) of 1.52%; separate fittings for each of the twelve conditions gave a similar maturation rate value (data not shown). Results are consistent with the RFP maturation rate reported in literature [67].

Steady state transfer function of promoters

Promoters input-output function at the steady-state was measured using the protocol described in Sec. 2.6.2. The RFP reporter was assembled downstream of the two promoters of interest, P_{TetR} and P_{Lux}, and their activity was measured upon different inducers concentrations. The P_{Lux} promoter was tested in pSB3K3 with the BBa_B0034 RBS upstream of RFP, since this is the context in which P_{Lux} will be used in the final circuit (except RFP that is replaced by *aiiA*). The P_{TetR} promoter in the final circuit will be used in pSB4C5 upstream of *luxI* with the BBa_B0030 RBS. However, for subsystem characterization, it is also used in pSB3K3 to drive *aiiA* (with the BBa_B0034 RBS). For these reasons, P_{TetR} was tested in these two contexts, where the enzyme CDSs were replaced with RFP. For each induction and each subsystem, at least three independent replicates were considered. After calculating the synthesis rate per cell signal, in terms of red fluorescence arbitrary units A.U._r per cell per minute, each of the replicates was normalized for the output of the BBa_J23101 promoter in pSB3K3, used as internal standard reference. Fig. 3.9 shows the activity of P_{TetR} in pSB4C5 with BBa_B0030 upon increasing aTc concentration; the signal starts increasing at 1 ng/mL aTc and reaches its maximum (1.0261 A.U. cell^{-1} min^{-1}, that corresponds to $\sim$ 0.57 A.U._r cell^{-1} min^{-1}) at 10 ng/mL or greater concentrations; accordingly, Tab. 3.3 shows that the half saturation constant for the P_{TetR} promoter is 3.1102 ng/mL. Analogously, Fig. 3.10 shows the results of P_{TetR} in the pSB3K3 with BBa_B0034 condition, with the identified Hill parameters reported in Tab. 3.3; in this case, the half saturation constant is 1.8676 ng/mL and the maximum activity increases, as expected, to 2.9938 A.U. cell^{-1} min^{-1} (which corresponds to a $\overline{S_{cell}}$ signal of $\sim$ 1.75 A.U._r cell^{-1} min^{-1}). A basal activity near zero was observed when no inducer was added. The P_{Lux} promoter transfer function is shown in Fig. 3.11. The output signal of the genetic device begins to be significantly different from zero at $3OC_6$-HSL concentrations greater than 5-10 nM. Parameters values are summarized in Tab. 3.3.

Activation dynamics of promoters

Promoters transcription activity is modulated by inducers supplementation which increases the measurable output with a dynamic activation that the

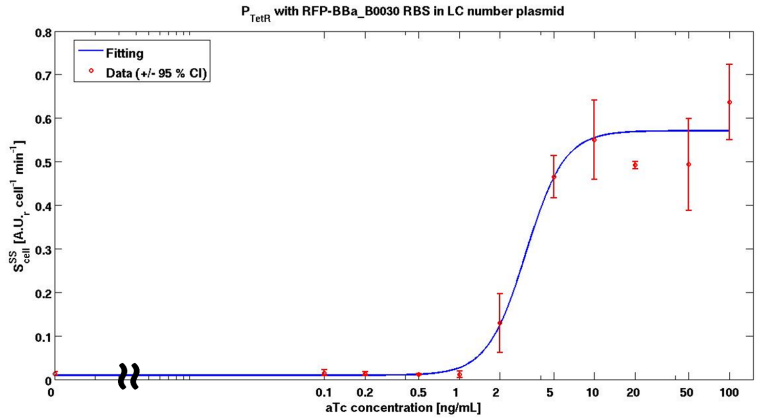


Figure 3.9: **Static transfer function for P_{TetR}.** Input-output transfer function for P_{TetR} promoter with BBa_B0030, in pSB4C5 plasmid: data were fitted with a Hill activation function.

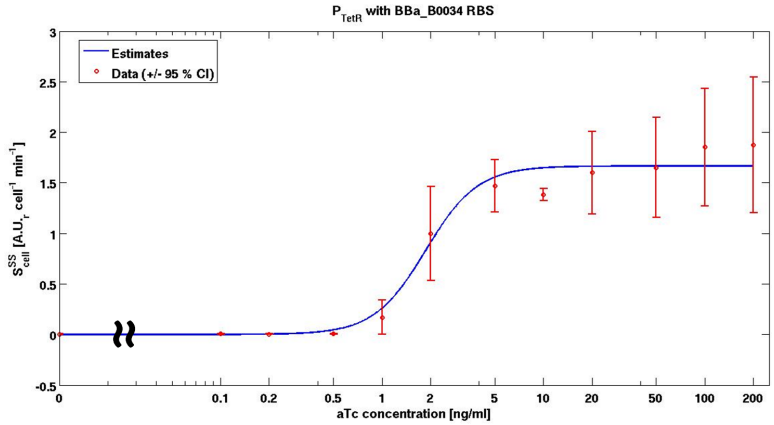


Figure 3.10: **Static transfer function for P_{TetR}.** Input-output transfer function for P_{TetR} promoter with BBa_B0034, in pSB3K3 plasmid: data were fitted with a Hill activation function.

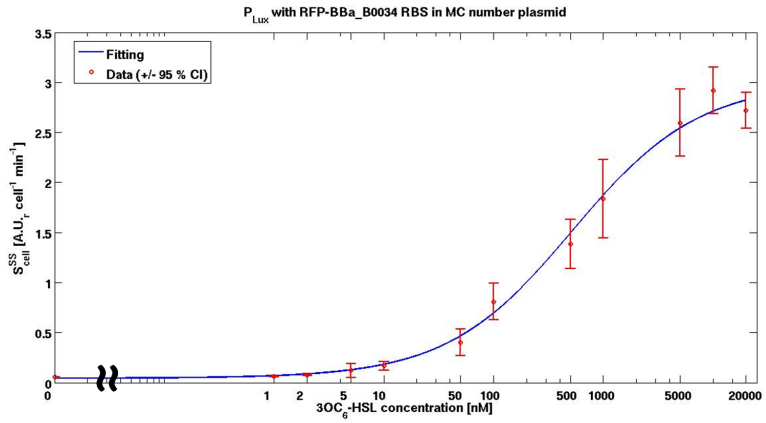


Figure 3.11: **Static transfer function for P_{Lux}.** Static transfer function for P_{Lux} promoter: data were fitted with a Hill function.

Table 3.3: **P_{TetR} and P_{Lux} input-output transfer function parameters.** Estimated parameters for the two promoters of interest; P_{TetR} transcriptional activity was evaluated with RBS BBa_B0030 in pSB4C5 low copy number plasmid and with RBS BBa_B0034 in pSB3K3 medium copy number plasmid, while P_{Lux} activity was evaluated with BBa_B0034 in pSB3K3 medium copy number plasmid. Parameters are reported together with the respective CV in brackets.

P_{TetR} with RBS BBa_B0030 in pSB4C5			
$\alpha_{P_{TetR}}$	$\delta_{P_{TetR}}$	$k_{P_{TetR}}$	$\eta_{P_{TetR}}$
1.0261 (1.36%)	0.0165 ($\gg$ 100%)	3.1102 (1.18%)	3.0373 (2.37%)
P_{TetR} with RBS BBa_B0034 in pSB3K3			
$\alpha_{P_{TetR}}$	$\delta_{P_{TetR}}$	$k_{P_{TetR}}$	$\eta_{P_{TetR}}$
2.9938 (0.67%)	6 e-11 ($\gg$ 100%)	1.8676 (3.37%)	2.7063 (7.37%)
P_{Lux} with RBS BBa_B0034 in pSB3K3			
$\alpha_{P_{Lux}}$	$\delta_{P_{Lux}}$	$k_{P_{Lux}}$	$\eta_{P_{Lux}}$
5.3930 (0.45%)	0.0147 ($\gg$100%)	529.6090 (0.02%)	0.7574 (8.50%)

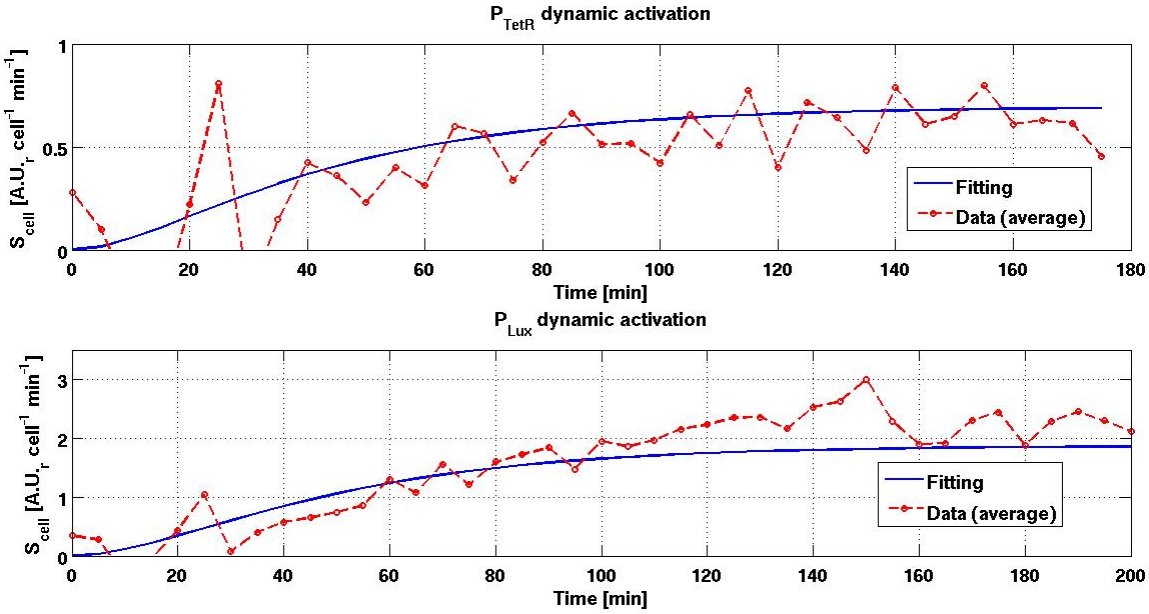


Figure 3.12: **Activation dynamics for P_{TetR} and P_{Lux} promoters.** Data fitting for dynamic activation profiles of the two promoters of interest: P_{TetR} driving RFP expression with RBS BBa_B0030 induced with 100 ng/mL aTc and P_{Lux} driving RFP expression with RBS BBa_B0034 induced with 1μM 3OC$_6$-HSL.

model described in Sec. 2.6.2 tries to capture. In this model the maturation rate of the reporter as well as the steady state promoter parameters must be known or already estimated in order to identify the activation rate r_P. The dynamics was estimated for the two promoters used in the regulatory network on the basis of the synthesis rate per cell evaluated from growth and fluorescence data. The a, α_{p_X}, δ_{p_X}, k_{p_X} and η_{p_X} parameters were fixed to the previously estimated values. The P_{TetR} promoter was induced with 100 ng/mL aTc while the P_{Lux} was induced with 1 μM 3OC$_6$-HSL. Results are shown in Fig. 3.12 and estimated values for the promoters activation rates are reported in Tab. 3.4.

As shown in Tab. 3.4, it appears that P_{TetR} transcriptional activation from the basal level to its maximum is faster than the one of P_{Lux}, although its estimate is also affected by a much higher CV. Promoter activation dynamics reflect the dynamic response of the final circuit in the production of 3OC$_6$-HSL upon induction with aTc (in the case of P_{TetR}) and in the lactonase production upon induction with 3OC$_6$-HSL (in the case of P_{Lux}). Since the P_{TetR} promoter in pSB3K3 and in pSB1A2 was used for subsystems characterization only at the steady state, its dynamics was not estimated.

Table 3.4: **Estimated rate constants for P_{Lux} and P_{TetR} activation.** Estimated values for r_{pTetr} in pSB4C5 and $r_{P_{Lux}}$ in pSB3K3 induced with 100 ng/mL aTc and 1μM 3OC$_6$-HSL respectively; parameters are shown with their respective CV in brackets.

Promoter rate constant	**Estimated value [min^{-1}]**
$r_{P_{TetR}}$	0.0824 (45.80%)
$r_{P_{Lux}}$	0.0551 (21.68%)

3.4 Enzymes parameters identification

The characterization of the two enzymes of interest was carried out using the subsystems presented in Sec. 2.6.3, where the expression of the two proteins was driven by the P_{TetR} promoter. This was necessary as the 3OC$_6$-HSL concentration must not influence the input of the system in a self regulating fashion as it could result by placing the P_{Lux} promoter upstream of the two coding sequences. This way it was possible to evaluate the only production or degradation of 3OC$_6$-HSL by the enzyme of interest, given a set of experimental conditions (such as the induction of the promoter or the cultures growth).

3.4.1 LuxI parameters identification

The parameters linked to the production of the signalling molecule by the LuxI enzyme are presented in Sec. 2.6.3.
In the model presented in Eq. 2.6, the k_{LuxI} parameter is a function of the aTc concentration, since it determines the intracellular level of enzyme. From these experiments, in which the cultures growth and the production of 3OC$_6$-HSL were monitored over time, the initial OD_{600} and 3OC$_6$-HSL, the growth rate of the cultures, μ, and the production rate $k_{LuxI}(aTc)$ were simultaneously estimated. Fig. 3.13 is an example in which the P_{TetR} promoter, driving the transcription of *luxI* in a medium copy number plasmid, was induced with 100 ng/mL aTc. As the culture grows in its induced state, the 3OC$_6$-HSL, measured by the biosensor (Sec. 2.7), increases, as no enzymatic degradation is present and the spontaneous decay of the signalling molecule can be neglected on this time scale. The $k_{LuxI}(aTc)$ parameter was estimated for different aTc concentrations (0, 0.5, 1, 2, 4, 8, 100 ng/mL) in low copy number plasmid while the only 100 ng/mL concentration was used to test the circuit put in the medium copy number plasmid.

Once the parameter linked to the production of 3OC$_6$-HSL was identified in several independent experiments, the subsequent step consisted in the estimation of the parameters that express the relation between per cell enzyme

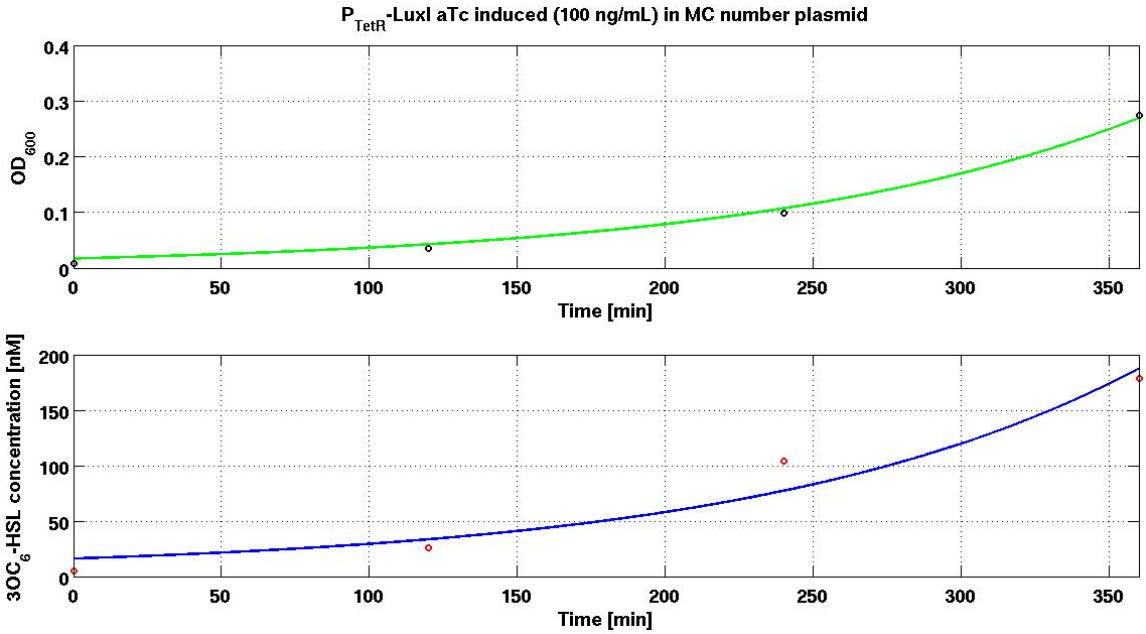


Figure 3.13: **Results for LuxI $3OC_6$-HSL production rate.** Growth and $3OC_6$-HSL production for culture bearing a medium copy number plasmid encoding for LuxI enzyme, induced with 100 ng/mL aTc; data were fitted using the model presented in Sec. 2.6.3.

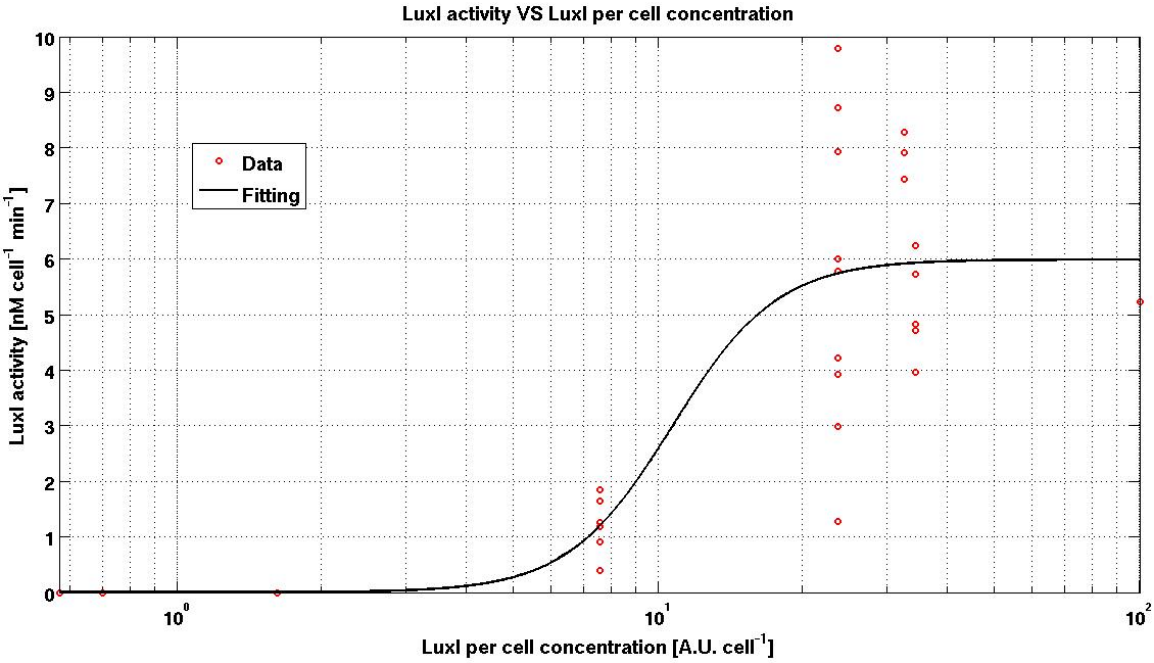


Figure 3.14: **LuxI activity as a function of LuxI per cell amount.** LuxI activity is presented as a function of LuxI amount per cell (reported in a $\log_{10}$ scale), which was estimated from P_{TetR} activation parameters at the steady state; the activity corresponding to the maximum amount per cell was estimated for BBa_B0030 as RBS and in a medium copy number plasmid.

level and the activity of LuxI, via a Hill function. In this case the enzyme amount was calculated on the basis of the parameters of P_{TetR} activation at the steady state and used to identify $k_{LuxI,max}$, $k_{M,LuxI}$ and η_{LuxI}. Results are shown in Fig. 3.14.

The estimated parameters are reported in Tab. 3.5 with their respective CVs. As expected from the lack of data for activities between the minimum and the maximum, η_{LuxI} resulted to be affected by great uncertainty. Although other data points at intermediate intracellular levels could help the identification procedure, such task was not performed. The main reason is that the intracellular levels of 7.5907 and 23.6736 A.U. correspond to aTc inductions of 2 and 4 ng/mL, which is a very narrow concentration window; other induction experiments spanning such window are not expected to improve fitting. The use of another inducible system could overcome such steep effect by providing estimated intracellular enzyme levels that weakly change as a function of induction.

Table 3.5: **Estimated parameters for LuxI activity.** Parameter values for $k_{LuxI,max}$, $k_{M,LuxI}$ and η_{LuxI}; estimated parameter CV is reported in brackets.

Parameter	Estimated value	Measurement Unit
$k_{LuxI,max}$	5.9875 (13.62%)	nM cell^{-1} min^{-1}
$k_{M,LuxI}$	10.7426 (39.14%)	A.U. cell^{-1}
η_{LuxI}	3.9812 (98.08%)	-

3.4.2 AiiA parameters identification

Similarly to LuxI characterization, cells expressing AiiA enzyme were induced with increasing aTc concentrations (0, 0.2, 0.5, 1, 2, 4, 8, 100 ng/mL) , as reported previously (Sec. 2.6.3). The $k_{AiiA}(aTc)$ parameter was estimated in cells bearing the AiiA expression system in the medium copy plasmid pSB3K3. The full induction condition (100 ng/mL) was also tested in the high copy number plasmid pSB1A2. The decay of exogenous $3OC_6$-HSL, added at the starting point of the experiment, as well as the cultures growth, were measured (Fig. 3.15). These data were used to simultaneously estimate the parameters of the model reported in Eq. 2.11, i.e., the initial OD_{600} and $3OC_6$-HSL, the growth rate of the cultures, μ, and the enzyme-dependent $3OC_6$-HSL degradation rate $k_{AiiA}(aTc)$. Then, with the obtained $k_{AiiA}(aTc)$, it was possible to identify the parameters of the Hill activation function of the AiiA enzyme (Eq. 2.12). In particular, as it was carried out for LuxI, the enzyme activity was plotted against the estimated intracellular concentration of AiiA, determined by the aTc concentration, to identify the $k_{AiiA,max}$, $k_{M,LuxI}$ and η_{LuxI} parameters. Their estimated values are reported in Tab. 3.6.

As it is evident in Fig. 3.16, data are noisy and the estimation procedure resulted hard to perform especially because the transition from the minimum to the maximum activity is determined for inductions that, in terms of concentration of aTc, are not so different. In particular, activity rises from zero to high values (0.2583 cell^{-1} min^{-1}) for estimated AiiA levels of 2.7614 to 15.6443 A.U. cell^{-1}, respectively. They correspond to aTc concentrations of 0.5 and 1 ng/mL, respectively, which are not so different. As reported previously for LuxI, which shows the same problem, the AiiA activity investigation was not carried out between these two concentrations. This steep behaviour could be overcome by using different inducible systems. The illustrated behaviour results in high uncertainty in the $k_{M,AiiA}$ and η_{AiiA} parameter identification, as reported in Tab. 3.6.

