The pursuit of truth and beauty
is a sphere of activity in which
we are permitted to remain
children all our lives.

A. Einstein

Abstract (English)

Synthetic Biology is a new emerging discipline, which lies between life sciences and engineering and introduces the principles of abstraction, modularity and standardization in the biology world. The application of such principles allows the design of standard biological systems to program living cells for the realization of a desired function. These systems consist of DNA sequences, which can be properly combined to realize the genetic instructions for cell behaviour customization. The cell can execute these instructions, thus implementing the desired function.
Every DNA component can be regarded as a 'basic part', which can be physically combined with other basic parts to create the final system. The BioBrick™ standard assembly has been proposed as a standard procedure to construct complex circuits from basic parts. Each part behaves as a biological 'brick' for the design of complex genetic programs. BioBrick™ parts are collected in an open access repository, called the Registry of Standard Biological Parts.
According to the engineering principles applied to Synthetic Biology, for each part an extensive documentation on its functioning should be provided, in order to enable the rational design of complex circuits on the basis of the information available for individual modules.
Within this frame, this thesis work aims at studying the effects of copy number variation on the behaviour of gene networks, with the aid of both experimental data and computational tools. These effects deeply impact on the design and fine tuning process of biological systems to such an extent that copy number variation can be considered as a control knob to be regulated in order to achieve the desired behaviour.
These investigations require the use of synthetic devices able to maintain the studied circuits at a given number of copies. Commonly used plasmids present in the Registry of Standard Biological Parts enable a 40-fold copy number variation (from 5 copies per cell for low copy vectors to $\sim$ 200 copies per cell for high copy vectors) and plasmid-free strains with the desired genetic circuit integrated in single copy in the genome can be constructed by using, as an example, the procedures described in this thesis work. In order to avoid money- and time-consuming trial-and-error approaches for the fine tuning of

the gene network copy number, mathematical models can be used to simulate the effects of this variation *in silico*.
All these aspects are investigated and discussed in this thesis work.
In Chapter 1 the current status of Synthetic Biology is presented. Functional modules are described and the standard assembly and measurement procedures are illustrated. The importance of plasmids, which are the main genetic tool for copy number regulation, is introduced. The role of mathematical modelling in the design and fine tuning of gene networks is discussed.
In Chapter 2 two case studies are presented to highlight the nonlinearities of system behaviour due to DNA copy number. First, a modularity study on basic parts, then the characterization of a synthetic bacterial lysis device highlight copy number-related effects which affect the predicability of system behaviour.
In Chapter 3 the design and characterization of a BioBrick™-compatible integrative vector for *Escherichia coli* is presented. This tool allows the construction of gene networks present in a single copy in the bacterial genome and has been widely utilized in this thesis work. Quantitative analysis of the activity of different constitutive promoters is then reported in the genomic and plasmid contexts.
In Chapter 4 the quantitative characterization of a simple inducible device in four different copy number contexts is presented. An empiric mathematical model is here proposed, to fully characterize the device in all the investigated contexts.
In Chapter 5 a mechanistic mathematical model of the simple inducible system is proposed, in order to study the effects of copy number variation of both promoter and transcription factor, individually and in concert.
In Chapter 6 the overall conclusions of this thesis work are drawn.

Abstract (Italian)