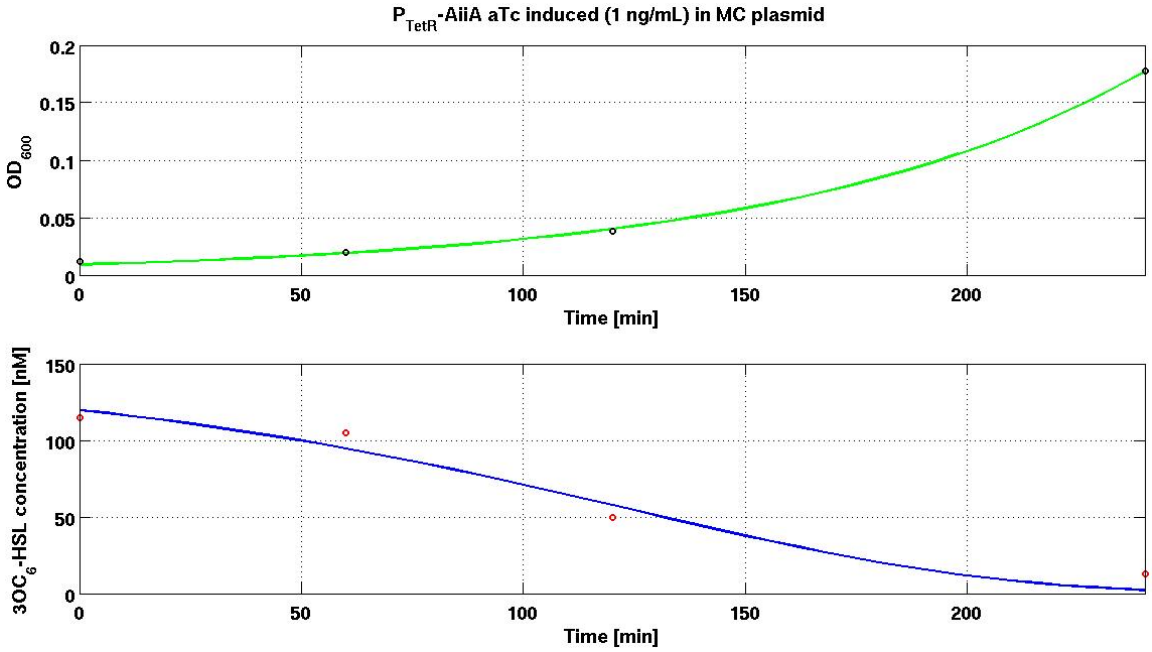


Figure 3.15: **Results for AiiA $3OC_6$-HSL degradation rate.** Growth and $3OC_6$-HSL degradation for cultures bearing the AiiA lactonase expression system inducible by P_{TetR}, induced with 1 ng/mL aTc; data were fitted using the model presented in Eq. 2.11.

Table 3.6: **Estimated parameters for AiiA activity.** Parameters value for $k_{AiiA,max}$, $k_{M,AiiA}$ and η_{AiiA}; parameters are reported as average value and percentage coefficient of variation.

Parameter	Estimated value	Measurement Unit
$k_{AiiA,max}$	0.4020 (10.81%)	$\text{cell}^{-1}\ \text{min}^{-1}$
$k_{M,AiiA}$	13.6334 ($\gg$ 100%)	A.U. cell^{-1}
η_{AiiA}	4.9716 ($\gg$ 100%)	-

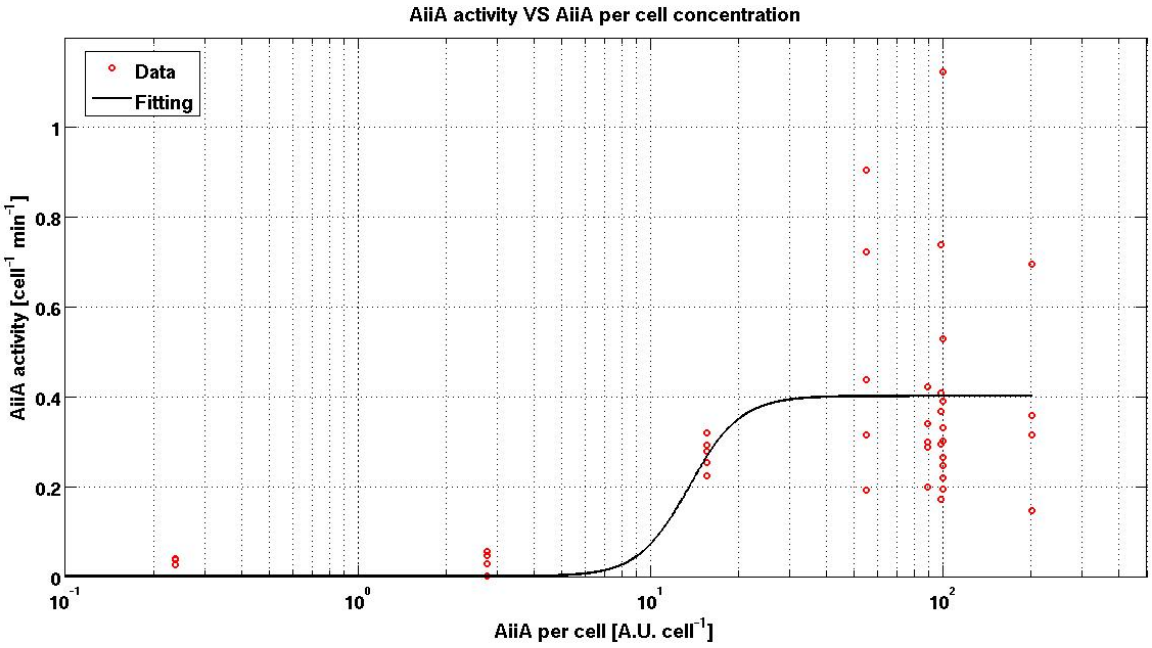


Figure 3.16: **AiiA activity as a function of the estimated intracellular level of enzyme.** The activity of the enzyme is reported to the amount of enzyme per cell (reported in $\log_{10}$ scale) as described in Sec. 2.6.3.

3.5 Uncertainty propagation results via Monte Carlo method

Since the identified parameters are affected by uncertainty, a Monte Carlo-based strategy was adopted in order to propagate this uncertainty towards the output of the proposed genetic controller. This process consists in estimating the distribution of model parameter values and then providing a distribution of the full circuit output that can be used to perform comparisons with experimental values. The description of the Monte Carlo approach used in this work is reported in Sec. 2.9. Here, only the results will be reported and commented.

The Monte Carlo method was applied to the estimates of the RFP maturation rate; in this case the simultaneous fitting of the 4 conditions considered (2 constitutive promoters and 2 inducible promoters, 3 replicates for each one, see Sec. 3.3.1), was carried out. 10^4 simulations were used to estimate the maturation rate and obtain its distribution which is reported in Fig. 3.17. Each run the RFP maturation rate was jointly estimated for all the 12 replicates, while static gain, A, and delay, τ, were separately assigned to each of the tested conditions. A Kolmogorov-Smirnov test was performed and the maturation rate of RFP resulted to be normally distributed around its mean value, which is 0.0157 min^{-1}, with a CV of 1.54%.

By propagating the uncertainty of RFP maturation rate, the static promoters functions were studied applying the Monte Carlo method in order to evaluate the variance of the Hill equation parameters. As regards the P_{TetR} promoter, the distribution of its parameters is shown in Fig. 3.18. Performing Kolmogorov-Smirnov test on each parameter, it resulted that only $\alpha_{P_{TetR}}$ is normally distributed, while the others are not; in particular, $\delta_{P_{TetR}}$ shows a peak near zero as the leakage of this promoter can be considered negligible.

As reported in Tab. 3.7 parameter estimates are close to the ones resulting from the *lsqnonlin* routine (reported in Tab. 3.3), even though CVs are slightly different.

Table 3.7: P_{TetR} **promoter parameters in low copy number plasmid average and percentage coefficient of variation.** For each of the parameters of the input-output curve of P_{TetR} promoter in low copy number plasmid, the average and the percentage coefficient of variation is reported after applying Monte Carlo method.

$\alpha_{P_{TetR}}$	$\delta_{P_{TetR}}$	$k_{P_{TetR}}$	$\eta_{P_{TetR}}$
1.0274 (3.28)	0.0193 (95.92)	3.1311 (8.70)	3.1858 (22.68)

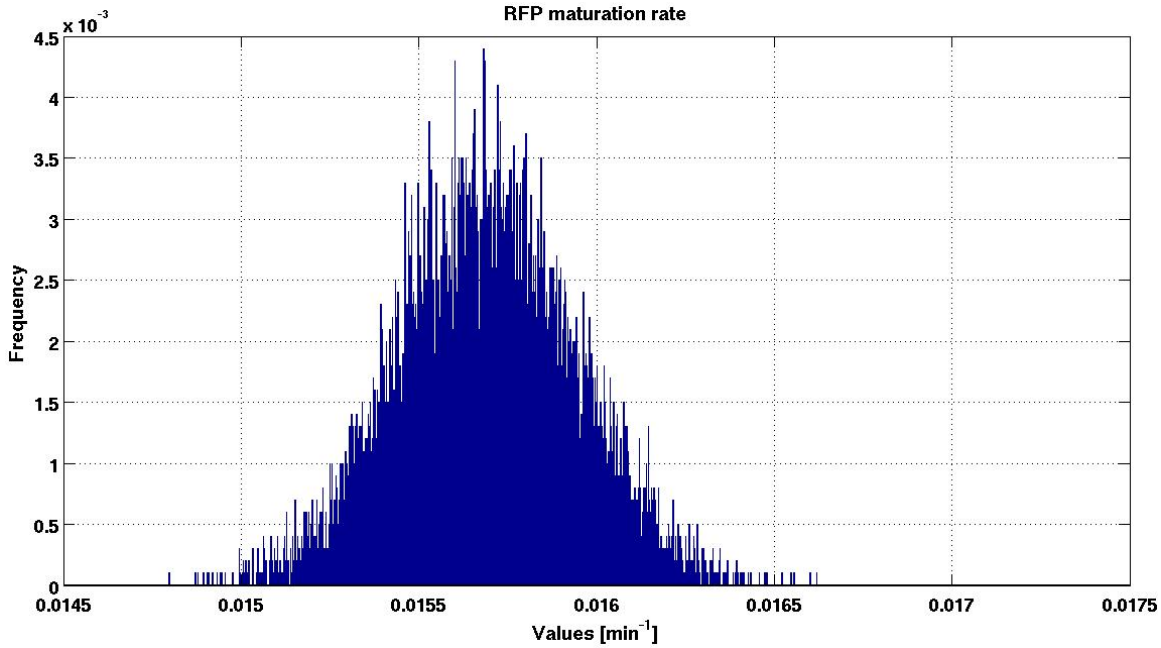


Figure 3.17: **RFP maturation rate distribution.** RFP maturation rate distribution obtained generating 10^4 samples through Monte Carlo method.

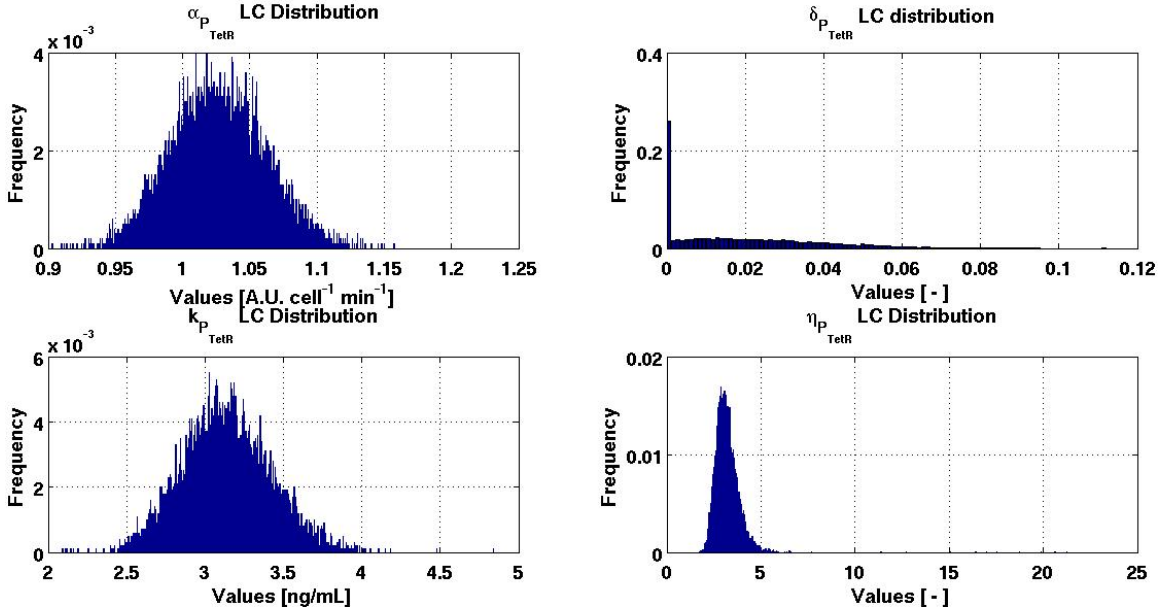


Figure 3.18: P_{TetR} **parameters distributions in low copy number plasmid after applying Monte Carlo method.** P_{TetR} promoter parameters distributions in low copy number plasmid are reported: each histogram was obtained from 10^4 samples.

Finally, a plot of the results of the Monte Carlo algorithm against experimental measurements is given (Fig. 3.19). In some cases, as for the highest inductions, data are very noisy and, in some cases, as for the highest inductions, they are out of the evaluated 95% confidence intervals.

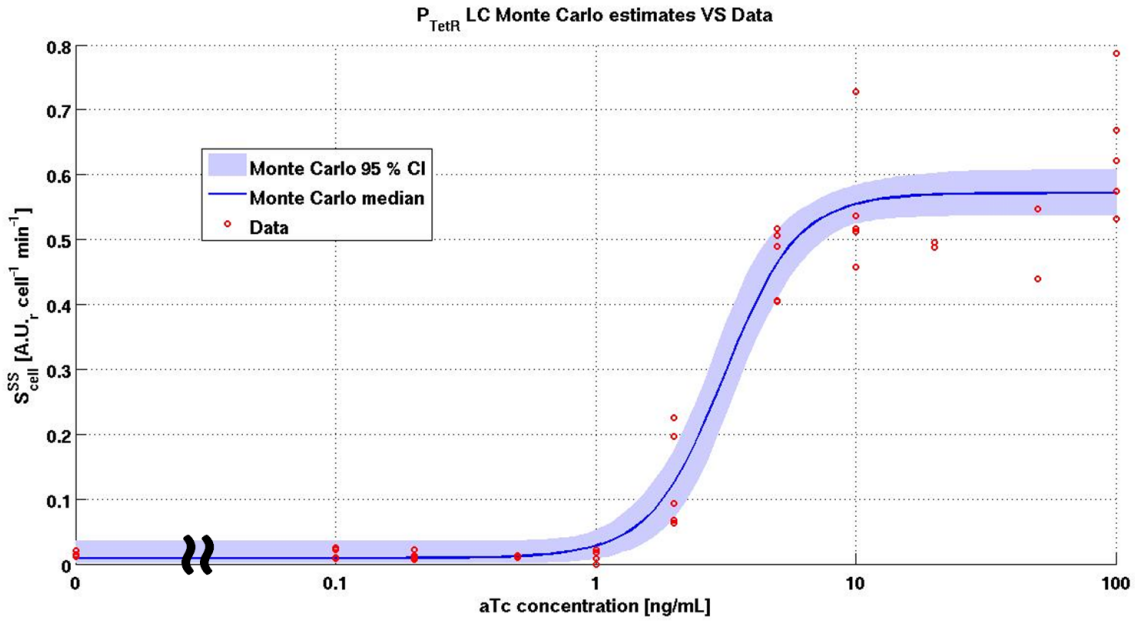


Figure 3.19: P_{TetR} **in low copy number plasmid input-output curve estimated with Monte Carlo algorithm against data.** P_{TetR} promoter input-output curve obtained from 10^4 samples; blue line represents the median of the output evaluated with Monte Carlo strategy, while light blue areas represent its 95% confidence intervals.

The P_{TetR} promoter was also tested in the medium copy number plasmid pSB3K3 and the same methodology was used to propagate the uncertainty of the parameters of its Hill activation function. In this case, only $\alpha_{P_{TetR}}$ resulted to be normally distributed (Fig. 3.20), according to the performed Kolmogorov-Smirnov test. The values and CVs of these parameters are reported in Tab. 3.8: while $\alpha_{P_{TetR}}$ and $k_{P_{TetR}}$ are similar to the ones reported in Sec. 3.3.1, $\delta_{P_{TetR}}$ and $\eta_{P_{TetR}}$ show very different CVs. In particular the CV of $\delta_{P_{TetR}}$ decreases while the one of $\eta_{P_{TetR}}$ increases; because of this effect, predictions of Monte Carlo approach are the ones depicted in Fig. 3.21.

Finally, the same strategy was applied for the P_{Lux} promoter. As shown in Fig. 3.22, similarly to P_{TetR}, the leakage, namely $\delta_{P_{Lux}}$ parameter, has a peak near zero; none of the parameters is normally distributed according to the Kolmogorov-Smirnov test. Average values and CVs are reported in Tab. 3.9,

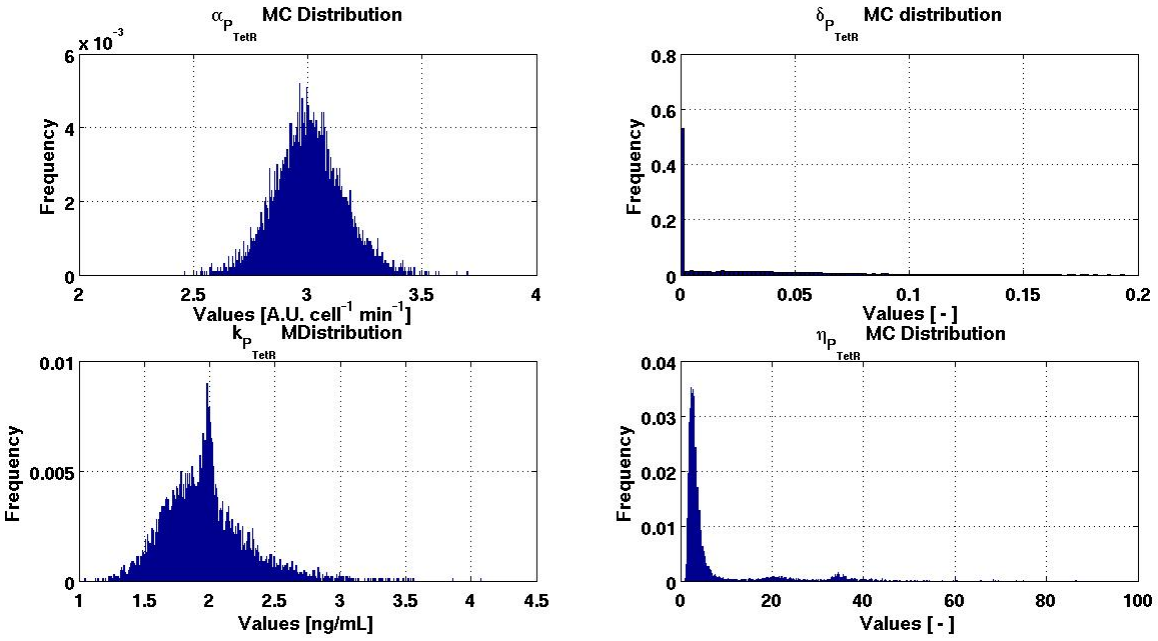


Figure 3.20: P_{TetR} **parameters distributions in medium copy number plasmid after applying Monte Carlo method.** P_{TetR} promoter parameters distributions in medium copy number plasmid are reported: each histogram was obtained from 10^4 samples.

Table 3.8: P_{TetR} **promoter parameters in medium copy number plasmid average and percentage coefficient of variation.** For each of the parameters of the input-output curve of P_{TetR} promoter in medium copy number plasmid, the average and the percentage coefficient of variation is reported after applying Monte Carlo method.

$\alpha_{P_{TetR}}$	$\delta_{P_{TetR}}$	$k_{P_{TetR}}$	$\eta_{P_{TetR}}$
3.0032 (4.63)	0.02 (>100)	1.9531 (15.43)	5.4960 (>100)

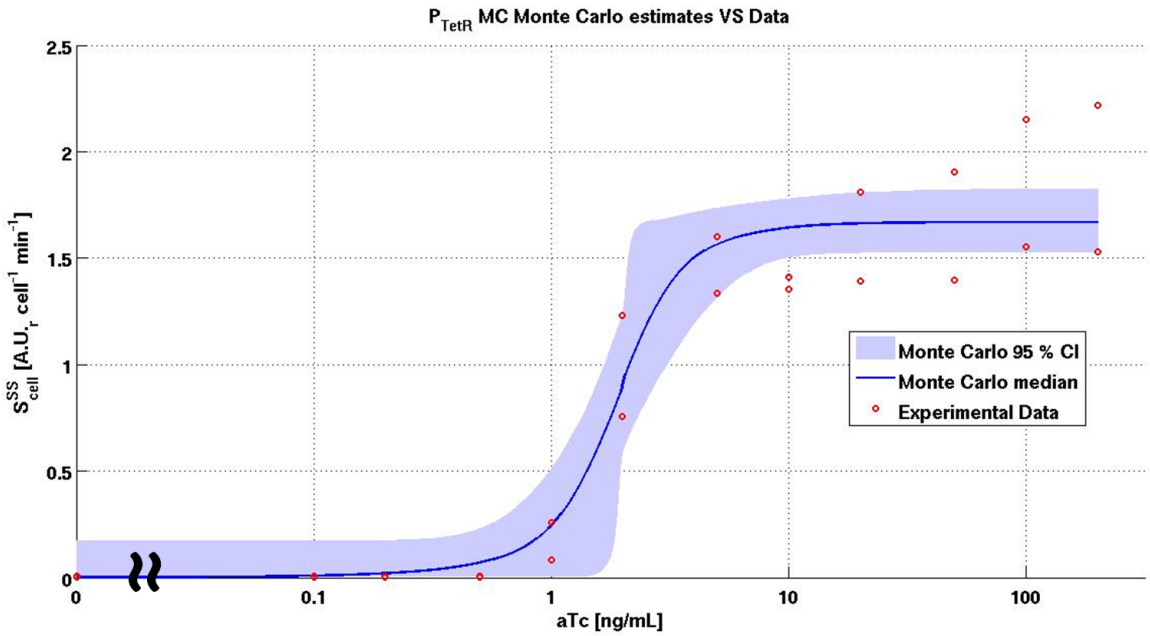


Figure 3.21: P_{TetR} **in medium copy number plasmid input-output curve estimated with Monte Carlo algorithm against data.** P_{TetR} promoter input-output curve obtained from 10^4 samples; blue line represents the median of the output evaluated with Monte Carlo strategy, while light blue areas represent its 95% confidence intervals.

while the Monte Carlo 95% confidence intervals of the output are reported together with data in Fig. 3.23.

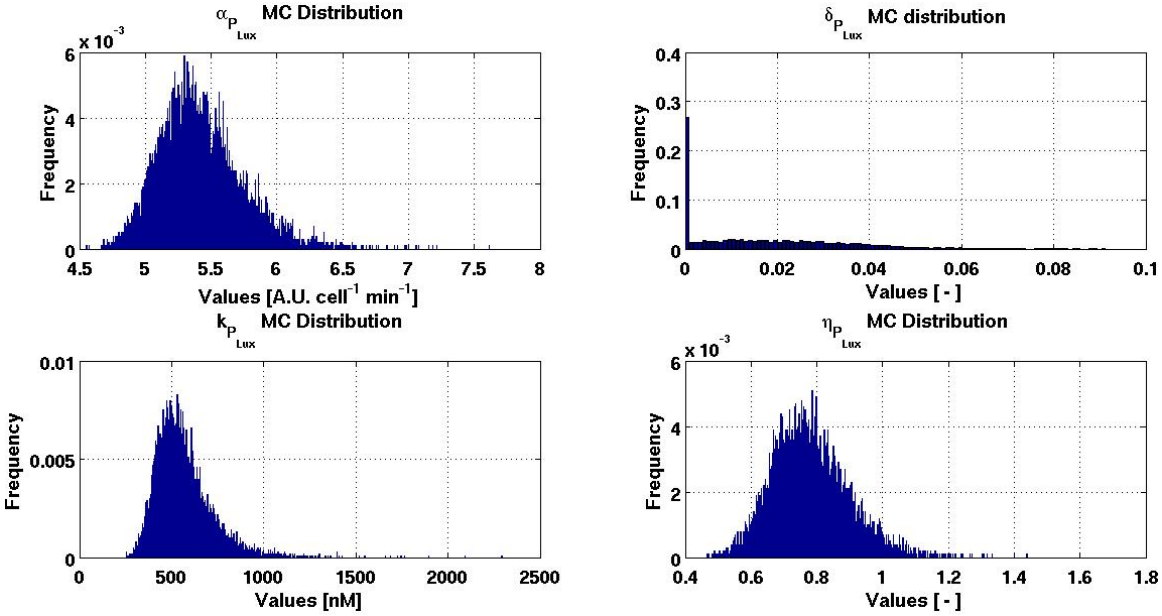


Figure 3.22: P_{Lux} **parameters distributions after applying Monte Carlo method.** P_{Lux} promoter parameters distributions are reported: each histogram was obtained from 10^4 samples.

Table 3.9: P_{Lux} **promoter parameters average and percentage coefficient of variation.** For each of the parameters of the input-output curve of P_{Lux} promoter the average and the percentage coefficient of variation is reported after applying Monte Carlo method.

$\alpha_{P_{Lux}}$	$\delta_{P_{Lux}}$	$k_{P_{Lux}}$	$\eta_{P_{Lux}}$
5.4136 (5.72)	0.0173 (95.76)	559.3358 (26.93)	0.7778 (14.25)

Also in this case Monte Carlo estimates were compared to the *lsqnonlin* routine: parameter values were similar but coefficients of variation were different. In particular, the estimated CV of $\alpha_{P_{Lux}}$, $k_{P_{Lux}}$ and $\eta_{P_{Lux}}$ increase while, conversely, the CV of $\delta_{P_{Lux}}$ decreases compared to the values obtained without the Monte Carlo method (compare with values in Tab. 3.3). As before, in Fig. 3.23, Monte Carlo predictions are compared to the experimental data.

Subsequently, the promoters dynamic activation was considered and investigated using the Monte Carlo method, generating 10^4 samples for each of the

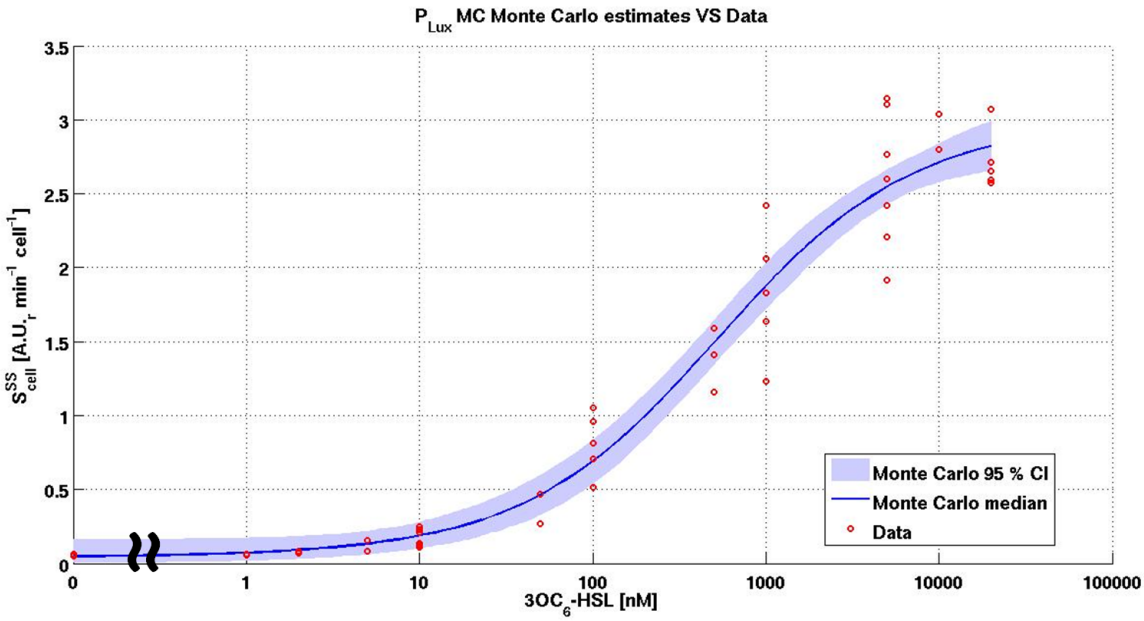


Figure 3.23: P_{Lux} **input-output curve estimated with Monte Carlo algorithm against data.** P_{Lux} promoter input-output curve obtained from 10^4 samples; blue line represents the median of the output evaluated with Monte Carlo strategy, while light blue areas represent its 95% confidence intervals.

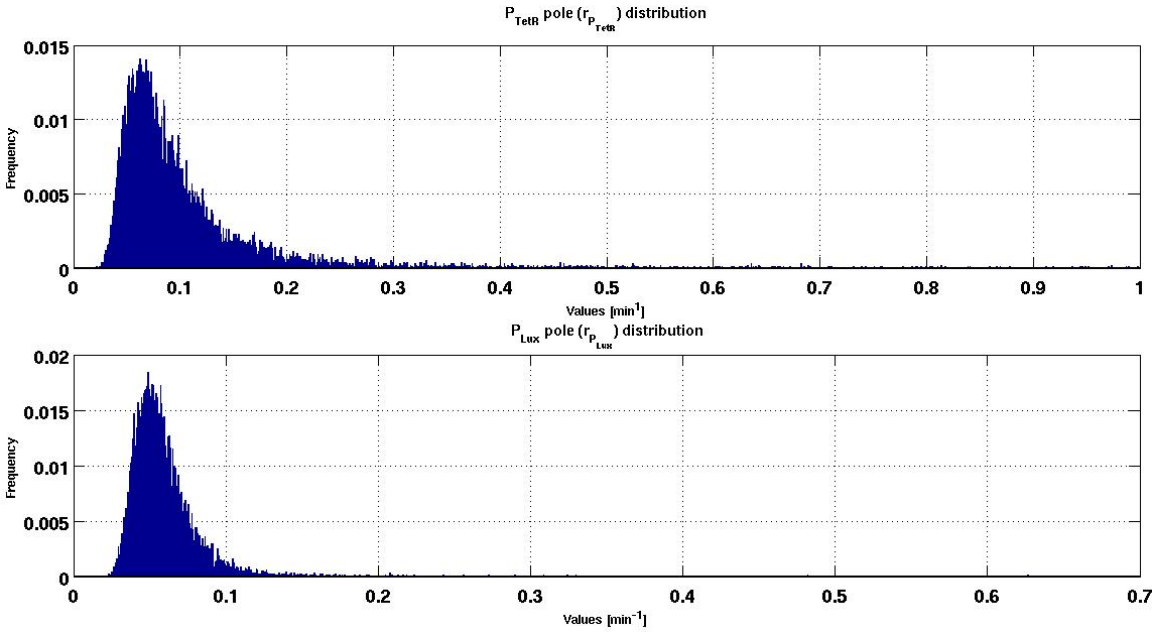


Figure 3.24: **Poles distribution for dynamic promoters activation.** Distributions of $r_{P_{TetR}}$ and $r_{P_{Lux}}$, obtained generating 10^4 samples through Monte Carlo method.

two promoters. The promoters parameters and the RFP maturation rate used to estimate the values of the poles $r_{P_{TetR}}$ and $r_{P_{Lux}}$ were sampled from the previously shown distributions in order to propagate the uncertainty of the parameters estimation from the previous steps. The differential equation system shown in Eq. 2.3 was identified in order to obtain the poles distributions. As considering $r_{P_{TetR}}$ and $r_{P_{Lux}}$ greater than 1 does not affect the activation dynamics of the two promoters (data not shown), this value was set as the upper limit. In this way, the distributions shown in Fig. 3.24 were obtained, where $r_{P_{Lux}}$ has a narrower distribution than $r_{P_{TetR}}$. The mean values and their CVs are reported in Tab. 3.10. Parameter values are slightly higher, as well as their CVs, compared to the estimation without the Monte Carlo method (compare to values in Tab. 3.4).

Median and 95% confidence intervals obtained with the Monte Carlo strategy were reported in Fig. 3.25 to compare the results of Monte Carlo simulations and data. In both cases the 95% confidence interval is larger, i.e., the uncertainty is higher, in the transient than at steady state, due to the uncertainty affecting the poles value together with the uncertainty affecting the maturation rate of RFP.

The attention was then focused on the enzymes parameters which were estimated using the Monte Carlo method again, by sampling the promoters

Table 3.10: **P_{Lux} and P_{TetR} poles mean and coefficients of variation.** $r_{P_{TetR}}$ and $r_{P_{Lux}}$ average and coefficients of variation obtained from 10^4 Monte Carlo simulations.

Promoter pole	Estimated value [min^{-1}]
$r_{P_{TetR}}$	0.1057 (85.19%)
$r_{P_{Lux}}$	0.0586 (37.29%)

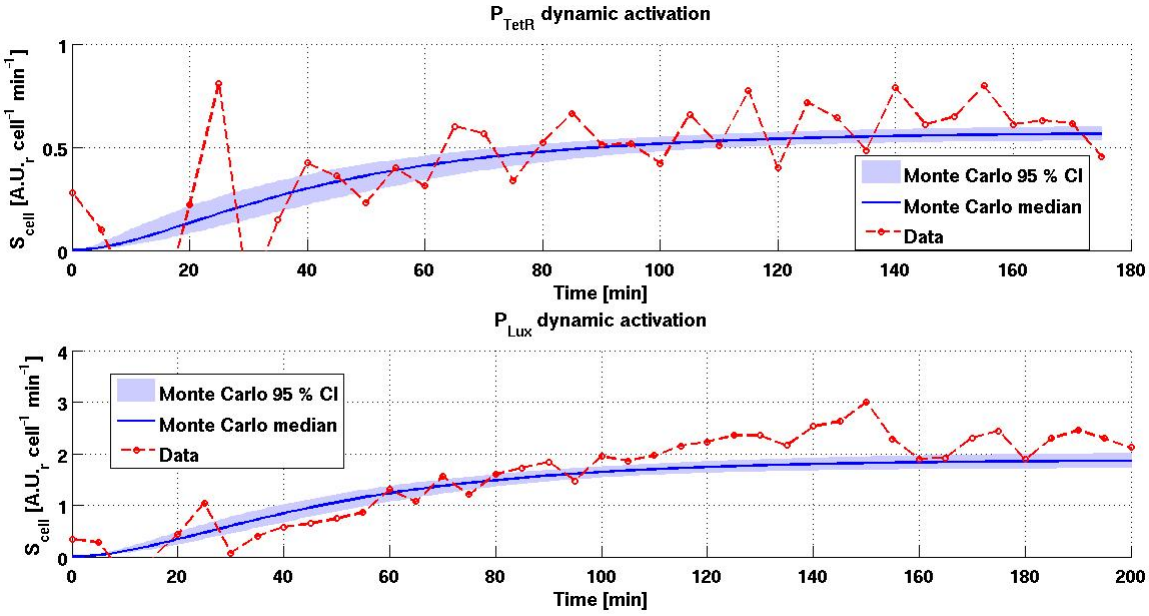


Figure 3.25: **Dynamic promoters activation: Monte Carlo predictions against data.** Monte Carlo 95% confidence intervals (light blue area) and median (blue line), obtained from 10^4 simulations, are reported together with experimental data of promoters activation dynamics.

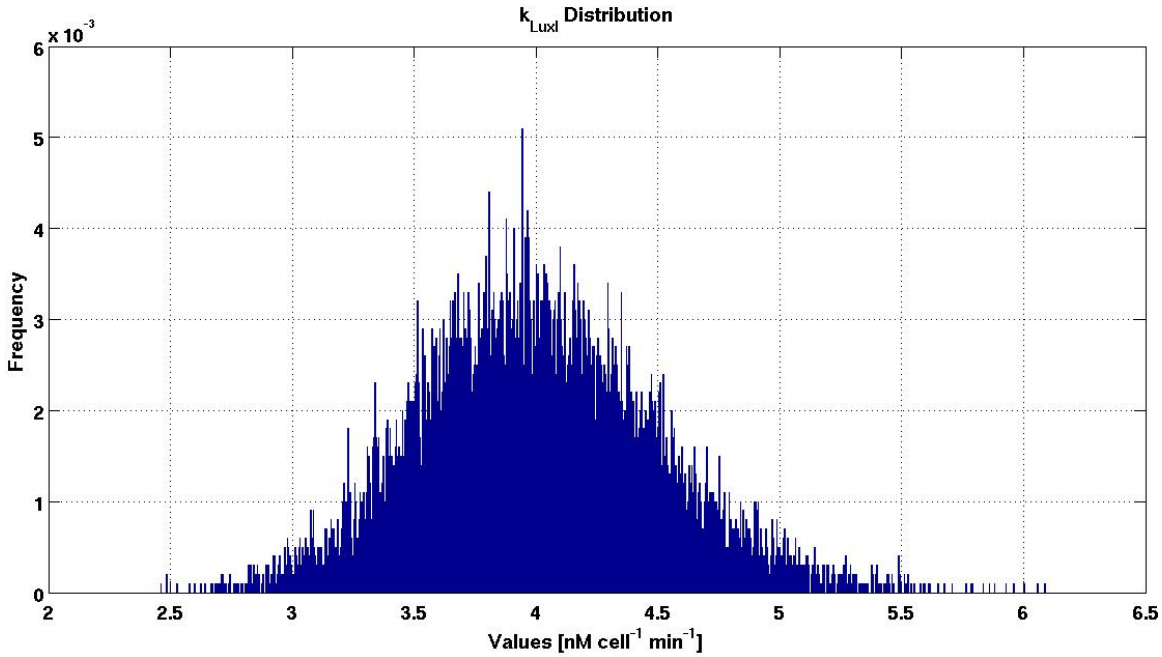


Figure 3.26: **Example of $k_{LuxI}(aTc)$ distribution.** Distribution of $k_{LuxI}(aTc)$ in cultures bearing P_{TetR} driven *luxI* in a low copy number plasmid, induced with 100 ng/mL aTc.

parameters from the previously shown distributions; in fact, promoters parameters are used to evaluate the concentration of enzyme per cell at the steady state (see Eq. 2.10 and Eq. 2.13). LuxI parameters were evaluated from 10^4 samples generated through the Monte Carlo method; each performed experiment was used to generate a distribution of the parameters of the model shown in Eq. 2.6. The distribution of $k_{LuxI}(aTc)$ is particularly important as it was used to evaluate, in the next step of uncertainty propagation, the parameters of enzyme activity as a function of the per cell level of enzyme: $k_{LuxI,max}$, $k_{M,LuxI}$ and η_{LuxI}. In particular, each individual LuxI assay generated a distribution for $k_{LuxI}(aTc)$ according to the Monte Carlo method performed on OD_{600} and $3OC_6$-HSL sampled data; then, such distributions were used in another Monte Carlo method step to estimate AiiA parameters as a function of the estimated intracellular level of enzyme. Fig. 3.26 shows an example of $k_{LuxI}(aTc)$ distribution, obtained in a single experiment where the culture bearing the subsystem with *luxI* was induced with 100 ng/mL of aTc.

While for high (4 to 100 ng/mL) aTc inductions the value of $k_{LuxI}(aTc)$ parameter appears to be normally distributed, for low aTc inductions (ranging from 0 to 2 ng/mL) the enzyme activity is null or very low and Monte Carlo estimates show a distribution with a peak near zero (data not shown). More-

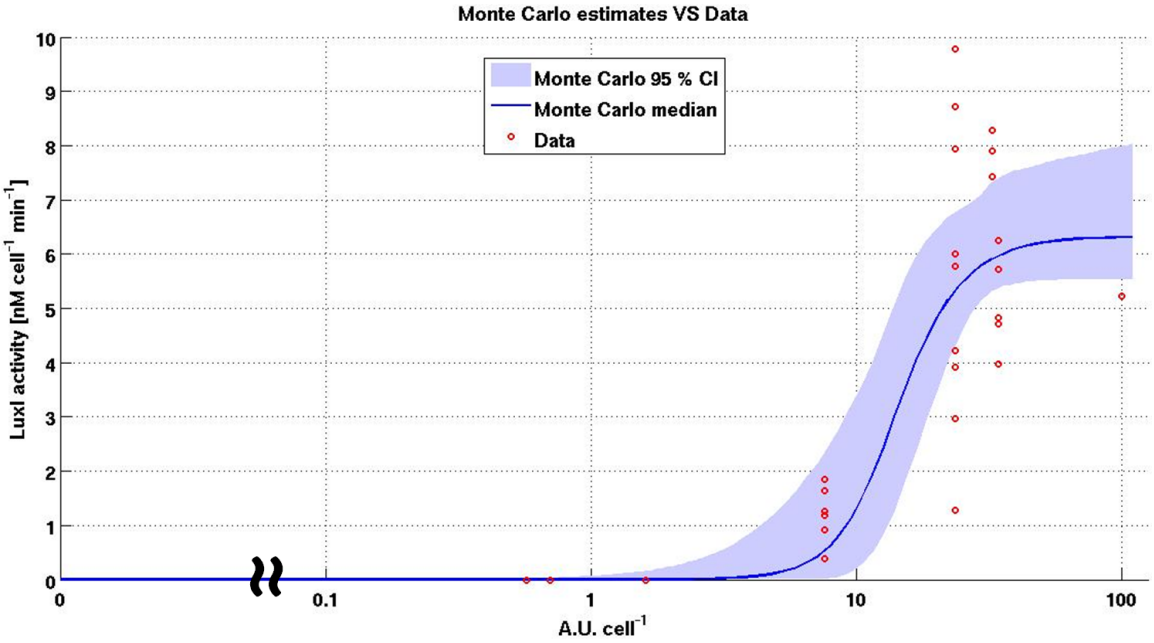


Figure 3.27: **Monte Carlo predictions and data for LuxI enzyme activity.** Median and 95% confidence intervals of enzyme activity from 10^4 Monte Carlo simulations compared to experimental data.

over, in order to obtain biologically plausible values for k_{LuxI}, its maximum was set to 10 nM cell^{-1} min^{-1}; this also prevented from obtaining parameter values that give integration problems in the *ode15s* routine. The activity of the enzyme, in terms of Hill function parameters (given their intracellular level), was then considered, sampling from the distribution of promoter Hill function parameters to estimate the per cell enzyme concentration and sampling from the previously described enzyme activity distributions. Again, 10^4 samples were generated through Monte Carlo method and the model parameters presented in Eq. 2.7 were identified. On the basis of these simulations, the 95% confidence intervals of the function that describes the enzyme activity in relation to its per cell level is shown in Fig. 3.27. The variability that affects $k_{LuxI,max}$ and $k_{M,LuxI}$ parameters, along with the uncertainty of the per cell enzyme concentration, results in a broad distribution of the output, especially for values around the half activation, even if experimental data points showed a broader distribution.

The identified parameters and their respective CVs are reported in Tab. 3.11. Estimates are not so different from the ones reported in Tab. 3.5, even though $k_{LuxI,max}$ and $k_{M,LuxI}$ slightly increase their values. CVs are different from the previously evaluated ones, especially for η_{LuxI} which is now 50% of the one

Table 3.11: **Parameters for LuxI enzyme activity estimated with Monte Carlo method.** After performing 10^4 simulations sampling from the estimated LuxI intracellular level and LuxI activity distributions previously obtained, average parameters values and their CVs were estimated.

Parameter	Estimated value	Measurement Unit
$k_{LuxI,max}$	6.4584 (11.34%)	nM cell^{-1} min^{-1}
$k_{M,LuxI}$	14.6515 (19.59%)	A.U. cell^{-1}
η_{LuxI}	4.1011 (41.71%)	-

reported in Tab. 3.5.

In the same way, the uncertainty of the parameters was propagated to the activity of AiiA, sampling from the distribution of the per cell enzyme level (given the aTc induction) and the activity of the lactonase. Monte Carlo method was initially applied to the independent experiments regarding $3OC_6$-HSL decay in liquid cultures, in order to obtain distributions of the $k_{AiiA}(aTc)$ parameter, as shown in the example of Fig. 3.28.

Also in this case enzyme intracellular concentration and enzyme activity were sampled 10^4 times from previously evaluated distributions and parameters estimated. In Tab. 3.12 parameters are reported with their CVs and the relation between the per cell enzyme level and the enzyme activity is depicted in Fig. 3.29. As $k_{M,AiiA}$ and η_{AiiA} are affected by a greater uncertainty than $k_{AiiA,max}$ the distribution of the output in the proximity of the half activity is clearly broader than for enzyme concentration at saturation. In this case parameter values are not so different from the ones reported in Tab. 3.6, but the CVs of $k_{M,AiiA}$ and η_{AiiA} are much smaller.