La Biologia Sintetica è una nuova ed emergente disciplina, che si colloca all'interfaccia tra le scienze della vita e l'ingegneria ed è responsabile dell' introduzione dei principi di astrazione, modularità e standardizzazione nel mondo della biologia. L'applicazione di questi principi consente la progettazione razionale di componenti biologici standard, utilizzabili per programmare l' attività cellulare al fine di realizzare le funzioni desiderate. Tutti i componenti standard sono costituiti da sequenze di DNA, opportunamente combinabili per realizzare istruzioni genetiche per ottenere un comportamento cellulare definito dall'utente. La cellula esegue tali istruzioni, implementando cosí la funzione desiderata.
Ogni sequenza di DNA può essere considerata come una 'parte di base' e può essere fisicamente combinata con altre parti per creare il sistema finale. Lo standard BioBrick™ è stato proposto come procedura standard per la costruzione di circuiti complessi a partire dalle parti di base. Ogni parte si comporta come un 'mattone' biologico all'interno di un circuito genetico più complesso. Tutte le parti BioBrick™ sono raccolte in una banca dati *open source*, chiamata Registry of Standard Biological Parts.
In accordo con i principi dell'ingegneria applicati alla Biologia Sintetica, per ogni parte deve essere fornita una estensiva documentazione quantitativa, al fine di permettere il progetto razionale di un circuito complesso sulla base delle informazioni disponibili per ogni modulo semplice. All'interno di questo contesto, questo lavoro di tesi si propone di studiare gli effetti della variazione del numero di copie di DNA sul comportamento di una rete genica, utilizzando un approccio sia sperimentale che modellistico. Tali effetti condizionano a tal punto il processo di progettazione e regolazione fine di circuiti sintetici che il numero di copie può essere considerato come una manopola di regolazione per ottenere il comportamento desiderato.
Gli studi in esame richiedono l'utilizzo di strumenti sintetici capaci di mantenere il circuito genetico ad un numero di copie desiderato. Nel Registry of Standard Biological Parts sono presenti dei plasmidi, comunemente utilizzati per la propagazione di circuiti sintetici, che consentono di mantenere un numero di copie che varia di circa 40 volte (da 5 per i plasmidi a basso numero di

copie a $\sim$ 200 per i plasmidi ad alto numero). Ceppi batterici privi di plasmidi con il costrutto desiderato integrato in singola copia nel genoma possono essere costruiti utilizzando, ad esempio, le procedure descritte in questa tesi.
Al fine di evitare lenti e dispendiosi approcci di tipo trial-and-error per la regolazione fine del numero di copie di DNA di un circuito, possono essere utilizzati dei modelli matematici per la predizione *in silico* degli effetti imputabili a tale variazione.
Tutti gli aspetti legati alla variazione del numero di copie fin qui illustrati sono stati studiati e vengono discussi in questo lavoro di tesi.
Nel Capitolo 1 è illustrato lo stato attuale delle ricerche nel campo della Biologia Sintetica. Vengono presentate le parti funzionali più comunemente utilizzate nella creazione di circuiti sintetici e vengono illustrati gli standard fisici di assemblaggio e gli standard di misura. Vengono introdotti i plasmidi, ovvero lo strumento genetico più diffuso per la regolazione del numero di copie di DNA di una rete genica. è inoltre discusso il ruolo della modellistica matematica nel progetto e nella regolazione fine di reti geniche sintetiche.
Nel Capitolo 2 vengono presentati due casi di studio che evidenziano non-linearità sull'uscita del sistema dovute al numero di copie di DNA. Sia uno studio sulla modularità di parti semplici che la caratterizzazione quantitativa di un dispositivo per la lisi batterica hanno evidenziato effetti dovuti al numero di copie di DNA sulla predicibilità del comportamento del sistema.
Nel Capitolo 3 viene illustrato il progetto e la caratterizzazione di un vettore integrativo, compatibile con lo standard BioBrick™, per il batterio *Escherichia coli*. Questo vettore è uno strumento per la costruzione di circuiti sintetici presenti in singola copia nel genoma batterico ed è stato ampiamente utilizzato in questo lavoro di tesi. L'analisi dell'attività di diversi promotori costitutivi sia in un contesto genomico che su plasmide è poi riportata.
Nel Capitolo 4 viene discussa la caratterizzazione quantitativa di un semplice dispositivo inducibile in quattro diverse condizioni di numero di copie di DNA. Un modello matematico empirico è stato utilizzato per caratterizzare appieno il funzionamento del dispositivo biologico nelle condizioni studiate.
Nel Capitolo 5 viene introdotto un modello meccanicistico del comportamento del dispositivo inducibile, che consente di studiare l'effetto della variazione, sia individuale che in sincrono, del numero di copie del fattore di trascrizione e del promotore.
Infine, nel Capitolo 6 sono riportate le conclusioni generali di questo lavoro di tesi.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Synthetic Biology