P_{TetR} in low copy plasmid and RBS BBa_B0030 and P_{TetR} in medium copy number plasmid with BBa_B0034 were used to tune the expression of *luxI* and *aiiA* genes respectively; during uncertainty propagation, their parameters were sampled from the previously obtained distributions for different aTc concentrations. When *luxI* and *aiiA* were expressed in pSB3K3 medium copy number and pSB1A2 high copy number plasmids respectively, the P_{TetR} promoter parameters were sampled from the distributions of the only full induction conditions (100 ng/mL aTc) that were previously tested with RFP.

Once the uncertainty was propagated, the distributions of the parameters were used to evaluate the output of the whole system. Again, parameters were sampled from their own distributions and used to predict the $3OC_6$-HSL concentration of the open- and close-loop. When performing these simulations, three parameters were set to a fixed value: the growth rate of cells, μ,

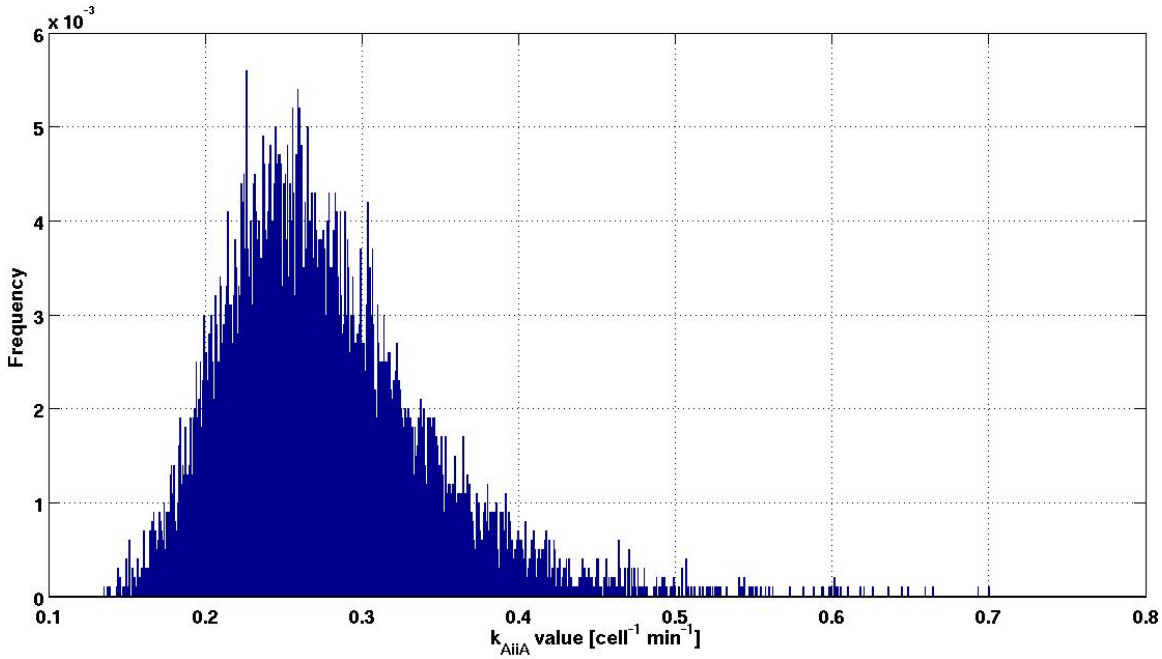


Figure 3.28: **Example of $k_{AiiA}(aTc)$ distribution.** Distribution of $k_{AiiA}(aTc)$ in cultures bearing P_{TetR} driven *aiiA* in a medium copy number plasmid, induced with 100 ng/mL aTc.

Table 3.12: **Parameters for AiiA enzyme activity estimated with Monte Carlo method.** After performing 10^4 simulations sampling from the estimated AiiA intracellular level and AiiA activity distributions previously obtained, average parameter values and their CVs were estimated.

Parameter	Estimated value	Measurement Unit
$k_{AiiA,max}$	0.4266 (8.93%)	cell^{-1} min^{-1}
$k_{M,AiiA}$	11.7440 (43.70%)	A.U. cell^{-1}
η_{AiiA}	3.0296 (64.73%)	-

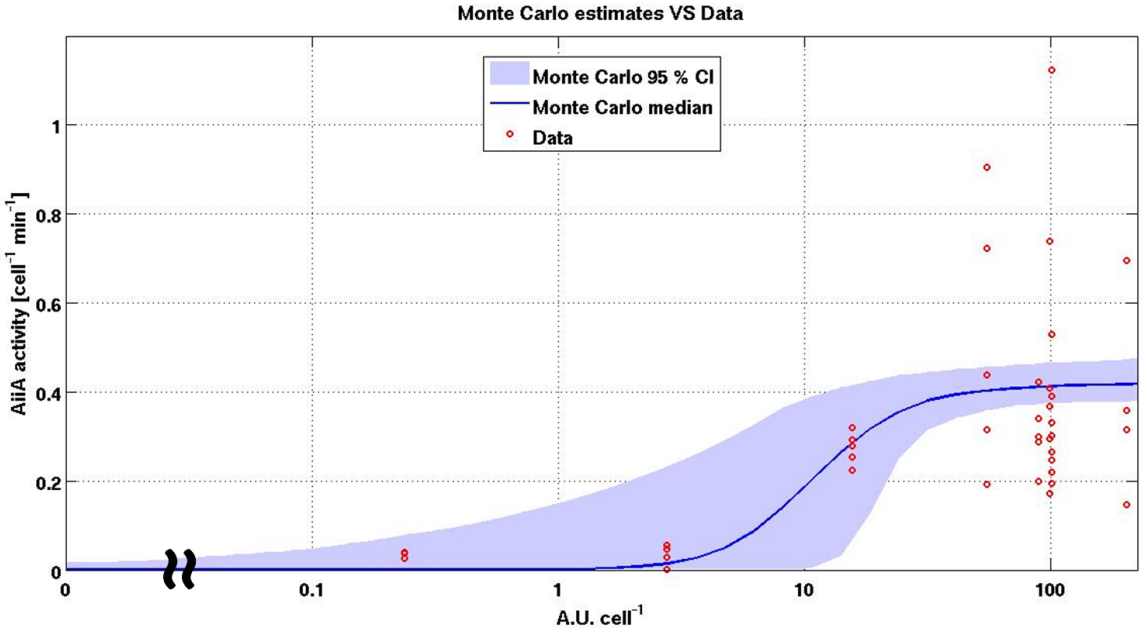


Figure 3.29: **Monte Carlo predictions and data for AiiA enzyme activity.** Median and 95% confidence intervals of AiiA enzyme activity from 10^4 Monte Carlo simulations compared to experimental data.

which was evaluated from growth data and set to 0.0125 min^{-1}, the $3OC_6$-HSL spontaneous degradation rate,γ_{HSL} which was evaluated in *ad-hoc* designed experiments (see App. A) and set to 2.2470 e-04 min^{-1}, and the enzyme degradation due to the presence of an LVA tag, γ_{LVA}, which was set to 0.0173 min^{-1}, namely, the value presented in [49]. Simulations were run and confidence intervals of the genetic circuit output were given and compared to the experimental measurements.

3.6 Simulations and experimental results

3.6.1 Results for growing liquid cultures in batch mode

The model of Eq. 2.1 is expected to provide reliable predictions only in the exponential growth phase of recombinant bacteria bearing the close-loop regulator circuit, because:

- the experimental measurements performed in the identification steps on the subsystems were all performed in that phase;
- promoters in stationary phase can have unpredictable activities;
- the hypothesis of intracellular species dilution due to cell growth is valid when bacteria duplicate at the constant rate μ, i.e., only in the exponential growth phase.

In order to perform long experiments for the close-loop regulator, exponential phase must be maintained and, for this reason, the system was tested in chemostatic condition. However, before presenting the experimental results in such condition and the related model predictions, the system was preliminarily tested in batch condition in a 15-mL tube (5 mL culture volume) and compared to the open-loop configuration. Results are reported in Fig. 3.30. Even though growth curves are comparable, the two circuits show a different behaviour, with the open-loop circuit reaching a higher $3OC_6$-HSL value than the close-loop. This is qualitatively in accordance with the expected behaviour of the circuits, since $3OC_6$-HSL should be feedback-regulated in the close-loop circuit by AiiA.

3.6.2 Results in chemostat

A chemostat (Sec. 2.8) was realized and used to study the genetic controller and its non regulated counterpart at several constant OD_{600} values. Using this experimental set-up, cells were maintained around a target OD_{600} (Fig. 3.31) during the time of the measurements (generally between 4.5 and 6.5 h). Both

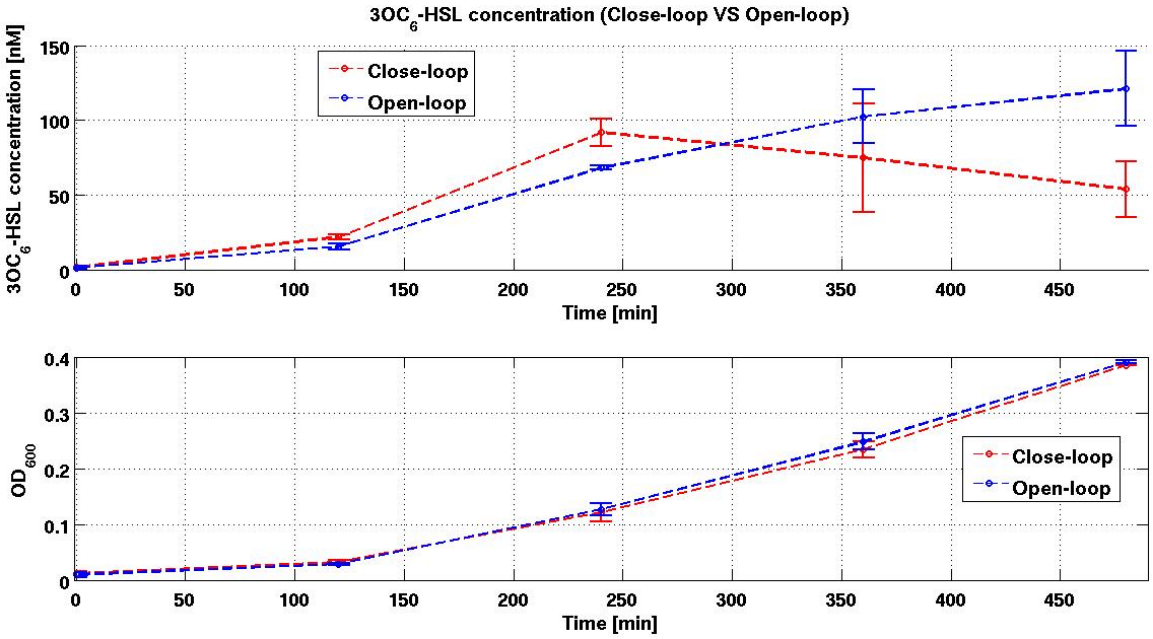


Figure 3.30: **Regulated VS non-regulated $3OC_6$-HSL production in batch cultures.** Comparison between close (red dashed line) and open-loop (blue dashed line) behaviour in growing liquid cultures induced with 100 ng/mL aTc; data points represent the average of at least two independent replicates with their 95% confidence intervals. In the two graphs both the $3OC_6$-HSL concentration and the cultures growth are reported.

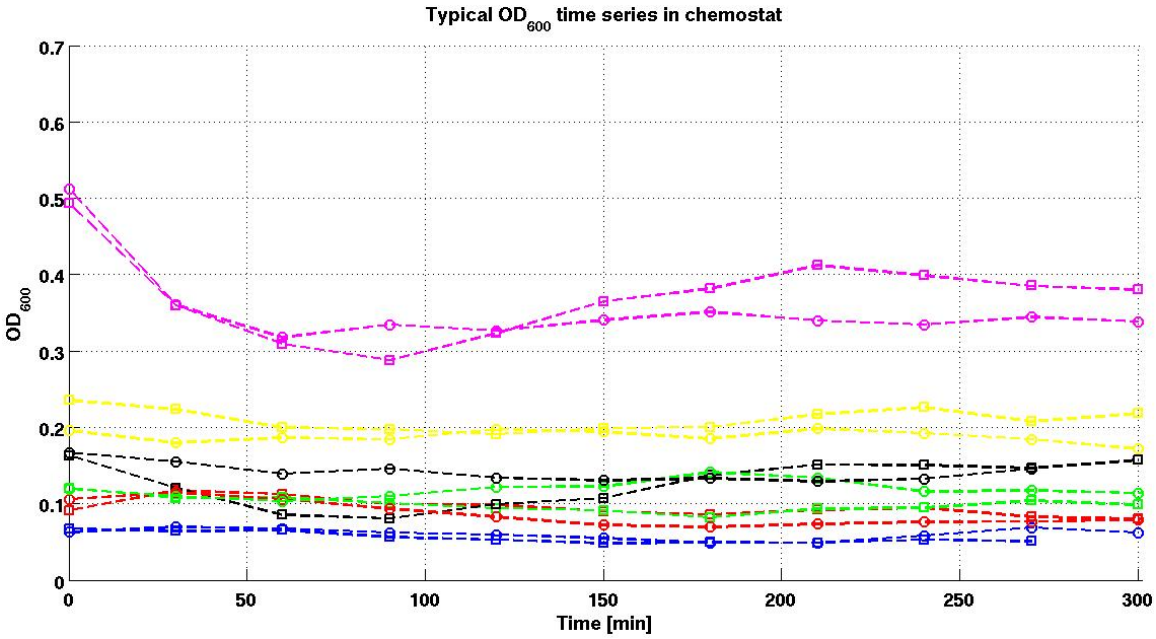


Figure 3.31: OD_{600} **over time in chemostat mode.** Time series of typical chemostat OD_{600}; in this experimental set-up culture density is maintained as constant as possible at a fixed OD_{600} by properly diluting cultures as explained in Sec. 2.8. Circles represent open-loop experiments while squares were used for close-loop experiments; similar colors were used for similar OD_{600} values.

the open- and the close-loop circuits were considered in order to verify if differences in $3OC_6$-HSL control occur and if such differences were quantitatively predicted by the developed mathematical model. In these experiments OD_{600} and $3OC_6$-HSL were measured, sampling from the bioreactor every 30 min; while OD_{600} was measured with Tecan Infinite F200, the signalling molecule was quantified using a biosensor, as described in Sec. 2.7.

The genetic open-loop circuit was investigated at the OD_{600}: 0.06, 0.09, 0.12, 0.21 and 0.38. From the model simulations, it was expected that the steady state concentration of the signalling molecule in the extracellular environment would increase proportionally to the optical density of the culture as in Fig. 3.3. Fig. 3.32 (panel A) shows the experimental data of steady state $3OC_6$-HSL in open-loop configuration as a function of bacterial density in chemostat and predicted values are also shown.

The expected behaviour was observed despite, at OD_{600} of 0.2, the model prediction underestimates the circuit output against the experimental value of 142 nM (see Fig. 3.32). Importantly, OD_{600} values higher than 0.3 were never tested in the subsystems characterization stage, since the exponential phase in 15-mL tubes ends up at a lower density than in the bioreactor (data not

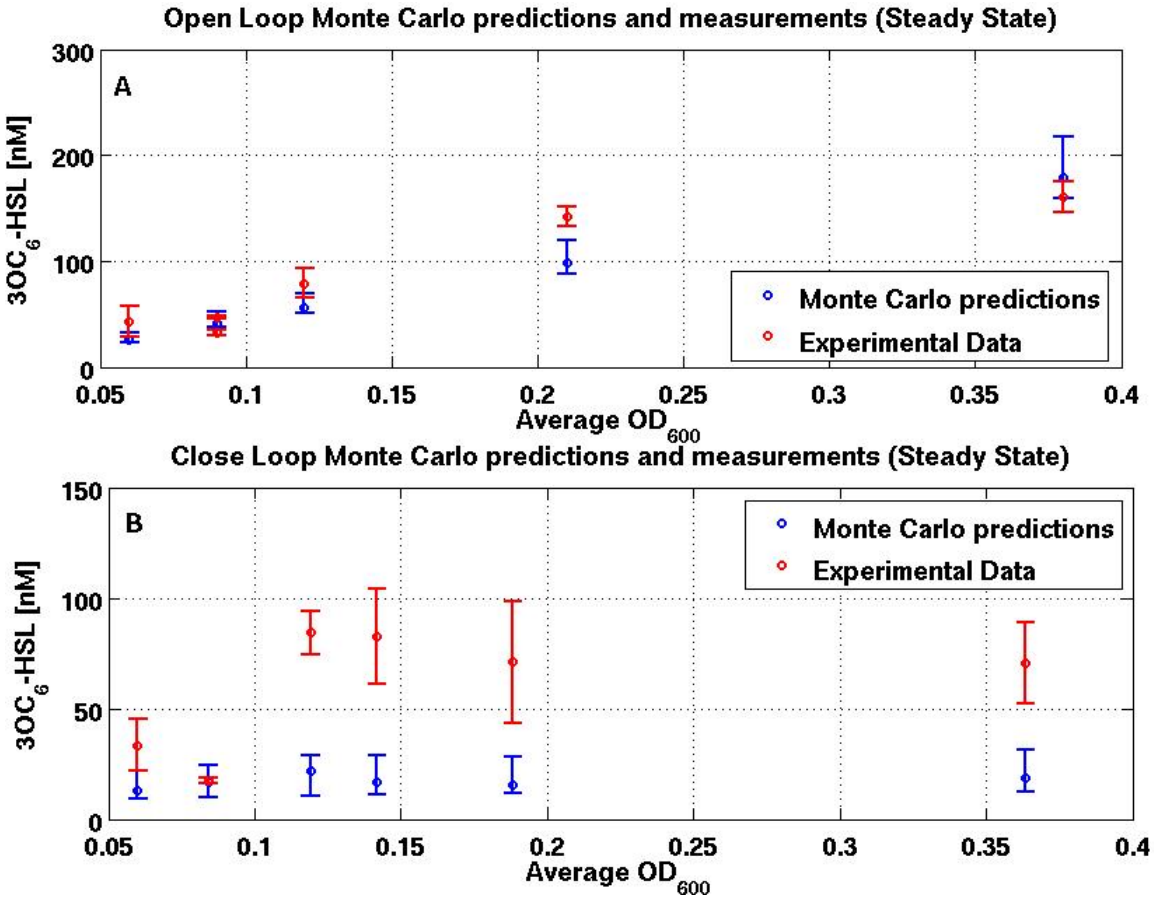


Figure 3.32: **Comparison between Monte Carlo predictions and experiments.** Data were compared to the predictions calculated with the Monte Carlo method: both, open and close-loop are depicted. For Monte Carlo simulations the average value and the intervals were evaluated on the distribution of the $3OC_6$-HSL steady state estimated concentration while for experimental data the average value and confidence intervals were evaluated on the basis of the last three measurements of each experiment (assumed to represent the steady state value).

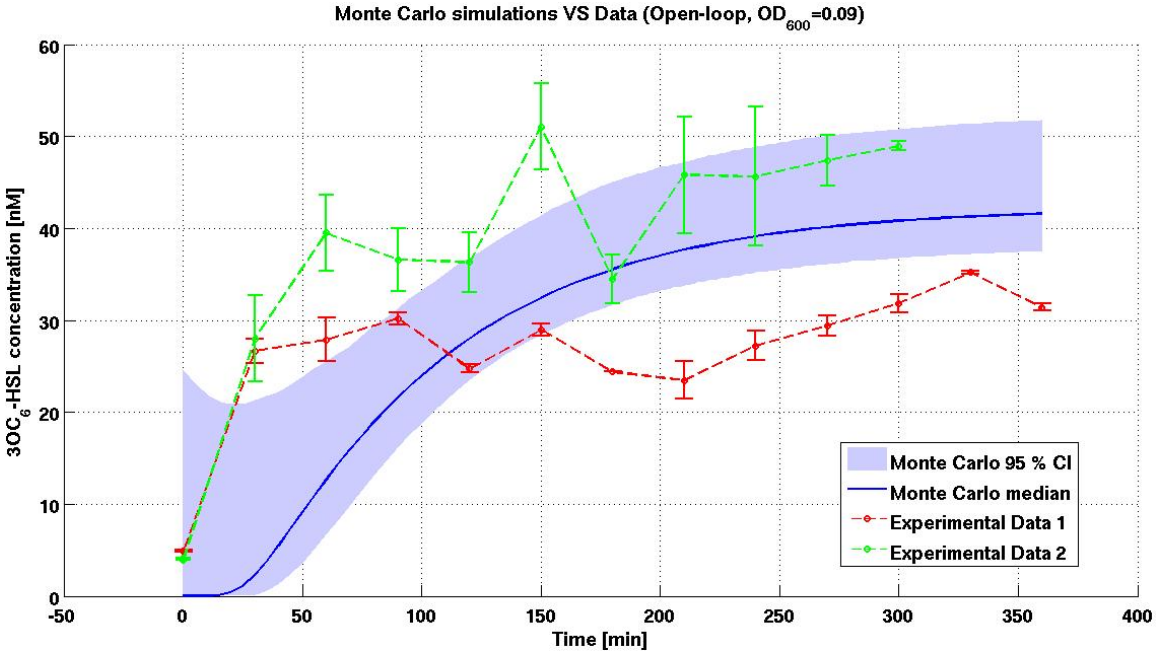


Figure 3.33: **Open-loop $3OC_6$-HSL concentration over time in chemostat experiments at OD_{600} 0.09.** Comparison between open-loop behaviour in chemostat at $OD_{600} = 0.09$ induced with 100 ng/mL aTc and model predictions based on Monte Carlo method; data points represent the average of two measurements with their 95% confidence intervals, while the plotted area represents the 95% confidence interval of the predicted output by the Monte Carlo method. Red and green dashed lines are two independent replicates, while the blue line represents the median obtained with Monte Carlo method.

shown), probably for the intensive oxygenation of cultures in chemostat, compared to closed tubes; however, for OD_{600} of 0.38 model predictions are quite in accordance with the experimental data (Fig. 3.34). Model predictions are generally consistent with experimental data, demonstrating that the identified model can correctly capture steady state values (see Fig. 3.33, green curve, for an example). However, in one case, a biological replicate gave an inconsistent steady state value (see Fig. 3.33, red dashed line). Finally, the dynamic behaviour is generally not well predicted, giving both slower and faster trends than observed in experimental data.