1.1 State of the art

Synthetic biology is an emergent discipline that combines the investigative nature of biology with the constructive nature of engineering [1]. Due to its relevance, this new research field could dramatically influence other scientific and engineering disciplines, as well as impact different aspects of daily life and society [2].
The conventional approach of genetic engineering to face complex problems typically involves the tweak of one or a few genes. Synthetic biology, by contrast, focuses on widespread modifications of the existing cellular architecture and on the design of elaborate systems from the ground up [1], encouraged by progress in the design of artificial gene networks, *de novo* DNA synthesis and protein engineering [3].
Even though scientific community is in accordance in defining the boundaries of Synthetic Biology, there is still a degree of misunderstanding about the true nature of the field [4]. The areas pertaining to Synthetic Biology can be circumscribed to four great knowledge domains: genetic circuitry, minimal genomes (such as *Mycoplasma mycoides*), protocells and chemical based Synthetic Biology (such as XNA) [5].
In the area of genetic circuits, bio-engineering can provide useful contributions by applying the key concepts of abstraction, modularity and standardization to cell biology and genetic engineering, thus coming to the rational design of gene networks [5]. The goal of Synthetic Biology, at least in this acception, is to extend or modify the behaviour of organisms and engineer them to perform new user-defined tasks.
One useful analogy to conceptualize both the goal and methods of Synthetic Biology is the computer engineering hierarchy illustrated in Fig. 1.1.

Within the hierarchy, every constituent part is embedded in a more com-

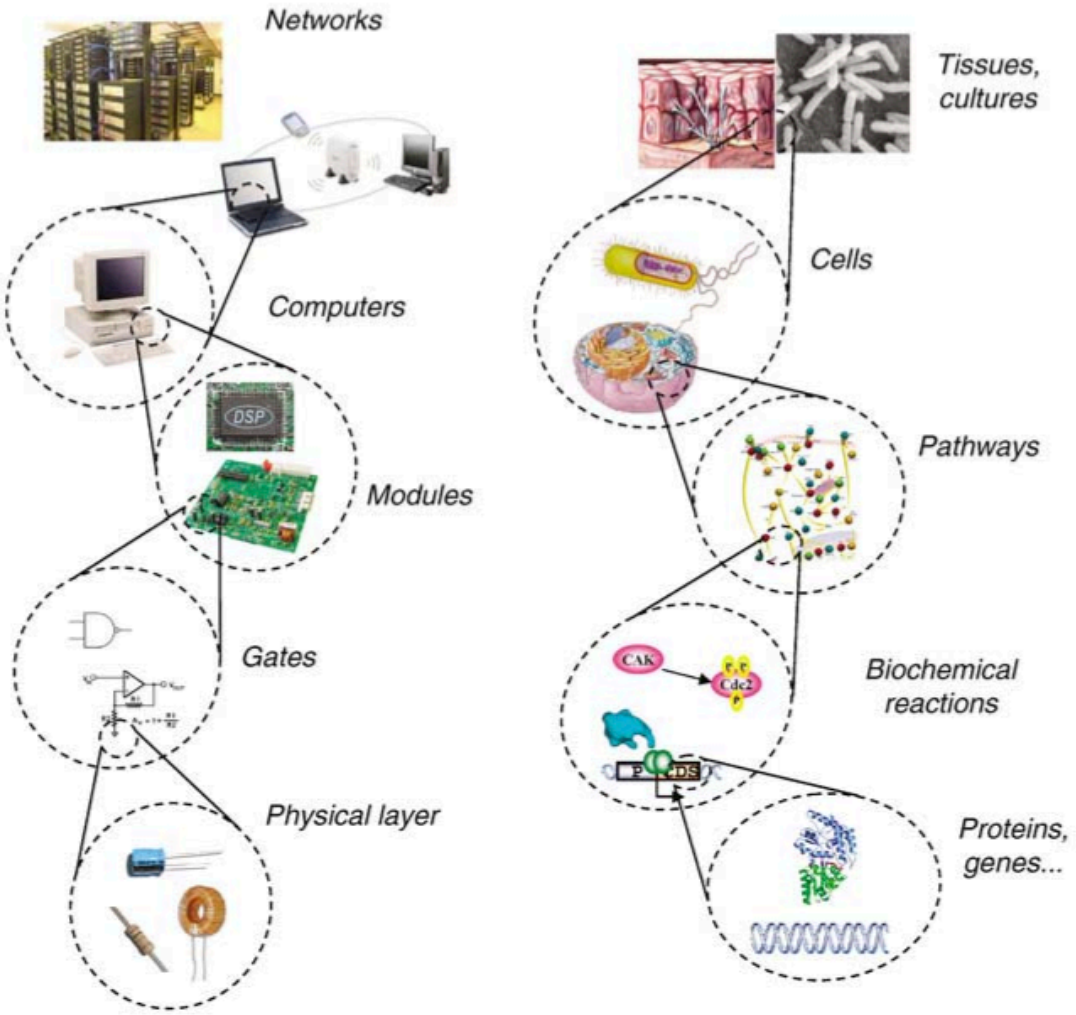


Figure 1.1: **A possible hierarchy for Synthetic Biology is inspired by computer engineering** Andrianantoandro et al. [2]

plex system that defines its working context. The design and realization of new functions is implemented using a bottom-up strategy. At the lower level of the hierarchy, DNA, RNA, proteins, and metabolites (including lipids and carbohydrates, amino acids, and nucleotides) are present. This level is analogous to the physical layer of transistors, capacitors and resistors in computer engineering. The device layer includes biochemical reactions that regulate the flow of information and manipulate physical processes. It is comparable to engineered logic gates that perform computations in a computer. The module layer includes libraries of biological devices, that can be combined to build-up complex pathways, functioning like integrated circuits. The inter-connection of these modules and the incorporation into a host organism allow the potentiation or modification of cell behaviour in a user-defined fashion.
Independently operating engineered cells can perform tasks of increasing complexity. Populations of communicating cells can implement even more complicated functions, like different computers can be connected in networks [2]. This analogy highlights the importance of modularity of a system. It can be achieved if its components can be functionally separated and recombined in new system. Complexity arises from the interactions between modules at different levels of organizational hierarchy [6].

1.2 Functional modules in Synthetic Biology

In Synthetic Biology, a *basic part* is the simplest element of a gene regulatory network. Commonly used basic parts are promoters, ribosome binding sites, coding sequences, transcriptional terminators and plasmids (Fig. 1.2).

At its simplest, a basic part is a single functional unit that cannot be divided further into smaller functional units.
Basic parts can be assembled together to make longer, more complex composite parts, called *devices*, that are collections of components able to implement a specific user-defined function in living cells, such as the expression of a recombinant protein if a given molecule is present in the extra-cellular environment. Different devices can be functionally connected to form *systems*, able to implement high-level tasks, such as the expression of different proteins at different pre-determined time intervals. These systems work when incorporated in host cells, that provide their native transcriptional and translational machinery to implement the artificially introduced function. Most commonly used *molecular chassis* to host artificial systems are gram-negative bacteria, gram-positive bacteria and yeasts. The most widespread organism used in Synthetic Biology is the gram-negative bacterium *Escherichia coli*. Hereinafter, the discussion will be focused on this host organism, but the key concepts can be generalized to different chassis.
In the following subsections, commonly used basic parts are described.