As regards the close-loop circuit, tests were carried out at the constant OD_{600} of 0.06, 0.08, 0.12, 0.14, 0.18 and 0.36. As shown in Fig. 3.32 (panel B), its steady-state prediction resulted more difficult than in the open-loop case. This was expected, since the close-loop circuit is the result of interconnection of modules never characterized together, while the open-loop circuit

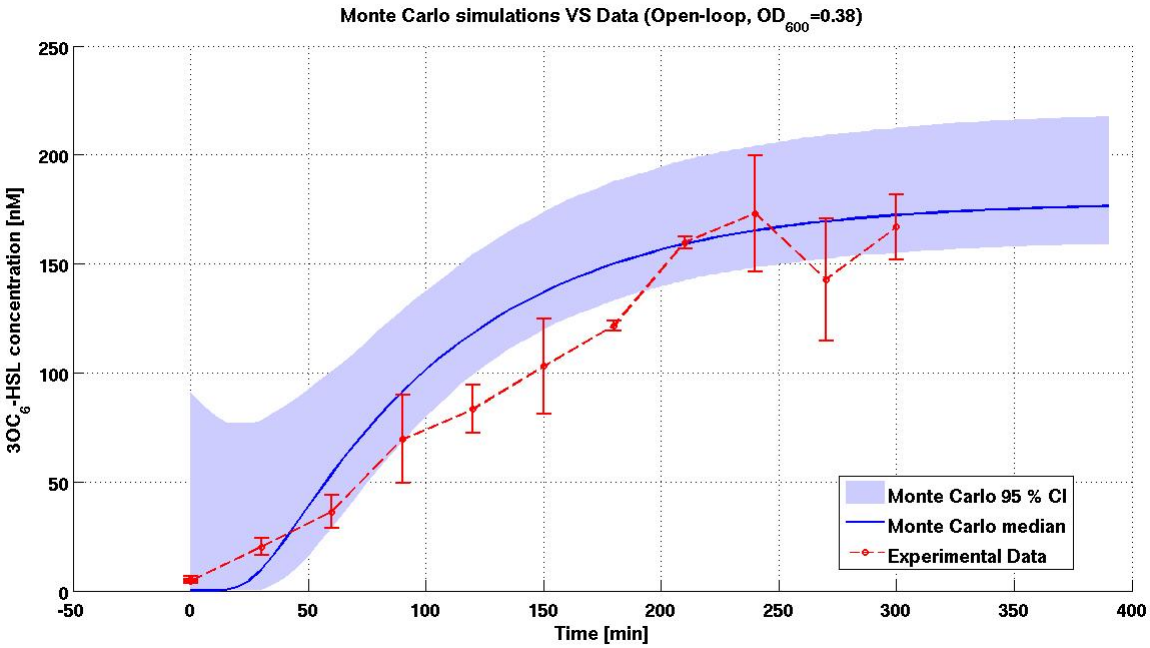


Figure 3.34: **Open-loop $3OC_6$-HSL concentration over time in chemostat experiments at OD_{600} 0.38.** Comparison between open-loop behaviour in chemostat at $OD_{600} = 0.38$ induced with 100 ng/mL aTc and model predictions based on Monte Carlo method; data points represent the average of two measurements with their 95% confidence intervals, while the plotted area represents the 95% confidence interval of the predicted output by the Monte Carlo method with the median plotted as a blue line.

was already used to obtain the LuxI parameters, although it was never tested in chemostatic conditions at different OD_{600} values. The close-loop circuit gave predictable results for low OD_{600} values (see for example Fig. 3.35), while for higher cell densities the $3OC_6$-HSL steady state is significantly underestimated, by about 3-fold (see for example Fig. 3.36). In particular, at OD_{600} of 0.12 the steady state is not significantly different from the one obtained for the open-loop circuit at the same density.
As it happened for the open-loop configuration, also in this case the dynamic behaviour is not well predicted, showing faster or slower $3OC_6$-HSL trends compared to experimental time series. Even though the absolute predicted steady state values are inconsistent with experimental data, their values did not significantly change for OD_{600} higher than 0.12, consistently with circuit qualitative predictions (see Fig. 3.3), while in the open-loop circuit steady state values showed an OD_{600}-dependent trend, as described above. This suggests that the feedback configuration of the close-loop circuit works, but its behaviour could not be quantitatively captured by the model. In particular, the qualitative observed trend of $3OC_6$-HSL steady state values in the close-loop circuit (i.e., low values for low OD_{600} values, constant values for higher OD_{600}) is well predicted by the mathematical model (see Fig. 3.37), although the absolute values of $3OC_6$-HSL and OD_{600} were different compared to Fig. 3.32.

In summary, the designed close-loop regulator was characterized and had a significantly different behaviour from the open-loop circuit, as expected, suggesting a successful design architecture. However, the steady state behaviour was not quantitatively consistent with model predictions and the dynamic behaviour had significant predictability problems, also in the open-loop configuration. The results suggest that more replicates at more cell densities would be useful to depict the actual experimental behaviour of the close-loop circuit; moreover, the reasons why the quantitative predictions are not consistent will have to be identified.
Several simulations were performed, varying parameter values (one at a time) concerning the degrading action of AiiA (i.e., $\alpha_{P_{Lux}}$, $k_{P_{Lux}}$, $k_{AiiA,max}$ and $k_{M,AiiA}$). Fig. 3.38 shows these results: parameters were changed but the predicted output did not seem to exactly reproduce the situation observed in experimental data (Fig. 3.32). This is particularly clear as in experimental data the steady state concentration moves from 30-40 nM for OD_{600} values smaller than 0.1, to the concentration of 60-100 nM for OD_{600} values greater than 0.1 (considering that the output of the open-loop is about 90 nM at OD_{600}=0.12 and about 150 nM when OD_{600} is 0.21). However, simulations can predict a steady state value of about 100 nM for high OD_{600} (0.2-0.4) when, for example, $k_{M,Aiia}$ is increased.
These discrepancies were probably due to the fact that the model did not take

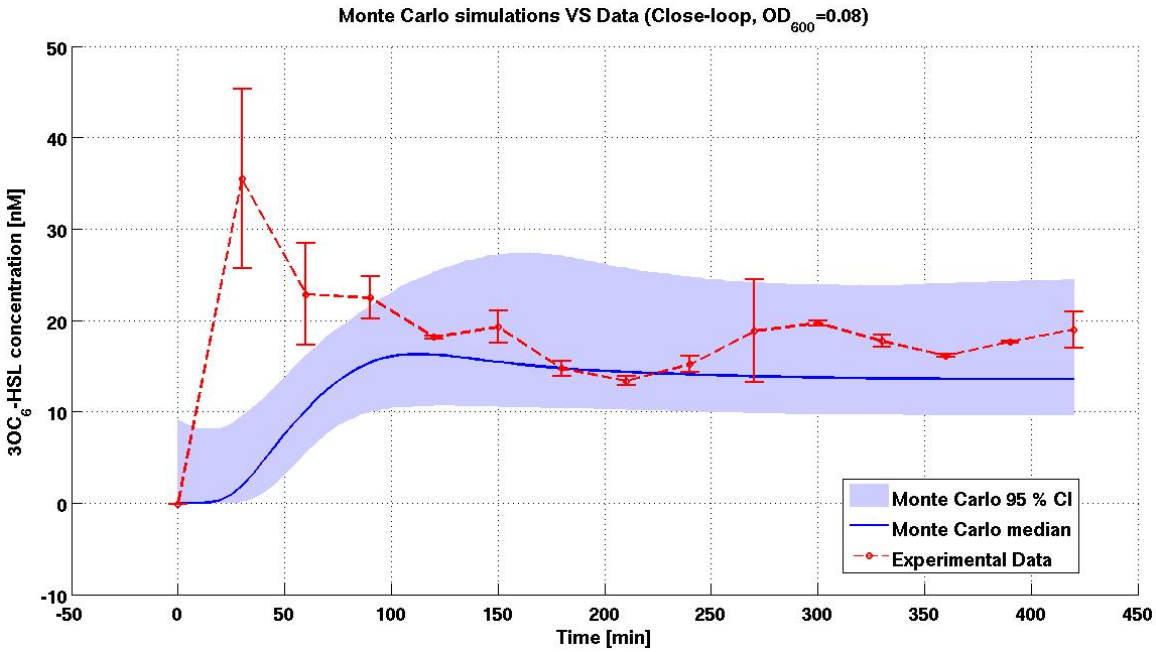


Figure 3.35: **Close-loop $3OC_6$-HSL concentration over time in chemostat experiments at $OD_{600} = 0.08$.** Comparison between close-loop behaviour in chemostat at $OD_{600} = 0.08$ induced with 100 ng/mL aTc and model predictions based on Monte Carlo method; data points represent the average of two measurements with their 95% confidence intervals, while the plotted area represents the 95% confidence interval of the predicted output by the Monte Carlo method with the median plotted as a blue line.

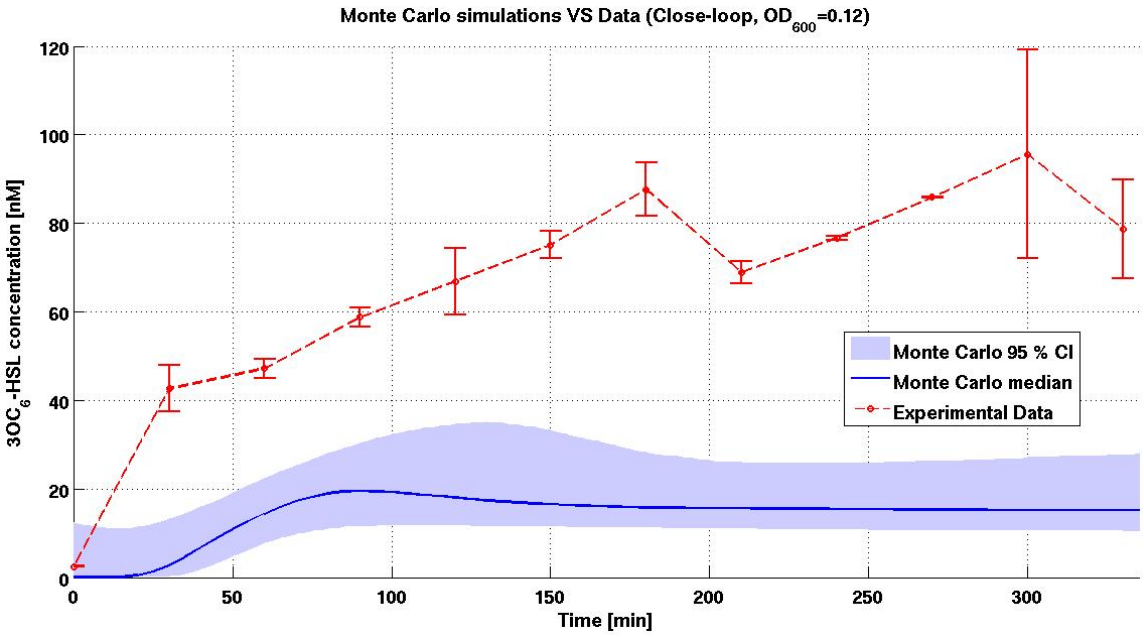


Figure 3.36: **Close-loop $3OC_6$-HSL concentration over time in chemostat experiments at $OD_{600} = 0.12$.** Comparison between close-loop behaviour in chemostat at $OD_{600} = 0.12$ induced with 100 ng/mL aTc and model predictions based on Monte Carlo method; data points represent the average of two measurements with their 95% confidence intervals, while the plotted area represents the 95% confidence interval of the predicted output by the Monte Carlo method with the median plotted as a blue line.

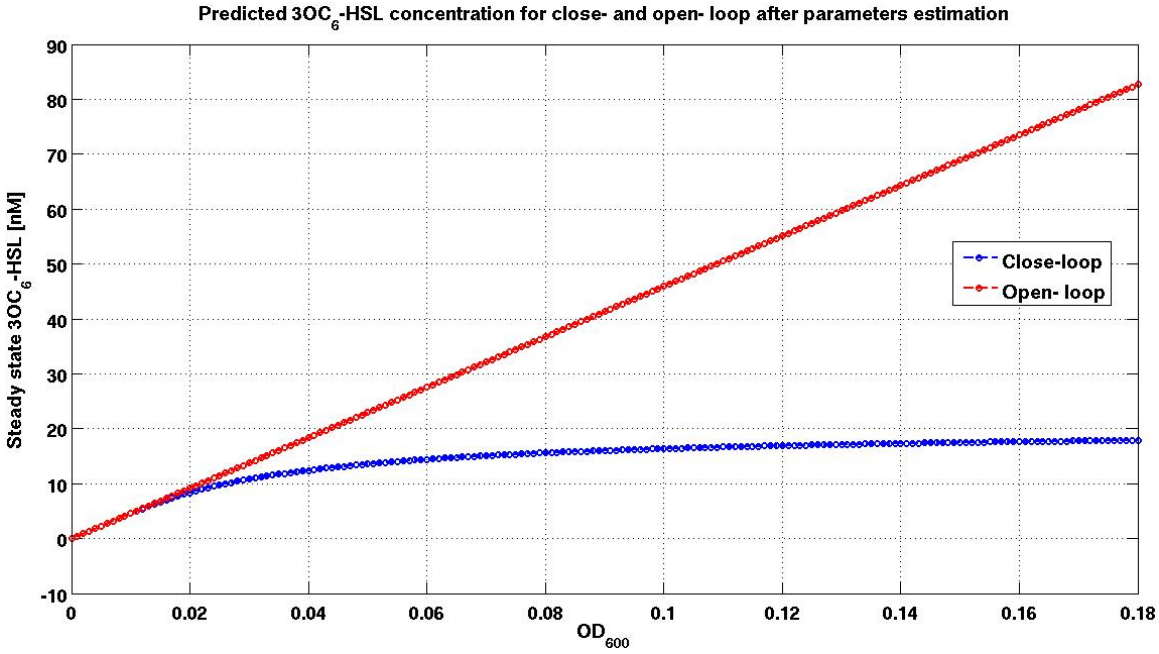


Figure 3.37: **$3OC_6$-HSL concentration predictions of close-loop with estimated parameters.** After parameters estimation on subsystems, simulations were carried out, using the nominal parameters estimated, in order to predict the $3OC_6$-HSL concentration at the steady state as a function of OD_{600}.

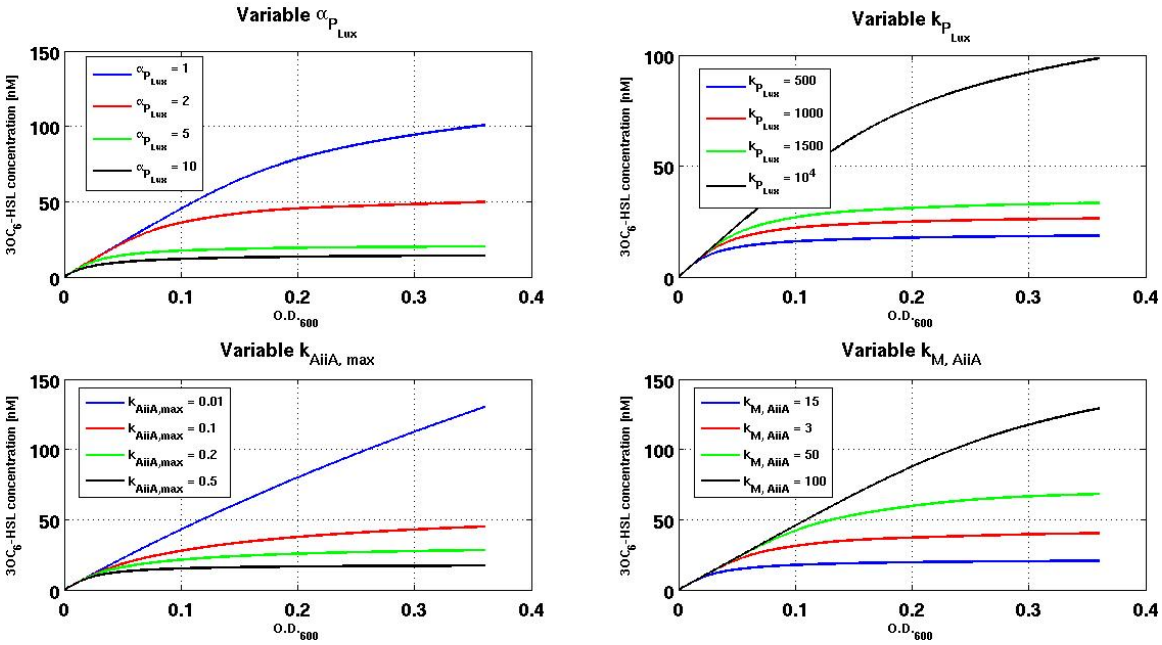


Figure 3.38: **Simulations for different parameter configurations.** $\alpha_{P_{Lux}}$, $k_{P_{Lux}}$, $k_{AiiA,max}$ and $k_{M,AiiA}$ were varied in order to verify the possibility to obtain the observed behaviour of the close-loop device. The other parameters were set to their nominal estimated values.

into account some details such as the molecular interactions between AiiA and $3OC_6$-HSL or density-dependent diffusion of $3OC_6$-HSL which might affect AiiA activity. However the overall results highlight the importance of mathematical modelling tools in Synthetic Biology and the difficulty in carrying out a rigorous bottom-up design process.

Chapter 4

Conclusions

Synthetic Biology has emerged in the last decade as an ambitious discipline which aims at operating on living systems on the basis of the engineering principles of standardization and modularity. Working in a biological context makes modularity difficult to achieve. The possibility to interconnect well characterized parts, obtaining a system with a predictable behaviour is still an ongoing issue in Synthetic Biology; in fact, biological parts could be affected by context dependent effects, such as retroactivity or fan out (described in Ch. 1), that can lead to unpredictability of the behaviour of a system resulting from the interconnection of pre-characterized modules.
This thesis has focused on the possibility to build up a genetic controller for the $3OC_6$-HSL signalling molecule in *E. coli* through a bottom-up approach. This close-loop circuit is composed by a module for the production of the LuxI enzyme, which synthesizes $3OC_6$-HSL, and by a second module, which expresses the AiiA enzyme, which degrades it, in a regulated manner. This regulation is implemented as a negative feedback, i.e., the production of AiiA is dependent on the $3OC_6$-HSL concentration. The open-loop configuration, by contrast, lacks the AiiA expression module. Mathematical modelling has been exploited as a tool to predict the behaviour of the composite system from the characterization of basic parts. In fact, describing the proposed system with a set of differential equations can help the assembly of the genetic controller and facilitate the identification of the parameters that mostly impact on the output, thus guiding the choice of some factors, such as RBS strength and/or plasmid copy number. The ODE models for the whole circuits (in both the close- and open-loop configurations) and their subparts are reported in Ch. 2. The parameters of the sub-models were identified from experiments in which batch cultures were monitored to measure the quantities of interest during their exponential growth (Ch. 3) and the estimated parameters were used to

predict the behaviour of the whole feedback system. The experimental data regarding the complete genetic controller and its non-regulated counterpart were collected by using the chemostat configuration, i.e., in continuous culture mode (Sec. 2.8), thus testing the bottom-up approach by comparing *in vivo* and *in silico* data.
As an element of novelty, the Monte Carlo method was coupled to the bottom-up approach. This was necessary as, in order to propagate the uncertainty of the subsystems parameters on the output of the composite circuit, the parameters distributions were needed. In order to obtain these distributions, the Monte Carlo approach has been applied at each step of parameter estimation, adding to the nominal prediction a gaussian noise with zero mean and variance estimated from the experimental data (for a comprehensive explanation, see Sec. 2.9). Thus, predictions were evaluated by sampling from parameters distributions.
The prediction of the steady state value reached by the open-loop circuit, in which the negative feedback on the $3OC_6$-HSL was absent, was in accordance with experimental measurements, thus confirming that the LuxI enzyme activity was similar both in the batch and chemostat mode. This circuit was used as a control of the designed close-loop controller. The close-loop circuit gave a more problematic prediction than the open-loop one, as the steady state behaviour in chemostat mode could not be predicted by the full identified model. This was probably due to non modelled secondary effects (e.g., $3OC_6$-HSL diffusion or AiiA enzyme interactions with the signalling molecule) that led to a constant underestimation (about 3-fold) of the close-loop circuit output, as shown in Ch. 3. However, the qualitative behaviour of the genetic circuit implementing the negative feedback on $3OC_6$-HSL was consistent with the model predictions; in fact, the close-loop reached a saturation value of $\sim$ 60-100 nM which was maintained for OD_{600} ranging from 0.12 to 0.36, while the open-loop circuit showed a linearly increasing steady state output as a function of the culture density, reaching a concentration >150 nM at OD_{600} of 0.38. For the described reasons, the designed architecture of the controller could be considered successful, but improvements should be performed to enhance the quantitative prediction capability of the developed mathematical model. Similarly, the dynamic behaviour of the controller, i.e., its response time, could not be precisely predicted, as described in Sec. 3.6.2, thus confirming the need of model refinement. Considering the steady state behaviour, even by tuning the model parameters (e.g., varying the model parameters of P_{Lux} activation or AiiA activity), it was not possible to simulate the observed behaviour, i.e., a steep transition of the steady state $3OC_6$-HSL concentration from 30-40 nM to 60-100 nM for OD_{600} values respectively of 0.09 and 0.12.
Hence, it has to be considered the possibility to redefine the mathematical

model in order to take into account the secondary effects above mentioned, even though this could lead to over-parameterization. Moreover, the mathematical model could be improved and used to identify those parameters (such as RBSs strength and plasmid copy number) that can lead to different steady state concentrations and dynamical behaviour. Finally, as the proposed circuit aims at behaving as a close-loop controller, a deep study on disturbance rejection has to be performed. In summary, this work has focused on many key aspects of Synthetic Biology and demonstrated that the bottom-up design of an interconnected circuit is not trivial, even though mathematical modelling can strongly support the whole design cycle.

Appendix A

Half-life measurements of chemical inducers for recombinant gene expression[1]

Regulated promoters are widely used for optimizing heterologous protein expression in host organisms, such as *E. coli*, and in several Synthetic Biology applications. They allow to trigger and modulate their transcription rate using exogenous chemical inducers in a concentration-dependent fashion. Thus, it is of great importance to know their stability in relation to the context of application, in order to obtain a robust and predictable degree of control on promoter activity.
Here, three popular inducers were considered: Isopropyl β-D-1thiogalactopyranoside (IPTG), anhydrotetracycline (aTc) and N-(3-oxohexanoyl)-L-homoserine lactone ($3OC_6$-HSL). A factorial study to identify the conditions that mostly impact on their decay rates was performed, using whole-cell biosensors. The conditions considered were: temperature (30°C - 37°C), pH (6 - 7), growth medium (LB - M9) and sterility (cultured broth - sterile broth). This can give important information on inducers stability and help to define the most suitable experimental set-up, as it was done studying the genetic negative feedback controller that was investigated at pH 6, as $3OC_6$-HSL is more stable.

[1]The content of this appendix has been published in [68]

A.1 Methods

A.1.1 Reagents and media

IPTG (I1284, Sigma Aldrich), aTc (631310, Clontech) and $3OC_6$-HSL (K3007, Sigma Aldrich) were used as inducers and stored as described in Sec. 2.1.2. LB and M9 supplemented media were prepared as described in Sec. 2.1.1. When appropriate, pH was adjusted with hydrochloric acid or sodium hydroxyde.