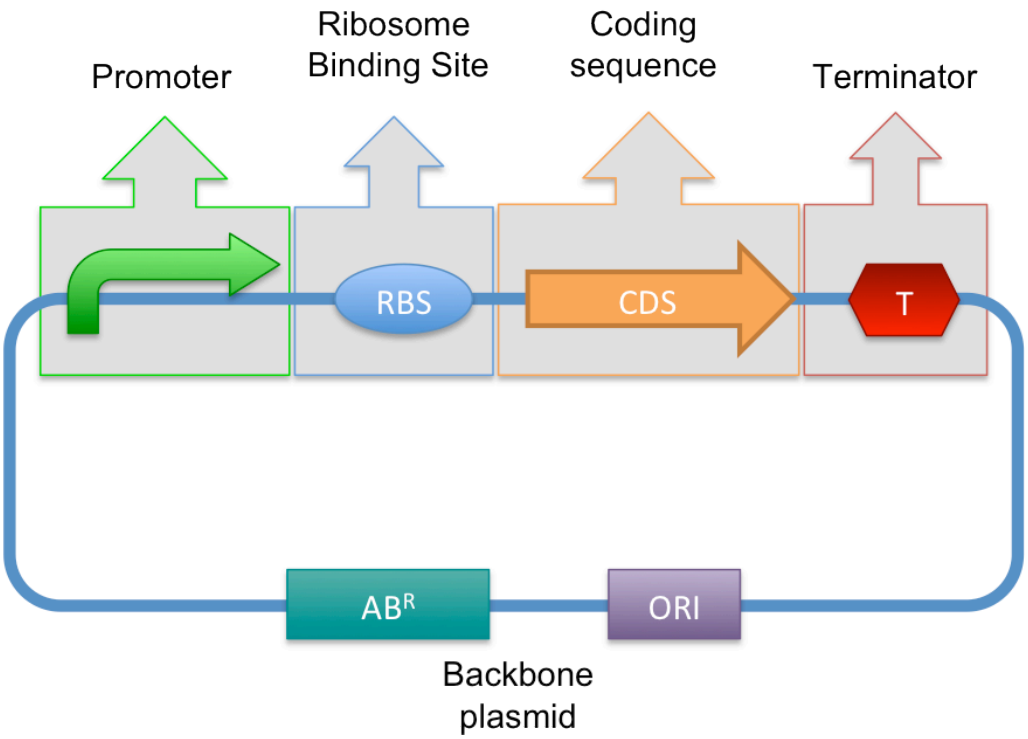


Figure 1.2: **Basic parts in Synthetic Biology** The plasmid backbone (blue line) bears a selection marker (AB^R, antibiotic resistance) and an origin of replication (ORI), which ensure the plasmid maintenance and propagation in the host cell. The grey-shaded components represent a protein-generator circuit, composed by a promoter (green),a ribosome binding site (blue), a coding sequence (orange) and a terminator (red).

1.2.1 Promoters

Promoters are DNA sequences which affect the frequency and location of transcription initiation by recruiting the cell transcriptional machinery and by leading to transcription of the downstream DNA sequence. The specific sequence of the promoter determines its affinity to the RNA polymerase and thus its "transcriptional strength" (i.e. a "strong" promoter leads to a high rate of transcription initiation). Specific sequences can affect the behaviour of promoters in terms of transcriptional strength and regulation. In *E. coli*, the -35' and -10' sequences have a fundamental role in recruiting the transcriptional machinery. "Operator sites" are small DNA sequences responsible for the activation or repression of transcriptional activity, depending on endogenous or exogenous regulator elements. Promoters whose activity can be tuned are known as regulated promoters, as opposed to "constitutive" promoters, which show a constant activity.

1.2.2 Ribosome Binding Sites (RBSs)

A ribosome binding site (RBS) is a RNA sequence, found in the 5' untranslated region (5'-UTR) to which ribosomes can bind and initiate translation. In prokaryotes, RBSs are composed by the Shine-Dalgarno sequence[1]. This sequence is complementary to the 16S ribosomal RNA (rRNA) and is located 3-12 bases upstream of the start codon[2] (the best location has been demonstrated to be 7-9 base pairs upstream). The Shine-Dalgarno base composition dramatically affects the efficiency of RBS in initiating translation. In bacteria, RBSs are effective control elements for translation initiation and thereby protein expression.