A.1.2 Biosensors

Fig. A.1 describes the recombinant *E. coli* bearing the biosensors. The IPTG-, aTc- and $3OC_6$-HSL- biosensor genetic devices, called BBa_J107010, BBa_I13521 and BBa_J107053 respectively, are BioBrick™ parts from the MIT Registry of Standard Biological Parts [22]. They have been assembled into the pSB3K3 medium-copy plasmid (with kanamycin-resistance marker and p15A replication origin) by using the BioBrick™ Standard Assembly procedure [23] and conventional molecular biology techniques [60]. Detailed description about plasmid construction, as well as full DNA sequences, can be found in the individual parts web pages of the Registry [22]. TOP10 (Invitrogen) *E. coli* strain was used as a host for cloning. MG1655-Z1 E. coli strain [65] was used as the final host for the biosensing plasmids, as its genome encodes constitutive LacI and TetR over-expression cassettes. Chemically competent TOP10 were transformed according to manufacturer's instructions, while MG1655-Z1 competent cells were prepared as described in [60] and were heat-shock transformed at 42°C. The described whole-cell biosensors were routinely stored at -80°C in 20% glycerol stocks and they were always grown in selective media with 20 mg/L of kanamycin.

A.1.3 Sampling

1 mM of IPTG, 50 ng/mL of aTc or 50 nM of $3OC_6$-HSL were added to 3 mL of sterile medium (sterile broth) or to 3 mL of a non-recombinant MG1655-Z1 E. coli culture (cultured broth), prepared by inoculating 3 mL of sterile medium (LB or M9) with a bacterial colony isolated from a streaked LB agar plate. Sterile and cultured broths, in 15-mL tubes, were incubated with shaking at 220 rpm at 30°C or 37°C. The pH of media was set at 6.0 or 7.0. As a result, each inducer was tested in 16 conditions (cultured/sterile broth; pH 6.0/7.0; 30/37°C; LB/M9 medium). At appropriate time points, 100 μL were taken from the cultured broth and centrifuged (13000 rpm, 1 min), bacteria-free supernatant was taken and stored at -20°C before analysis. For sterile broth,

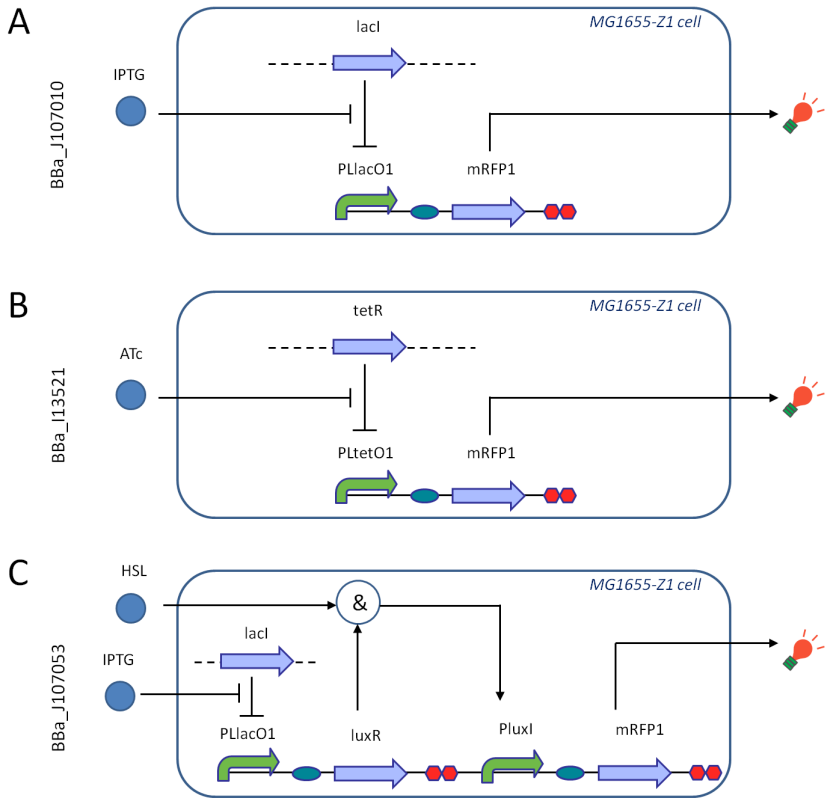


Figure A.1: **Biosensors used in this study.** IPTG- (A), aTc- (B) and $3OC_6$-HSL- (C) biosensor constructs and functioning. The BioBrick™ code (BBa_code) of the biosensor-encoding sequences and the microbial host are reported for each construct. All of them were used in the pSB3K3 medium-copy BioBrickTM vector backbone with the p15A replication origin and kanamycin resistance marker. Symbols: curved arrows represent promoters, straight arrows represent coding sequences, ovals represent RBSs, octagons represent transcriptional terminators, "&" means that both inducer and gene product are required to activate the promoter, circles represent chemical inducers, bulbs represent a red fluorescence output and finally dashed lines represent MG1655-Z1 genomic DNA. PLlacO1 and PLtetO1 are synthetic promoters that can be repressed by LacI and TetR respectively. PluxI is the wild type luxI promoter from *V. fischeri* that can be activared by the LuxR-$3OC_6$-HSL complex. In the $3OC_6$-HSL-biosensor (panel C), LuxR expression was always induced by 500 μM of IPTG, as gene expression is driven by the PLlacO1 promoter.

100 μL were taken and directly stored at -20°C. For each investigated condition, at least two independent tubes were prepared and sampled in different days. When required, cultured broths were prepared by inoculating a colony of the lacY-deficient strain DB3.1 [69, 70] or MC1061 [71] instead of MG1655-Z1. The growth of cultures was measured by optical density at 600 nm (OD_{600}) in a 96-well microplate (Greiner) in the Infinite F200 reader (Tecan). The presented OD_{600} measurements are background (sterile media)-subtracted absorbance values, relative to the pathlength of 200 μL of liquid in microplate.

A.1.4 Biosensor-based fluorescence assays

$3OC_6$-HSL was measured with the procedure described in Sec. 2.7. IPTG and aTc were measured via analogous procedures, using the specific biosensors described above.

A.1.5 Data analysis

The acquired time series were processed as described in Sec. 2.5 to obtain a value proportional to the average RFP protein synthesis rate per cell (S_{cell}), which is related to the inducer concentration in the wells. Briefly, raw absorbance and fluorescence time series were background-subtracted over time by using sterile medium and non-fluorescent culture as background to obtain the actual fluorescence (F) and absorbance (OD_{600}) signals in the well, proportional to the per-well RFP molecules and number of bacteria, respectively. The $dF/dt/OD_{600}$ signal was averaged at its steady state to obtain S_{cell}. The steady state is reached after a time that depends on the dynamics of each specific biosensor. S_{cell} is measured as arbitrary units (A.U.$_r$) of RFP per minute per cell.
The obtained standard calibration curve was fitted with the Hill function Eq. A.1.

$$S_{cell}^{SS} = \delta_{p_X} + \frac{\alpha_{p_X}}{1 + \left(\frac{k_{p_X}}{[I]}\right)^{\eta_{p_X}}} \tag{A.1}$$

α_{p_X} is the S_{cell} at maximum induction, k_{p_X} is the inducer concentration that yields $S_{cell} = \alpha_{p_X}/2$, η_{p_X} is the Hill coefficient, δ_{p_X} is the basic activity of the biosensor in absence of inducer and $[I]$ is the inducer concentration. The inducer concentration of the unknown samples was computed from their S_{cell} value via the identified Hill function representing the calibration curve. When required, the resulting value was multiplied by the applied dilution factor of the sample. Each inducer decay curve was expressed as percentage of its value at t=0 and, for each tested condition, the independently measured time series

were pooled in a single time series $c(t)$.
Assuming a first-order exponential decay for $c(t)$, its natural logarithm was fitted with the linear function shown in Eq. A.2.

$$\ln(c(t)) = ln(c(0)) - k \cdot t \tag{A.2}$$

In Eq. A.2, k is the degradation rate of the inducer. Fitting was performed with MATLAB R2011b (MathWorks, Natick, MA) via the *regress* routine, which also computes the 95% confidence intervals of k. When required, inducer half-life was computed as $\ln(2)/k$. The measurement unit of k is h^{-1}.
For each inducer, the linear model shown in Eq. A.3 with interactions was used to determine the experimental factors that significantly affect the decay rate.

$$\begin{aligned}
&\ln(c(t)) = \alpha + \alpha_{Temp} \cdot Temp + \alpha_{pH} \cdot pH + \alpha_{Med} \cdot Med + \alpha_{Ster} \cdot Ster + \\
&+\alpha_{Temp,pH} \cdot Temp \cdot pH + \alpha_{Temp,Med} \cdot Temp \cdot Med + \\
&+\alpha_{Temp,Ster} \cdot Temp \cdot Ster + \alpha_{pH,Med} \cdot pH \cdot Med + \\
&+\alpha_{pH,Ster} \cdot pH \cdot Ster + \alpha_{Med,Ster} \cdot Med \cdot Ster + \\
&-(\beta_t + \beta_{Temp} \cdot Temp + \\
&+\beta_{pH} \cdot pH + \beta_{Med} \cdot Med + \\
&+\beta_{Ster} \cdot Ster + \beta_{Temp,pH} \cdot Temp \cdot pH + \\
&+\beta_{Temp,Med} \cdot Temp \cdot Med + \beta_{Temp,Ster} \cdot Temp \cdot Ster + \\
&+\beta_{pH,Med} \cdot pH \cdot Med + \beta_{pH,Ster} \cdot pH \cdot Ster + \\
&+\beta_{Med,Ster} \cdot Med \cdot Ster) \cdot t
\end{aligned} \tag{A.3}$$

Eq. A.3 model describes $\ln(c(t))$ behaviour by using time (t) as continuous independent variable and pH, temperature ($Temp$), sterility ($Ster$) and medium (Med) as Boolean independent variables assuming the -1 or +1 value, where -1 corresponds to pH 6.0, 30°C, sterile broth and LB, while +1 corresponds to pH 7.0, 37°C, cultured broth and M9, respectively.
The MATLAB *regstats* routine was used to estimate the regression coefficients and their p-value (P). To test if an experimental factor (or interaction between two factors) significantly contributes to inducer degradation, its β coefficient was considered, i.e., the coefficient of the factor * time term. When $P<0.05$ for a given coefficient, the related factor (or interaction) is considered to significantly affect the inducer decay rate. The variability of $ln(c(t))$ explained by the *j-th* factor (SS_j) or an interaction term was quantified and compared with the residual unexplained variability of the (reference) null model in Eq. A.4.

$$\begin{aligned}\ln(c(t)) = \alpha + \alpha_{Temp} \cdot Temp + \alpha_{pH} \cdot pH + \alpha_{Med} \cdot Med + \alpha_{Ster} \cdot Ster + \\ + \alpha_{Temp,pH} \cdot Temp \cdot pH + \alpha_{Temp,Med} \cdot Temp \cdot Med + \\ + \alpha_{Temp,Ster} \cdot Temp \cdot Ster + \alpha_{pH,Med} \cdot pH \cdot Med + \\ + \alpha_{pH,Ster} \cdot pH \cdot Ster + \alpha_{Med,Ster} \cdot Med \cdot Ster + \\ - \beta_t \cdot t\end{aligned} \tag{A.4}$$

In particular, SS_{tot}, SS_j and the residual unexplained variability SSE can be calculated as follows:

$$SS_{tot} = \sum_i^N (y_i - y_{null,i})^2 \tag{A.5}$$

$$SS_j = SS_{tot} - \sum_i^N (y_i - y_{null,i} + \beta_j \cdot X_{i,j} \cdot t_i)^2 \tag{A.6}$$

$$SSE = \sum_i^N (y_i - y_{full,i})^2 \tag{A.7}$$

y_i are the experimental data, $y_{full,i}$ is the full model (Eq. A.3) output prediction computed in the *i-th* experimental point, $y_{null,i}$ is the null model (Eq. A.4) output prediction computed in the *i-th* experimental point, β_j is the regression coefficient of the *j-th* factor or interaction term and $X_{i,j}$ is the value (-1 or +1) of the *j-th* factor or interaction term in the *i-th* experimental point. When $X_{i,j}$ is an interaction term, it is computed as the product of the individual factors values (-1 or +1) in the *i-th* point. As a result, SS_j is the variability of the output that is explained by the $\beta_j \cdot X_j \cdot t$ term alone, thus quantifying the importance of the *j-th* factor or interaction term. SSE represents the output variability that remains unexplained by the full model (which includes all the considered factors and interaction terms). When appropriate, the linear model was used without considering cultured broth data and, in this case, sterility was not included as a factor.

A.1.6 Measurement of aTc with High Performance Liquid Chromatography (HPLC)

HPLC is known to enable aTc measurements via UV detection [72]. Here, a 10AD/vp HPLC (Shimadzu) was used with a Discovery C18 HPLC column 150x4.6 mm, 5 μm (Supelco) and a Diode Array UV Detector SPD-M10AVP (Shimadzu). The column was kept at 25°C. The flow rate of the mobile phase was 0.8 mL/min. Gradient elution was performed with solutions A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) following the

profile: 0-2 min 10% of B, 2-8 min 10% to 30% of B (linear increase), 8-12 min 30% to 80% of B (linear increase), 12-17 min 80% of B (constant) and 17-19 min 80% to 10% of B (linear decrease) and 19-25 min 10% (constant). Each single analysis was run for 25 min with additional 3 min for column conditioning. The injection of 100 μL was performed via automatic injector. A standard calibration curve was prepared in sterile M9 at pH 7.0. The LabSolutions software (Shimadzu) was used to analyze HPLC data. As identified by analyses of standard solutions, aTc retention time is 14.26 min and the maximum absorbance of aTc is observed at a wavelength of 430 nm. Experiments and sampling were performed as described above with the following exceptions: culture broth volume was 7 mL, the initial aTc concentration was 200 ng/mL, at appropriate time points 1.5 ml were taken and filter-sterilized (0.2 μm) for HPLC analysis and 100 μL were taken as described above for analysis via biosensor. All the samples were stored at -20°C before the analyses.

A.2 Results

Fig. A.2 shows calibration curves of the IPTG, aTc and $3OC_6$-HSL biosensors in a representative experiment. Each biosensor yields reproducible measurements of the same samples in different days with a CV of 7%, 21% and 11% for IPTG, aTc and $3OC_6$-HSL, respectively (data not shown). No relevant difference in biosensor activity can be detected when a known inducer amount was diluted in either exhausted or sterile medium (see Fig. A.3), thus demonstrating that these biosensors can be successfully used to detect inducer concentration in sterile or cultured broth conditions. Tab. A.1 reports the degradation rates of IPTG, aTc and $3OC_6$-HSL estimated from experimental data in all the investigated conditions, while Fig. A.4, Fig. A.5 and Fig. A.6 show all the data and fitted curves. IPTG is stable over 32 hours in all the sterile broth conditions (the 95% confidence interval of k contains zero in all the conditions, i.e., k is not distinguishable from zero from a statistical point of view), while in all the cultured broth conditions IPTG disappears in the supernatant (see Tab. A.1 and Fig. A.4). IPTG uptake by lactose permease (LacY) of the MG1655-Z1 strain could explain the observed phenomenon, as described in [73]. Correlation between OD_{600} and residual percent IPTG over time (see Fig. A.7 A) is consistent with this statement.

To further validate this, we measured IPTG in the LB pH 7.0 at 37°C cultured broth condition, inoculated with the DB3.1 or MC1061 strains, both deficient in LacY. Results (see Fig. A.7 B - C) showed that IPTG is not degraded in this condition and the inducer disappearance in the medium of MG1655-Z1 was actually due to uptake by LacY. These results confirm the high stability of the IPTG molecule, previously found in literature in different

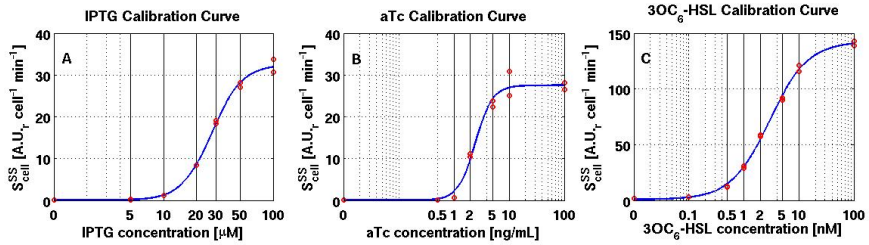


Figure A.2: **Calibration curve of the IPTG (A), aTc (B) and $3OC_6$-HSL (C) biosensors in a representative experiment.** Each curve was obtained in a single representative experiment. In all the three experiments, serial dilutions of the calibration samples were prepared in M9 at pH 7.0. The final inducer concentrations in the microplate well are reported in the X-axis, while the corresponding S_{cell} values (two replicates) are reported in the Y-axis (circles). Solid line represents data fitting with a Hill function, as described in Sec. 2.7.

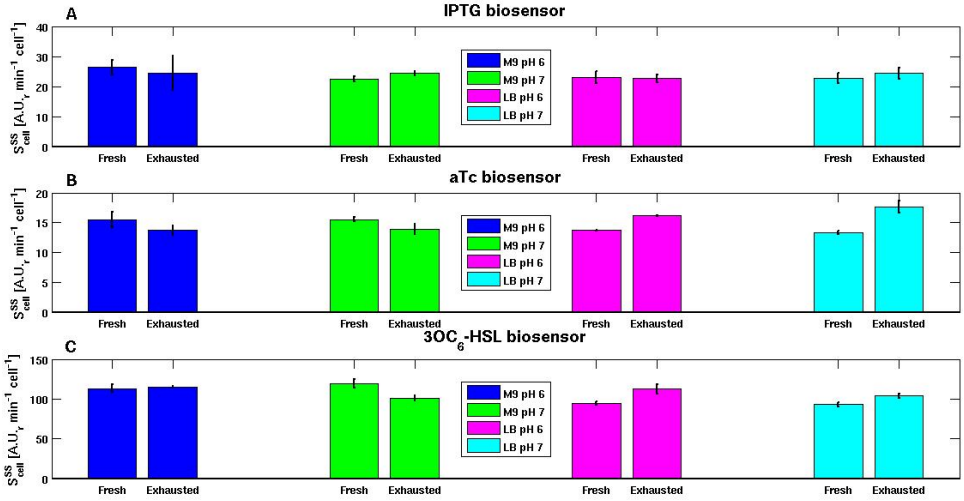


Figure A.3: **Effect of exhausted vs fresh medium on the activity of the IPTG (A), aTc (B) or $3OC_6$-HSL (C) biosensors.** The activity of the three biosensors was measured in terms of S_{cell} when assaying samples of known concentrations of IPTG, aTc or $3OC_6$-HSL. Inducers were diluted in different fresh (LB pH 6.0, LB pH 7.0, M9 pH 6.0 and M9 pH 7.0) or exhausted (from the supernatant of 32-h cultures grown at 37°C in LB pH 6.0, LB pH 7.0, M9 pH 6.0 and M9 pH 7.0) media. Assays were performed as described in the Methods section, except that the cultures used to prepare the exhausted media were grown in absence of inducers. The final concentration of IPTG, aTc and $3OC_6$-HSL in the microplate wells was 20 μM, 2 ng/mL and 5 nM, respectively. Bars represent the average of at least two independent replicates, while error bars represent the 95% confidence intervals of the average value.

Table A.1: **Degradation rate measurements, expressed as h^{-1}, for IPTG, aTc and $3OC_6$-HSL in all the tested conditions.** Estimated lower (L) and upper (U) 95% confidence intervals of the measured decay rates (k) are also reported. According to the first-order decay model (see Sec. A.1.5), negative values of k are due to experimental noise and indicate a non-significant decaying trend of the inducer. In this table, cultured broth conditions were tested with the MG1655-Z1 strain.

pH	**Temperature**	**Sterility**	**Medium**	k**(L;U) IPTG**	k**(L;U) aTc**	k**(L;U) 3OC$_6$-HSL**
6	30°C	Sterile	LB	0.001 (-0.002;0.005)	0.021 (0.018;0.024)	0.001 (-0.001;0.003)
6	30°C	Sterile	M9	-0.005 (-0.017;0.007)	0.029 (0.022;0.035)	0.021 (0.012;0.029)
6	37°C	Sterile	LB	0.001 (-0.002;0.004)	0.039 (0.03;0.049)	0.01 (0.007;0.014)
6	37°C	Sterile	M9	-0.002 (-0.01;0.006)	0.044 (0.042;0.047)	0.034 (0.024;0.044)
7	30°C	Sterile	LB	0.005 (0;0.01)	0.017 (0.016;0.019)	0.019 (0.013;0.024)
7	30°C	Sterile	M9	-0.001 (-0.005;0.004)	0.021 (0.012;0.031)	0.085 (0.069;0.102)
7	37°C	Sterile	LB	-0.004 (-0.011;0.002)	0.029 (0.019;0.038)	0.039 (0.037;0.042)
7	37°C	Sterile	M9	-0.003 (-0.01;0.003)	0.041 (0.036;0.046)	0.076 (0.037;0.114)
6	30°C	Cultured	LB	0.116 (0.09;0.143)	0.021 (0.007;0.035)	-0.006 (-0.009;-0.003)
6	30°C	Cultured	M9	0.037 (0.027;0.048)	0.026 (0.018;0.033)	0.007 (0.002;0.012)
6	37°C	Cultured	LB	0.115 (0.089;0.141)	0.051 (0.035;0.067)	0.027 (0.013;0.04)
6	37°C	Cultured	M9	0.114 (0.07;0.158)	0.053 (0.032;0.073)	0.013 (0.003;0.024)
7	30°C	Cultured	LB	0.096 (0.072;0.12)	0.02 (0.006;0.033)	0.063 (0.032;0.094)
7	30°C	Cultured	M9	0.096 (0.069;0.122)	0.031 (0.011;0.052)	0.062 (0.045;0.079)
7	37°C	Cultured	LB	0.075 (0.024;0.126)	0.042 (0.021;0.064)	0.091 (0.066;0.116)
7	37°C	Cultured	M9	0.093 (0.071;0.114)	0.058 (0.038;0.079)	0.088 (0.062;0.114)

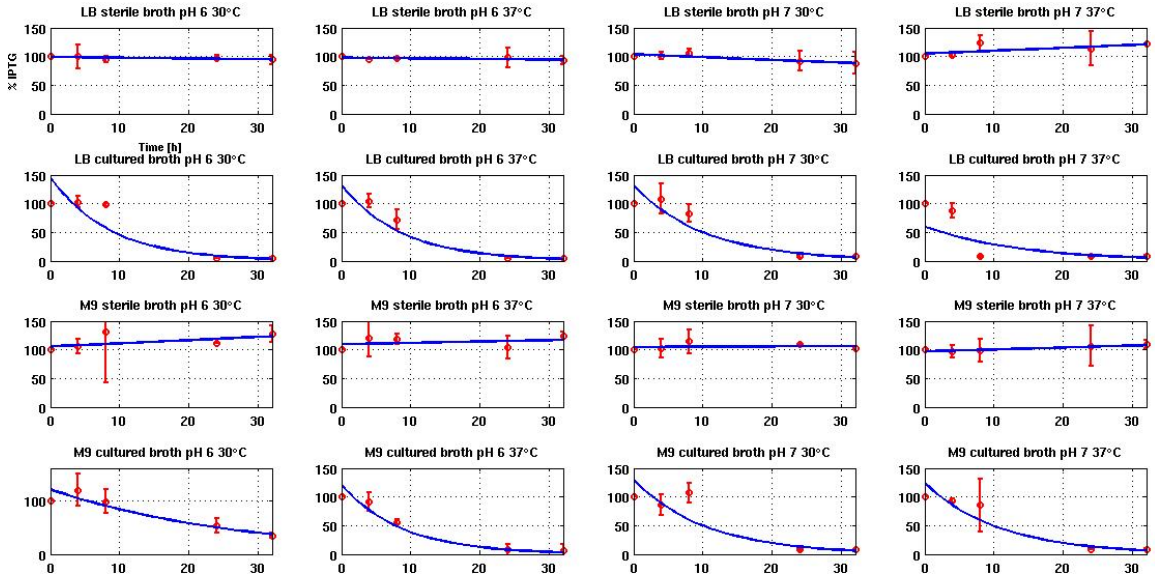


Figure A.4: **Data and fitting in IPTG degradation assays.** Fitting of IPTG measurements over time during the performed degradation assays. Data points represent the average concentration values, error bars represent the 95% confidence intervals of the average value of at least two independent experiments and solid lines are the fitted first-order exponential model curves.

conditions [74, 75, 73], and the role of lactose permease in IPTG uptake by LacY+ strains. None of the tested environmental factors have been found to affect IPTG stability.
aTc is degraded in all the tested conditions with an average half life of about 20 hours (see Tab. A.1). The factors that give a significant contribution in degradation were identified by fitting a linear regression model (see Sec. A.1.5). According to Tab. A.2, the statistically significant factors (see p-values) are temperature, sterility and medium. Temperature was found to be the main significant factor affecting degradation, while sterility and medium give a minor contribution (see Fig. A.8).