1.2.3 Coding Sequences

Protein coding sequences are DNA sequences that begin with a start codon, usually ATG (N-terminus of the corresponding protein) and end with a stop codon, usually a double stop codon TAA TAA (C-terminus of the corresponding protein). Coding sequences are transcribed into mRNA, which is translated into a polypeptide chain in turn. Every nucleotide triplet of a coding sequence, termed "codon", encodes an amino-acid in the polypeptide chain[3].
The N-terminus of a protein can host many features, such as protein export tags, lipoprotein cleavage and attachment tags. These features are indicated as *head domains*, since they occur at the beginning of a coding region.

[1] Shine-Dalgarno consensus sequence exists in bacteria, in archaea and in some chloroplastic and mitochondial transcripts

[2] *E. coli* recognises three different start codons: AUG, GUG, UUG [7]

[3] Different chassis may either map a given codon to a different sequence or may use different codons more or less frequently. Therefore some protein coding sequences may be optimized for use in a particular chassis.

A protein domain is a sequence of amino acids evolutionarily conserved as a unit, which can fold and function independently of the whole protein sequence. The DNA sequence of such domains must maintain *in-frame* translation[4]. Since these protein domains are within a coding sequence, they are called *internal domains*. Certain internal domains play a significant role in protein cleavage or splicing processes and are termed *special internal domains*.
The C-terminal domain of a protein contains at least one stop codon, that causes the translation cessation. Degradation tags are also required to be at the extreme C-terminus and are termed *tail domains*.

1.2.4 Terminators

Terminators are nucleotide sequences that usually occur at the end of a gene or operon and cause transcription cessation. In prokaryotes, terminators can be classified as ρ-independent or ρ-dependent terminators. ρ-independent terminators contain palindromic sequences that form a G-C rich stem loop, followed by a T-rich sequence. The stem loop interacts with the RNA polymerase and provokes its arrest. ρ-dependent terminators participate in sophisticated genetic regulatory mechanisms and cannot be specified by sequence.

1.2.5 Plasmids

Plasmids are circular, double-stranded DNA molecules, usually composed by few thousands of base pairs, able to self-replicate within the cell independently of the chromosomal DNA.
The essential sequences present in a plasmid are:

- an *origin of replication*, which ensures the plasmid maintenance in the cell at a given number of copies. Common origins of replication enable the maintenance of the plasmid from 4-5 copies per cell (low copy number origins) to more than 100 copies per cell (high copy number replication origins).

- a *selection marker*, in order to discriminate plasmid-bearing cells from plasmid-free cells. Commonly used selection markers are genes encoding for antibiotic resistances, which allow the selection of plasmid-bearing cells by supplementing the culture medium with the corresponding antibiotic. Other selection methods, not used in this thesis work, are based on auxotrophies.

- a *cloning site*, which is a sequence in the plasmid that contains a number

[4] *In-frame* translation occurs when the reading frame of the whole protein is not altered by the translation of the protein domain, i.e. when the number of nucleotides that belong to the protein domain is multiple of three.

of unique restriction sites[5]. The desired genetic-encoded functions can be inserted in the plasmid by using, as an example, the procedures described in Sec. 1.3.1, and then incorporated in the host cell.

Plasmid DNA can be easily extracted and purified from cells. It can also be routinely manipulated using common laboratory techniques and inserted into host cells. For this reason, plasmids are widely used in Synthetic Biology as delivery vectors for the desired genetic circuit.

1.3 Standards in Synthetic Biology

The rational design of customized genetic circuits requires an additional step: the quantitative characterization of parts that can be used to compose a system with predictable behaviour.

The standardization of measurement units and characterization methodologies can enable the re-use of BioBrick™ parts reliably. This process is based on the definition of *measurement standards* [8, 9, 10]. In addition, *physical standards* have been proposed to facilitate the construction of genetic circuits. Physical standards define a shareable interface for biological parts, which allows the assembly of these parts in complex circuits by using predetermined assembly schemes. On the other hand, measurement standardization requires the definition of techniques for the quantification of biological parts activity, satisfying the requisites of repeatability, reliability and shareability.
Like the assembly of an electronic circuit requires that all the components welded to a board (i. e. resistors, transistors, etc.) share a common structure and position to be properly fixed, similarly physical standardization of biological parts requires that all parts share a common "interface" to make the operations of DNA "cut and paste" simple and reproducible. One of the physical standards widely accepted in the field of Synthetic Biology is the *BioBrick*™ standard, proposed by Dr. Thomas Knight in 2003 [11]. It is based on the use of specific DNA sequences upstream and downstream of the part of interest, named respectively *prefix* and *suffix*. These sequences allow the assembly of two different *BioBrick*™ modules to form a new part, composed by the sequence of the starting modules and still compliant to the *BioBrick*™ standard, as described in Sec. 1.3.1.

BioBrick™ standard enables the physical assembly of individual parts in multi-component systems. The characterization of these parts requires an additional step: the standardization of tools, methodologies and measurement units that consent to study and re-use BioBrick parts reliably. An important result in this field is the definition of Relative Promoter Units (RPUs), standard units based on Polymerase per Second (PoPs) units, a physical quantity

[5] A restriction site is a specific, generally palindromic, 4-, 6- or 8-base pair long DNA sequence, recognized by an enzyme called *restriction endonuclease* as a DNA molecule cleavage site

that describes the activity of a transcriptional promoter [10]. Standard measurement techniques will be presented in Sec. 1.3.2.
A close attention must be paid to the direct application of engineering methodologies to the living material [12, 1]. In fact Synthetic Biology doesn't involve biological modules independent of the environment, but exploits the transcriptional and translational machinery of the host cell to implement the engineered task, also modifying essential processes. For these reasons, one of the main challenges of Synthetic Biology deals with the ability of fully predict the behaviour of even simple genetic circuits and build increasingly complex systems [2].

1.3.1 BioBrick™ parts, *Standard Assembly* and Registry of Standard Biological Parts

A biological part is a nucleic acid sequence that encodes a definable biological function, and a standard biological part is a part that has been refined in order to conform to one or more defined technical standards. After some attempts to standardize the components or processes underlying genetic engineering [13, 14] which did not provide a widely accepted solution, in 2003, Thomas Knight proposed the BioBrick™ standard for physical composition of biological parts [11]. Parts that conform to the BioBrick™ assembly standard are BioBrick™ standard biological parts. Unlike the previous experiences, BioBrick™ physical composition standard has been used by multiple groups [15, 10, 16, 17] and today has imposed itself as the gold standard. The main reason of this spread is that each summer since 2004 thousands of students develop and use BioBrick™ standard biological parts to engineer biological systems of their own design as a part of the International Genetically Engineered Machine competitions [18]. Additional technical standards defining BioBrick™ parts are set via an open standards setting process led by The BioBricks™ Foundation [19].
The key innovation of the BioBrick™ assembly standard is that any two BioBrick™ parts can be assembled in sequence, and the resulting composite object is itself a BioBrick™ part that can be combined with any other BioBrick™ parts. The idempotent physical composition standard underlying BioBrick™ parts enables the distributed creation of a collection of compatible biological parts and the shareability of parts among researchers working in laboratories in different parts of the world [16].

The Registry of Standard Biological Parts [20] represents the proof that such concepts of physical standardization, dissemination and re-usability of parts have been accepted and adopted worldwide. It is both an online open-access data base of biological parts and a physical repository (stored in the MIT laboratories) of such parts (in form of dry DNA, after proper quality-testing procedures), that can be requested and are free distributed to the research