In particular, 42% of the total variability is explained by temperature, while sterility and medium together account for only 11%. The decay rate of aTc is 2-fold higher at 37°C ($\beta + \beta_{Temp} = 0.045h^{-1}$) than at 30°C ($\beta - \beta_{Temp} = 0.023h^{-1}$), which corresponds to half-lives of 15 h and 30 h, respectively, without considering the other factors. Conversely, degradation is only 1.26-fold higher in cultured broth or M9 ($\beta + \beta_{Ster} = 0.038h^{-1}$, $\beta + \beta_{Med} = 0.038h^{-1}$) than in sterile broth or LB ($\beta - \beta_{Ster} = 0.03h^{-1}$, $\beta - \beta_{Med} = 0.03h^{-1}$). Conclusions do not substantially change if cultured broth conditions are excluded from the study (see Table Tab. A.3 and Fig. A.9): temperature is confirmed to

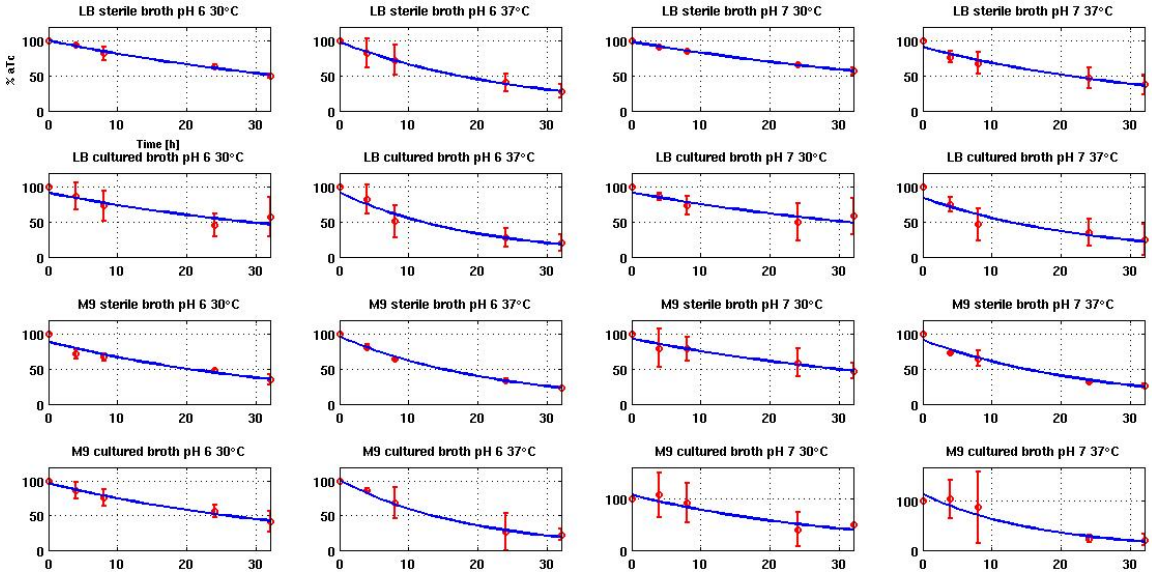


Figure A.5: **Data and fitting in aTc degradation assays.** Fitting of aTc measurements over time during the performed degradation assays. Data points represent the average concentration values, error bars represent the 95% confidence intervals of the average value of at least two independent experiments and solid lines are the fitted first-order exponential model curves.

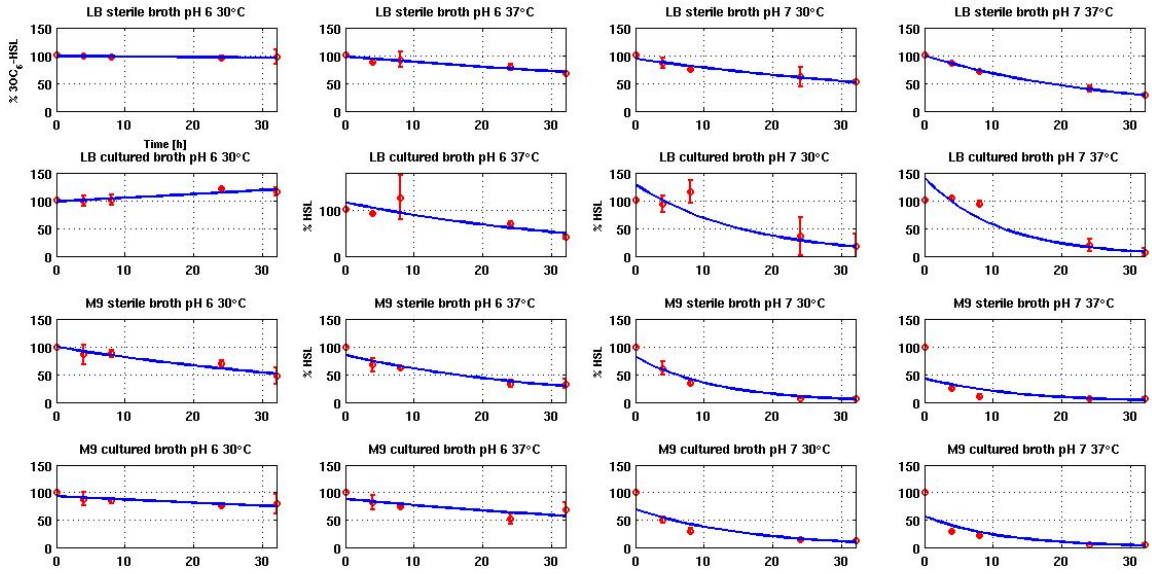


Figure A.6: **Data and fitting in $3OC_6$-HSL degradation assays.** Fitting of $3OC_6$-HSL measurements over time during the performed degradation assays. Data points represent the average concentration values, error bars represent the 95% confidence intervals of the average value of at least two independent experiments and solid lines are the fitted first-order exponential model curves.

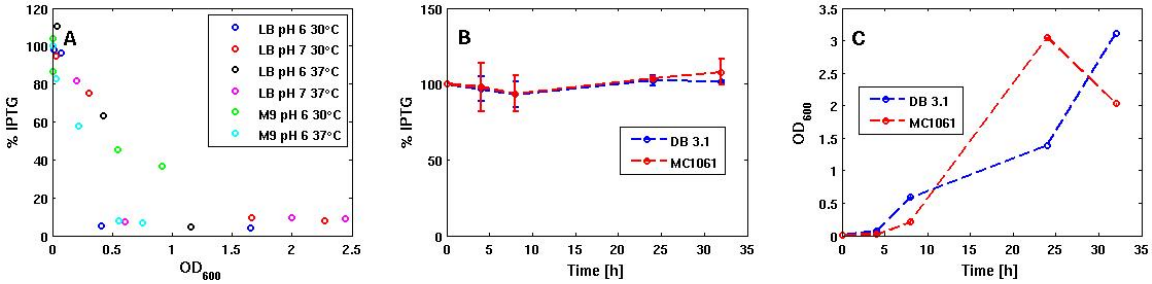


Figure A.7: **IPTG measured in the supernatant of cultures of strains with or without lactose permease, as a function of cell density.** Measured IPTG in supernatants of MG1655-Z1 (lacY+) cultured broth conditions as a function of cell density, expressed as OD_{600} (A). Measured IPTG in supernatants of DB3.1 and MC1061 (lacY-) cultured broth conditions in LB pH 7.0 at 37°C over time (B). Growth of DB3.1 and MC1061, expressed as OD_{600}, in a representative experiment performed in LB pH 7.0 at 37°C (C). In all the panels, circles represent data points, with colours specified in the legends, and error bars represent the 95% confidence intervals of the average.

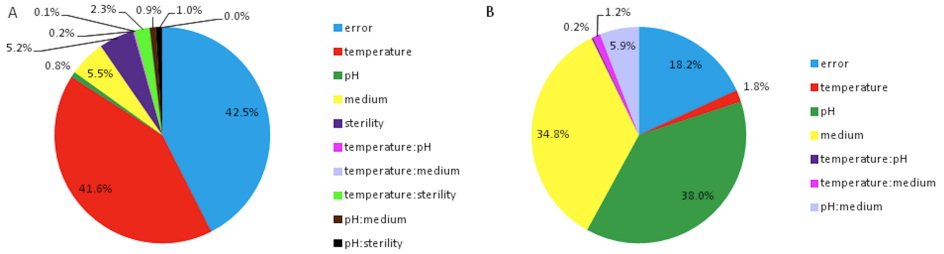


Figure A.8: **Quantification of the main factors and their interaction terms contributing to the degradation of aTc (A) and $3OC_6$-HSL (B).** The pie charts show the variability of the linear model output explained by each of the considered factors and interaction terms (SS_j, Eq. A.6). 100% of the pie corresponds to the residual error of the null model (SS_{tot}, Eq. A.5), while the error term represents the residual error that remains unexplained by the full linear model (SSE, Eq. A.7). Colon indicates an interaction between two factors. Only sterile broth conditions are considered for $3OC_6$-HSL and, as a result, the sterility factor and its related interaction terms are not present in panel B.

Table A.2: **Parameters of the linear model.** Estimated regression coefficients and p-values are reported for the study of aTc. The regression coefficients β with ($p<0.05$, t-test) are significantly different from zero and identify the main factor or interaction terms that significantly contribute to inducer decay.

aTc		
Parameter	**Identified Value**	**p-value**
α	4.0562	1.2e-165
α_{Temp}	0.0003582	0.99
α_{pH}	0.002346	0.93
α_{Med}	0.02576	0.31
α_{Ster}	0.01122	0.66
$\alpha_{Temp,pH}$	-0.01358	0.59
$\alpha_{Temp,Med}$	0.02222	0.38
$\alpha_{Temp,Ster}$	0.004408	0.86
$\alpha_{pH,Med}$	0.02403	0.34
$\alpha_{pH,Ster}$	0.01493	0.56
$\alpha_{Med,Ster}$	0.04917	5.4e-2
β_t	0.03398	2.7e-52
β_{Temp}	0.01076	1.5e-12
β_{pH}	-0.001510	0.28
β_{Med}	0.003901	5.4e-3
β_{Ster}	0.003825	6.4e-3
$\beta_{Temp,pH}$	-0.0006518	0.64
$\beta_{Temp,Med}$	0.0004395	0.75
$\beta_{Temp,Ster}$	0.002527	0.07
$\beta_{pH,Med}$	0.001577	0.26
$\beta_{pH,Ster}$	0.01638	0.24
$\beta_{Med,Ster}$	0.0002971	0.83

Table A.3: **Parameters of the linear model for aTc considering only the sterile broth conditions.** Estimated regression coefficients and p-values are reported for the study of aTc in sterile broth. The regression coefficients β with ($p<0.05$, t-test) are significantly different from zero and identify the main factor or interaction terms that significantly contribute to inducer decay.

Parameter	Identified Value	p-value
α	4.551	1.48e-98
α_{Temp}	-0.004049	0.83
α_{pH}	-0.01258	0.51
α_{Med}	-0.02341	0.23
$\alpha_{Temp,pH}$	-0.01977	0.31
$\alpha_{Temp,Med}$	0.02053	0.29
$\alpha_{pH,Med}$	0.01120	0.56
β_t	0.03016	3.5e-39
β_{Temp}	0.008235	4.1e-11
β_{pH}	-0.003148	3.6e-3
β_{Med}	0.003604	9.7e-4
$\beta_{Temp,pH}$	-0.0003665	0.73
$\beta_{Temp,Med}$	0.0007120	0.50
$\beta_{pH,Med}$	0.0004307	0.68

be the most important factor, while the other statistically significant factors (in this case pH and medium) give a much lower contribution.

HPLC was used to confirm the obtained results of aTc decay in one representative condition (M9 cultured broth at 30°C, pH 7.0), yielding consistent aTc measurements and degradation rates (see Fig. A.10).
$3OC_6$-HSL spans a wide range of decay rates, from zero (i.e., no detectable degradation in 32 hours) to a half-life of about 8 hours (see Tab. A.1). Degradation was firstly studied in sterile broth conditions by fitting a linear regression model, since pH is already known to affect inducer decay and in cultured broth conditions bacteria can change medium acidity during growth (see Fig. A.11). Tab. A.4 shows that the identified significant factors affecting $3OC_6$-HSL degradation are medium and pH with a significant interaction between them. They explain 38% (pH), 35% (Med) and 6% (pH:Med) of the output variability, respectively (see Fig. A.8). In particular, $3OC_6$-HSL degradation is negligible in the LB pH 6.0 condition ($\beta - \beta_{pH} - \beta_{Med} + \beta_{pH,Med} =\sim 0$ h^{-1}). Degradation increases in the LB pH 7.0 ($\beta + \beta_{pH} - \beta_{Med} - \beta_{pH,Med} = 0.029$ h^{-1}) and M9 pH 6.0 ($\beta - \beta_{pH} + \beta_{Med} - \beta_{pH,Med} = 0.029$ h^{-1}) conditions, while in M9 pH 7.0 degradation further increases because of the individual contri-

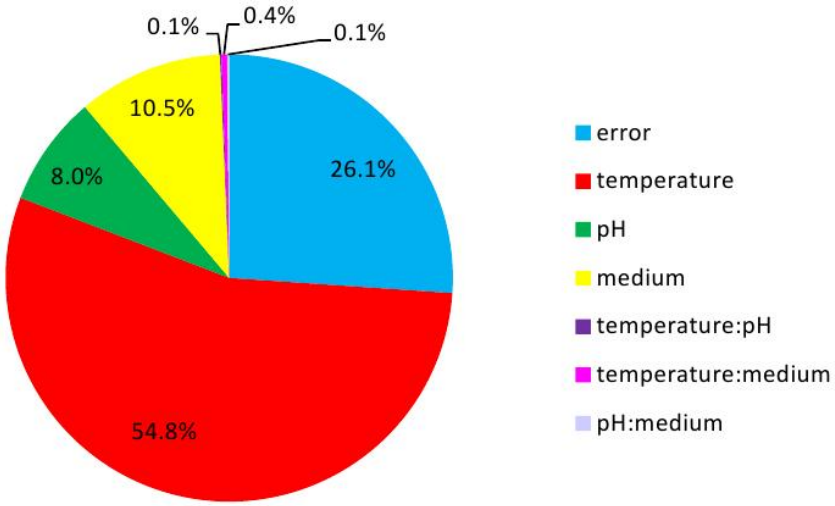


Figure A.9: **Quantification of the main factors and their interaction terms affecting the degradation of aTc in sterile broth.** The pie chart shows the variability explained by each of the considered factors and interaction terms (SS_j, Eq. A.6). 100% of the pie corresponds to the residual error of the null model (SS_{tot}, Eq. A.5), while the error term represents the residual error that remains unexplained by the full linear model (SSE, Eq. A.7). Colon indicates an interaction between two factors.

Table A.4: **Parameters of the linear model for $3OC_6$-HSL decay.** Estimated regression coefficients and p-values are reported for the study of $3OC_6$-HSL considering only sterile broth conditions. The regression coefficients β with ($p<0.05$, t-test) are significantly different from zero and identify the main factor or interaction terms that significantly contribute to inducer decay.

$3OC_6$-HSL		
Parameter	**Identified Value**	**p-value**
α	4.441	9.4e-74
α_{Temp}	-0.09555	3.5e-2
α_{pH}	-0.1136	1.3e-2
α_{Med}	-0.1403	2.4e-3
$\alpha_{Temp,pH}$	-0.05275	0.24
$\alpha_{Temp,Med}$	-0.01082	1.8e-2
$\alpha_{pH,Med}$	-0.1057	0.02
β_t	0.03560	1.8e-22
β_{Temp}	0.004155	0.091
β_{pH}	0.01911	4.4e-11
β_{Med}	0.01827	1.8e-10
$\beta_{Temp,pH}$	-0.001483	0.54
$\beta_{Temp,Med}$	0.003336	0.17
$\beta_{pH,Med}$	0.007504	2.9e-3

butions of pH, medium and their interaction ($\beta + \beta_{pH} + \beta_{Med} + \beta_{pH,Med} = 0.08$ h^{-1}). Taken together, the results indicate that pH and medium can tune $3OC_6$-HSL half life from ∼9 hours (M9 pH 7.0) to a time that is much longer than the performed experiments, which are 32-hour long (LB pH 6.0). $3OC_6$-HSL decay data from cultured broth conditions were then considered. Fig. A.12) shows the $3OC_6$-HSL degradation rates (k) reported in Tab. A.1, measured in the eight cultured broth conditions, as a function of the average pH of a culture in those conditions, computed from the data shown in Fig. A.11. As expected, in both the LB and M9 conditions there is a significant pH-dependent trend ($p<0.05$, t-test). Moreover, k is systematically higher in M9 than in LB for a wide range of pH. These results confirm the significant contribution of pH and growth medium on $3OC_6$-HSL decay. However, since the linear regression model was fitted on sterile broth data, the precise prediction of $3OC_6$-HSL degradation in cultured broth may not be inferred by the model, although a similar medium- and pH-dependent trend is qualitatively observed.

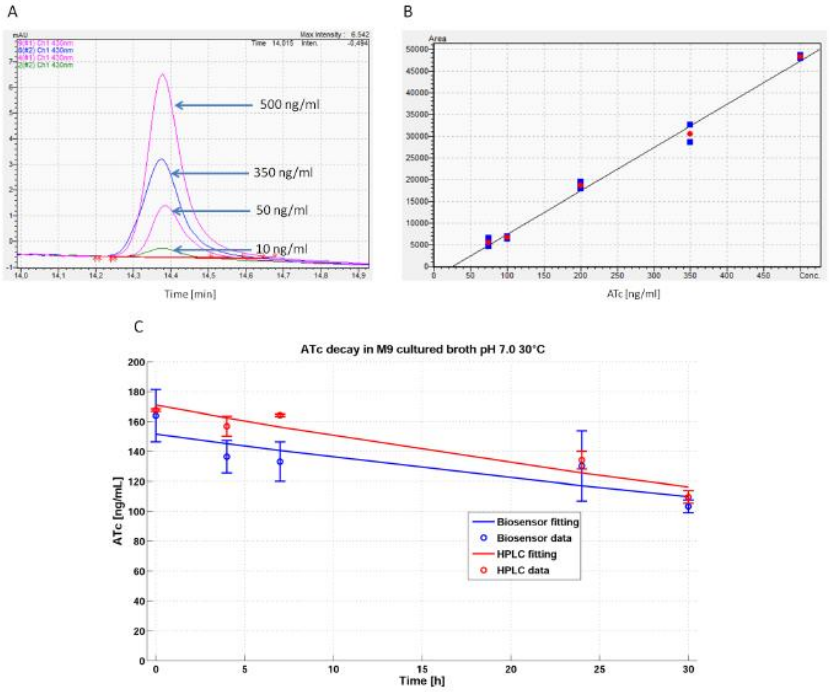


Figure A.10: **aTc measurements via HPLC.** Panel A shows the specific peak of aTc, starting at the retention time of 14.26 min, for samples with different inducer concentrations. No detectable peak is observed in samples without aTc (data not shown). Panel B shows the standard calibration curve, expressing aTc concentration vs aTc peak area, and the corresponding regression line (R2>0.99). Red circles are the average values and blue squares indicate the 95% confidence intervals of the average value of three independent measurements. The linearity range lower limit of the curve is 50 ng/mL (data not shown). Samples of panels A and B were prepared by adding known amounts of aTc to sterile M9 at pH 7.0. Panel C shows the result of an aTc decay experiment in M9 cultured broth at 30°C, pH 7.0, when measured via HPLC (red) or aTc biosensor (blue). Circles represent the average values and error bars are the 95% confidence intervals of the average values of two independent experiments. Lines represent the fitted first-order exponential model curves. Data and fitted curves are shown in absolute units (ng/mL of aTc). Results show good agreement between HPLC and biosensor measurements (CV of 8%) and fitted decay rates (k=0.013±0.004 for HPLC and k=0.011±0.006 for biosensor).

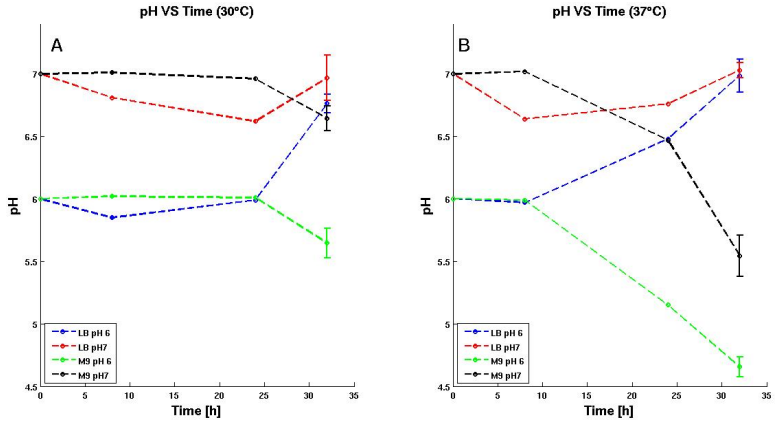


Figure A.11: **pH of cultured broths.** Because the presence of a growing bacterial population can affect the pH of the medium, pH measurements were taken at specific time points in cultured broth experiments. Data points represent the measured pH value of a representative experiment or the average value of 8-10 replicates. In the latter case, error bars represent their 95% confidence intervals. Panel A: growth at 30°C; panel B: growth at 37°C.

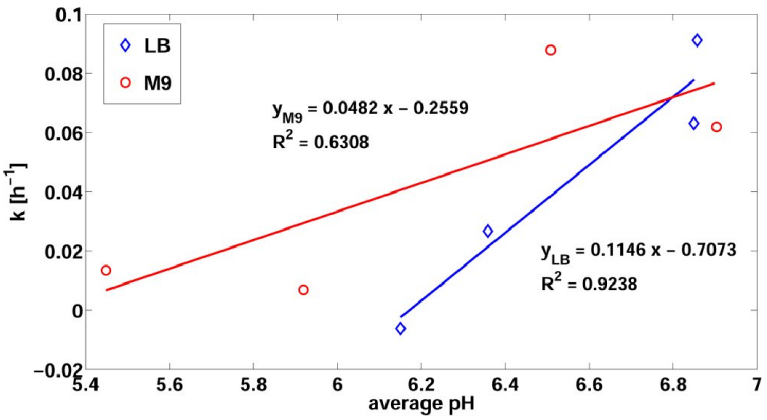


Figure A.12: **3OC$_6$-HSL degradation rate (k) as a function of the average pH of bacterial culture in LB or M9.** Data points represent the estimated decay rates as a function of pH, taken from Tab. A.1 and Fig. A.11, respectively. Solid lines represent data fitting via regression line with the indicated equation and R^2.

A.3 Conclusions

The decay rates of the commonly used chemical inducers IPTG, aTc and $3OC_6$-HSL have been studied in different conditions. Temperature has been found to be the major factor affecting aTc, increasing degradation rate by 2-fold from 30°C to 37°C. Medium and pH have been found to highly affect $3OC_6$-HSL in sterile broth, with a significant interaction: while no significant degradation occurs in LB pH 6.0 over a 32-h experiment, the increase of pH from 6.0 to 7.0 or the use of M9 instead of LB yield a significant degradation, which further increases in the M9 pH 7.0 condition. Data from cultured broth experiments qualitatively confirmed the medium and pH contribution. This study of $3OC_6$-HSL degradation confirms the important contribution of pH previously found in literature [76, 77]. Temperature was previously identified as a key factor for $3OC_6$-HSL decay, yielding a significantly higher degradation rate at 37°C than at 22°C [76], but in the temperature conditions used here (30°C and 37°C), no difference in degradation was detected. Finally, no significant degradation was observed for IPTG among the tested conditions. The knowledge of degradation rate in several contexts can support the rational design of synthetic biological systems by improving the predictability of induction effects, especially for prolonged experiments.

Bibliography

[1] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[2] S. Cohen, A. Chang, H. Boyer, and R. Helling. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences*, 70(11):3240–3244, 1973.

[3] F. H. C. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[4] R. Kitney, J. Calvert, R. Challis, J. Cooper, A. Elfick, P. S. Freemont, J. Haseloff, M. Kelly, and L. Paterson. Synthetic Biology: scope, applications and implications. Technical report, The Royal Academy of Engineering, 2009.

[5] S. A. Benner and A. M. Sismour. Synthetic Biology. *Nature Reviews Genetics*, 6(7):533–543, 2005.

[6] L. Serrano. Synthetic Biology: promises and challenges. *Molecular Systems Biology*, 3(1), 2007.

[7] E. Adrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic Biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2(1), 2006.

[8] M. Heinemann and S. Panke. Synthetic Biology: putting engineering into biology. *Bioinformatics*, 22(22):2790–2799, 2006.

[9] P. Chopra and A. Kamma. Engineering life through Synthetic Biology. *In silico biology*, 6(5):401–410, 2006.

[10] P. E. M. Purnick and R. Weiss. The second wave of Synthetic Biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, 2009.

[11] F. Moser, N. J. Broers, S. Hartmans, A. Tamsir, R. Kerkman, J. A. Roubos, R. Bovenberg, and C. A. Voigt. Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synthetic Biology*, 1(11):555–564, 2012.

[12] C. A. Voigt. Genetic parts to program bacteria. *Current opinion in biotechnology*, 17(5):548–557, 2006.

[13] D. Endy. Foundations for engineering biology. *Nature*, 438(7067):449–453, 2005.

[14] C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, and J. D. Newman. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 2013.

[15] F. Zhang, J. M. Carothers, and J. D. Keasling. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature biotechnology*, 30(4):354–359, 2012.

[16] C. R. Fischer, D. Klein-Marcuschamer, and G. Stephanopoulos. Selection and optimization of microbial hosts for biofuels production. *Metabolic engineering*, 10(6):295–304, 2008.

[17] C. Bi, P. Su, J. Müller, Y. C. Yeh, S. R. Chhabra, H. R. Beller, S. W. Singer, and N. J. Hillson. Development of a broad-host Synthetic Biology toolbox for Ralstonia eutropha and its application to engineering hydrocarbon biofuel production. *Microbial cell factories*, 12(1):107, 2013.

[18] J. A. Gimpel, E. A. Specht, D. R. Georgianna, and S. P. Mayfield. Advances in microalgae engineering and Synthetic Biology applications for biofuel production. *Current opinion in chemical biology*, 2013.

[19] Y. Y. Chen, K. E. Galloway, and C. D. Smolke. Synthetic Biology: advancing biological frontiers by building synthetic systems. *Genome biology*, 13(2):240, 2012.

[20] http://openwetware.org.

[21] H. M. Salis, E. A. Mirsky, and C. A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology*, 27(10):946–950, 2009.

[22] http://parts.igem.org/mainpage.

[23] T. F. Jr Knight. Idempotent vector design for standard assembly of biobricks. Technical report, MIT Artificial Intelligence Laboratory, 2003.

[24] R. P. Shetty, D. Endy, and T. F. Jr Knight. Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering*, 2(1):1–12, 2008.

[25] B. Canton, A. Labno, and D. Endy. Refinement and standardization of synthetic biological parts and devices. *Nature biotechnology*, 26(7):787–793, 2008.

[26] J. R. Kelly, A. J. Rubin, J. H. Davis, C. M. Ajo-Franklin, J. Cumbers, M. J. Czar, K. de Mora, A. L. Glieberman, D. L. Monie, and D. Endy. Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of biological engineering*, 3(1):4, 2009.

[27] J. H. Davis, A. J. Rubin, and R. T. Sauer. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic acids research*, 39(3):1131–1141, 2011.

[28] H. M. Salis. The ribosome binding site calculator. *Methods in Enzymology*, 498:19–42, 2011.

[29] G. Cambray, J. C. Guimaraes, V. K. Mutalik, C. Lam, Q. A. Mai, T. Thimmaiah, J. M. Carothers, A. P. Arkin, and D. Endy. Measurement and modeling of intrinsic transcription terminators. *Nucleic acids research*, 41(9):5139–5148, 2013.

[30] L. Pasotti, N. Politi, S. Zucca, M. G. Cusella De Angelis, and P. Magni. Bottom-up engineering of biological systems through standard bricks: A modularity study on basic parts and devices. *PloS one*, 7(7):e39407, 2012.

[31] H. M. Sauro. Modularity defined. *Molecular systems biology*, 4(1), 2008.

[32] S. Jayanthi, K. S. Nilgiriwala, and D. Del Vecchio. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synthetic Biology*, 2013.

[33] R. Kwok. Five hard truths for Synthetic Biology. *Nature*, 463(7279):288–290, 2010.

[34] K. H. Kim and H. M. Sauro. Measuring the degree of modularity from gene expression noise in gene regulatory circuits. *arXiv preprint arXiv:0910.5522*, 2009.

[35] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modular cell biology: retroactivity and insulation. *Molecular systems biology*, 4(1), 2008.

[36] D. Del Vecchio. A control theoretic framework for modular analysis and design of biomolecular networks. *Annual Reviews in Control*, 37(2):333–345, 2013.

[37] K. H. Kim, D. Chandran, and H. M. Sauro. Toward modularity in Synthetic Biology: Design patterns and fan-out. In *Design and Analysis of Biomolecular Circuits*, pages 117–138. Springer, 2011.

[38] K. H. Kim and H. M. Sauro. Fan-out in gene regulatory networks. *Journal of biological engineering*, 4:16, 2010.

[39] J. B. Lucks, L. Qi, W. R. Whitaker, and A. P. Arkin. Toward scalable parts families for predictable design of biological circuits. *Current opinion in microbiology*, 11(6):567–573, 2008.

[40] D. Sprinzak and M. B. Elowitz. Reconstruction of genetic circuits. *Nature*, 438(7067):443–448, 2005.

[41] D. Chandran, W. B. Copeland, S. C. Sleight, and H. M. Sauro. Mathematical modeling and Synthetic Biology. *Drug Discovery Today: Disease Models*, 5(4):299–309, 2009.

[42] F. Ceroni, S. Furini, A. Stefan, A. Hochkoeppler, and E. Giordano. A synthetic post-transcriptional controller to explore the modular design of gene circuits. *ACS Synthetic Biology*, 1(5):163–171, 2012.

[43] J. C. Anderson, C. A. Voigt, and A. P. Arkin. Environmental signal integration by a modular and gate. *Molecular systems biology*, 3(1), 2007.

[44] B. Wang, R. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like Synthetic Biology. *Nature communications*, 2:508, 2011.

[45] V. K. Mutalik, J. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q. A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature methods*, 10(4):354–360, 2013.

[46] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, 2012.

[47] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

[48] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[49] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*, 64(6):2240–2246, 1998.

[50] J. Stricker, S. Cookson, M. R. Bennet, W. H. Mather, L. S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.

[51] T. Danino, O. Mondragon Palomino, L. Tsimring, and J. Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.

[52] S. Basu, R. Mehreja, S. Thiberge, M. T. Chen, and R. Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6355–6360, 2004.

[53] S. Zucca, L. Pasotti, G. Mazzini, M. G. Cusella De Angelis, and P. Magni. Characterization of an inducible promoter in different DNA copy number conditions. *BMC bioinformatics*, 13(Suppl 4):S11, 2012.

[54] A. P. Burgard, P. Pharkya, and C. D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6):647–657, 2003.

[55] W. J. Holtz and J. D. Keasling. Engineering static and dynamic control of synthetic pathways. *Cell*, 140(1):19–23, 2010.

[56] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular systems biology*, 2(1), 2006.

[57] F. Menolascina, M. di Bernardo, and D. di Bernardo. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica*, 47(6):1265–1270, 2011.

[58] M. Dragosits, D. Nicklas, and I. Tagkopoulos. A Synthetic Biology approach to self-regulatory recombinant protein production in Escherichia coli. *Journal of biological engineering*, 6:2, 2012.

[59] L. You, R. S. Cox III, R. Weiss, and F. H. Arnold. Programmed population control by cell-cell communication and regulated killing. *Nature*, 428(6985):868–871, 2004.

[60] J. Sambrook and D. Russell. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, third edition, 2001.

[61] Y. H. Dong, J. L. Xu, X. Z. Li, and L. H. Zhang. Aiia, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of Erwinia carotovora. *Proceedings of the National Academy of Sciences*, 97(7):3526–3531, 2000.

[62] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research*, 25(6):1203–1210, 1997.

[63] F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of Escherichia coli k-12. *Science*, 277(5331):1453–1462, 1997.

[64] R. S. Cox III, M. G. Surette, and M. B. Elowitz. Programming gene expression with combinatorial promoters. *Molecular systems biology*, 3(1), 2007.

[65] S. C. Sleight, B. A. Bartley, J. A. Lieviant, and H. M. Sauro. Designing and engineering evolutionary robust genetic circuits. *Journal of biological engineering*, 4(1):12, 2010.

[66] M. J. Waites, N. L. Morgan, J. S. Rockey, and G. Higton. *Industrial microbiology: an introduction*. Wiley. com, 2009.

[67] G. Jach, M. Pesch, K. Richter, S. Frings, and J. F. Uhrig. An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nature methods*, 3(8):597, 2006.

[68] N. Politi, L. Pasotti, S. Zucca, M. Casanova, G. Micoli, M. G. Cusella De Angelis, and P. Magni. Half-life measurements of chemical inducers for recombinant gene expression. *Journal of biological engineering*, 2014.

[69] P. Bernard and M. Couturier. Cell killing by the f plasmid ccdb protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology*, 226(3):735–745, 1992.

[70] T. Miki, J. Ae Park, K. Nagao, N. Murayama, and T. Horiuchi. Control of segregation of chromosomal DNA by sex factor f in Escherichia coli: Mutants of DNA gyrase subunit a suppress letD (ccdB) product growth inhibition. *Journal of molecular biology*, 225(1):39–52, 1992.

[71] T. Durfee, R. Nelson, S. Baldwin, G. Plunkett, V. Burland, B. Mau, J. F Petrosino, X. Qin, D. M. Muzny, M. Ayele, R. A. Gibbs, B. Csorgo, G. Posfai, G. M. Weinstock, and F. R. Blattner. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *Journal of bacteriology*, 190(7):2597–2606, 2008.

[72] K. Tsuji and J. F. Goetz. HPLC as a rapid means of monitoring erythromycin and tetracycline fermentation processes. *The Journal of antibiotics*, 31(4):302–308, 1978.

[73] A. Fernández-Castané, C. E. Vine, G. Caminal, and J. López-Santín. Evidencing the role of lactose permease in IPTG uptake by Escherichia coli in fed-batch high cell density cultures. *Journal of Biotechnology*, 157(3):391–398, 2012.

[74] A. Fernández, J. Ruiz, G. Caminal, and J. López-Santín. Development and validation of a liquid chromatography-mass spectrometry assay for the quantitation of IPTG in E. coli fed-batch cultures. *Analytical chemistry*, 82(13):5728–5734, 2010.

[75] A. Fernández-Castané, G. Caminal, and J. López-Santín. Direct measurements of IPTG enable analysis of the induction behavior of E. coli in high cell density cultures. *Microbial cell factories*, 11(1):1–9, 2012.

[76] E. A. Yates, B. Philipp, C. Buckley, S. Atkinson, S. R. Chhabra, R. E. Sockett, M. Goldner, Y. Dessaux, M. Cámara, H. Smith, and P. Williams. N-acylhomoserine lactones undergo lactonolysis in a pH-, temperature-,

and acyl chain length-dependent manner during growth of Yersinia pseudotuberculosis and Pseudomonas aeruginosa. *Infection and immunity*, 70(10):5635–5646, 2002.

[77] A. L. Schaefer, B. L. Hanzelka, M. R. Parsek, and E. P. Greenberg. Detection, purification, and structural elucidation of the acylhomoserine lactone inducer of Vibrio fischeri luminescence and other related molecules. *Methods in enzymology*, 305:288–301, 2000.

List of publications

Articles in peer reviewed journals

- N. Politi, L. Pasotti, S. Zucca, M. Casanova, G. Micoli, M. G. Cusella De Angelis and P. Magni. Half-life measurements of chemical inducers for recombinant gene expression. *Journal of Biological Engineering*, Accepted

- S. Zucca, L. Pasotti, N. Politi, M. G. Cusella De Angelis and P. Magni. A standard vector for the chromosomal integration and characterization of BioBricks™ in *Escherichia coli*. *Journal of Biological Engineering*, 7(1):1-13, 2013

- L. Pasotti, N. Politi, S. Zucca, M. G. Cusella De Angelis and P. Magni. Bottom-Up Engineering of Biological Systems through Standard Bricks: A Modularity Study on Basic Parts and Devices. *PLoS ONE*, 7(7):e39407, 2012.

Contributions to conference proceedings

- N. Politi, S. Zucca, L. Pasotti, M. G. Cusella De Angelis, P. Magni. Modelling and implementation of a negative feedback mechanism in *E. coli*, *Proceedings of Synthetic Biology 6.0*. London, UK, July 09-11, 2013, The BioBricks Foundation

- L. Pasotti, N. Politi, S. Zucca, M. Casanova, M. G. Cusella De Angelis, P. Magni. Interconnection of quantitatively characterized genetic devices to engineer predictable biological functions. *Proceedings of Synthetic Biology 6.0*, London, UK, July 09-11, 2013, The BioBricks Foundation

- S. Zucca, L. Pasotti, N. Politi, M. G. Cusella De Angelis, P. Magni. In vivo and in silico approaches to study the effects of copy number variation on gene networks behaviour. *Proceedings of Synthetic Biology 6.0*, London, UK, July 09-11, 2013, The BioBricks Foundation

- N. Politi, S. Zucca, L. Pasotti, M. G. Cusella De Angelis and P. Magni, Controlling $3OC_6$-HSL concentration by a negative feedback-loop genetic circuit in *E. coli*. *Congresso Nazionale di Bioingegneria - Atti*. Roma, IT, June 26-29, 2012

- L. Pasotti, N. Politi, S. Zucca, M. G. Cusella De Angelis and P. Magni. Studio quantitativo di sistemi modello per validare la modularità di componenti in biologia sintetica. *Congresso Nazionale di Bioingegneria - Atti*. Roma, IT, June 26-29, 2012

- G. Bertoni, N. Politi, S. Zucca, F. Sampietro, E. Baldini, D. Bianchini, N. Franceschi, V. Ghio, T. Goggia, E. Pasi, D. Sartori, M. G. Cusella De Angelis and P. Magni. CTRL + E.: Signalling is Nothing Without Control. *Proceedings of iGEM 2011, Regional Jamboree*. Amsterdam, NL, October 1-2, 2011, The iGEM Foundation

- N. Politi, L. Pasotti, M. G. Cusella De Angelis, P. Magni. Steps Towards the Design of Complex Quorum Sensing Networks Through Standard Biological Parts:Characterization of a Synthetic LuxR-HSL Repressible Promoter. *Proceedings of BITS 2011*. Pisa, IT, June 20-22, 2011, p. 130-131, PISA:Edizioni Plus

- N. Politi, L. Pasotti, S. Zucca, M. Stuppia, M. G. Cusella De Angelis, P. Magni. Characterization of a synthetic luxR-HSL repressible promoter for the design of novel activation-repression circuit in *E. coli*. *Proceedings of Synthetic Biology 5.0*. Stanford, USA, June 15-17, 2011, The BioBricks Foundation

- S. Zucca, L. Pasotti, N. Politi, G. Zambianchi, M. G. Cusella De Angelis and P. Magni. Characterization of promoter strength varying the DNA copy number from one to hundreds of copy. *Proceedings of Synthetic Biology 5.0*. Stanford, USA, June 15-17, 2011, The BioBricks Foundation

- L. Pasotti, S. Zucca, M. Meroso, G. Zambianchi, N. Politi, M. G. Cusella De Angelis and P. Magni. Engineering *E. coli* chromosome through BBa_K300000, a BioBrick integrative base vector. *Proceedings of Synthetic Biology 5.0*. Stanford, USA, June 15-17, 2011, The BioBricks Foundation

- S. Zucca, L. Pasotti, G. Zambianchi, M. Lupotto, M. Meroso, R. Barini, S. Bisi, N. Politi, A. Ranieri, F. Sampietro, C. Tassorelli, M. Zocco, M. G. Cusella De Angelis and P. Magni. ProteInProgress. *Proceedings of iGEM 2010*. Boston, USA, November 6-8, 2010, The iGEM Foundation

Ringraziamenti

Giunto alla fine di questi tre intensi anni di Dottorato di Ricerca, vorrei ringraziare chi ha reso possibile la realizzazione e il compimento di questo percorso. In primis il Prof. Paolo Magni che, oltre ad avermi seguito nelle due precedenti fatiche della tesi triennale e di quella specialistica, mi ha dato la preziosa opportunità di poter ulteriormente arricchirmi e crescere durante questo periodo, occupandomi di una tematica stimolante come la Biologia Sintetica.
Vorrei, inoltre, ringraziare la Prof. ssa Maria Gabriella Cusella De Angelis per essere stata sempre disponibile ed aver reso possibile concretamente lo svolgimento del mio Dottorato e delle attività di ricerca che porto avanti con i miei colleghi, fornendoci gli spazi e le attrezzature necessari.
Un doveroso grazie va a tutti i Professori, in particolare al Prof. Riccardo Bellazzi. Un ricordo particolare, inoltre, per il Prof. Mario Stefanelli che, con le sue parole, i suoi insegnamenti e i suoi aneddoti, ha sempre rappresentato una fonte di ispirazione.
Non posso, ora, non citare i miei colleghi che, più che colleghi, ormai reputo veri amici. Un grosso ringraziamento va sicuramente a loro, Susanna, Michela e Lorenzo, che non mi hanno fatto mai mancare il loro importante sostegno e apporto, rendendo quest'esperienza positiva anche dal punto di vista umano. Grazie a Susanna e a Lorenzo per le "tirate d'orecchie" e a Michela per avermi risparmiato qualche levataccia.
Non posso dimenticarmi di Gabriele e della Dott. ssa Laura Benedetti, che in questi anni, "vissuti" in laboratorio, sono stati sempre molto disponibili e precisi ogni qual volta ho avuto bisogno di aiuto.
Un saluto va anche a chi ha condiviso con me i cinque anni di Università e ai ragazzi che, in questi anni, hanno partecipato ai progetti iGEM, in special modo a Federica, Riccardo e Tommaso ("mio" unico tesista).
In questi tre anni ho avuto anche l'opportunità di vivere e confrontarmi con i colleghi del Laboratorio di Informatica Biomedica, grazie ai quali ho vissuto

bellissime esperienze sia in ambito Universitario che extra-Universitario. Un grazie particolare va a Nadia per le nostre numerosissime chiacchierate.

Infine, un grazie a chi c'è sempre stato: a tutti i miei amici che in questi anni mi hanno aiutato regalandomi momenti di puro divertimento per alleggerire i periodi più stressanti. A mia zia Laura e mio zio Davide, ai miei cugini Greta e Matteo, che sono per me come fratelli.
E un grazie a chi mi ha sempre sostenuto, appoggiato e guidato nelle mie scelte, dandomi la possibilità concreta di arrivare fino a qui: la mia famiglia.