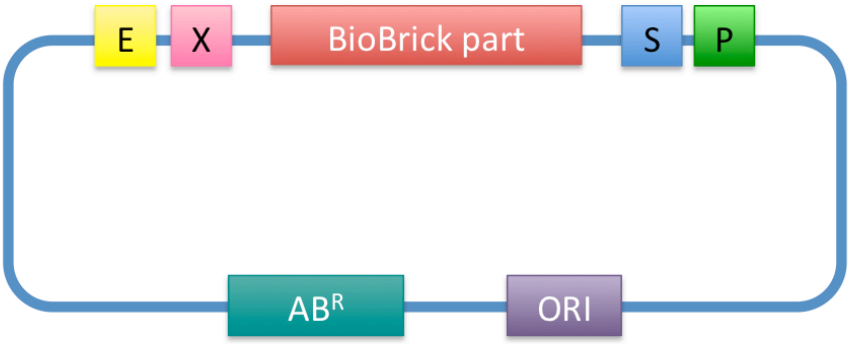


Figure 1.3: **BioBrick™ standard compliant part** The BioBrick™ part is preceded by a prefix sequence, composed by the EcoRI (E) and XbaI (X) restriction sites and is followed by a suffix, with SpeI (S) and PstI (P) restriction sites.

groups worldwide.

The Registry currently maintains a collection of over 6000 BioBrick standard biological parts [21]. Every part in the Registry has a BioBrick™ code number, that serves as the unique identifier of the part. The Registry maintains information about each part including its sequence, function, and, if available, user experiences and quality control. DNA encoding each BioBrick™ standard biological part is stored and can be propagated in *Escherichia coli* plasmid-based vectors [22].

BioBrick™ standard is based on the use of specific DNA sequences upstream and downstream of the cloning site, containing the part of interest. These sequences, named respectively *prefix* and *suffix*, contain peculiar restriction sites.

As illustrated in Fig. 1.3, the prefix is composed by *EcoRI* (E) and *XbaI* (X) restriction sites, while the suffix is composed by *SpeI* (S) and *PstI* (P) restriction sites. The region from the prefix to the suffix is named *cloning site* and hosts the desired part. The DNA cut operated by these four restriction endonucleases produces protruding or *sticky* ends, that can be easily joined back together or with other protruding ends generated by the same restriction endonuclease, thus allowing the easy cut and paste of DNA portions. The activity of ligase, a specific enzyme, is required to complete the joining process. The cleavage modes for E, X, S and P and the produced sticky ends are reported in Fig. 1.4.

The BioBrick™ assembly is based on the presence of two *isocaudamer* restriction endonucleases, X and S. Their cleavage generates identical sticky ends, that can be connected together. The new sequence does not match any of the original restriction sites and cannot be re-digested by any of the four enzymes

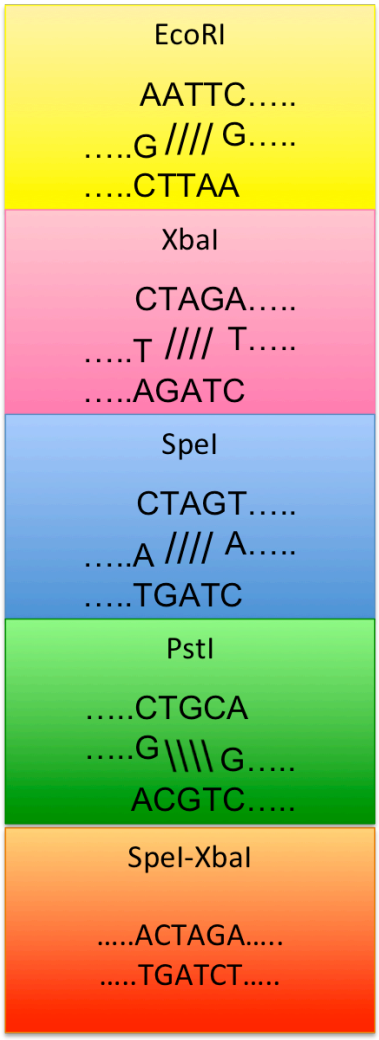


Figure 1.4: **E, X, S and P cleavage sites and mixed scar formed from X-S combination** The cleavage modes are indicated for the four restriction enzymes EcoRI (yellow box), XbaI (pink box), SpeI (blue box) and PstI (green box). In the orange box, the mixed site formed by the combination of XbaI and SpeI sticky ends is shown. This mixed site cannot be further recognized by any of the four original enzymes.

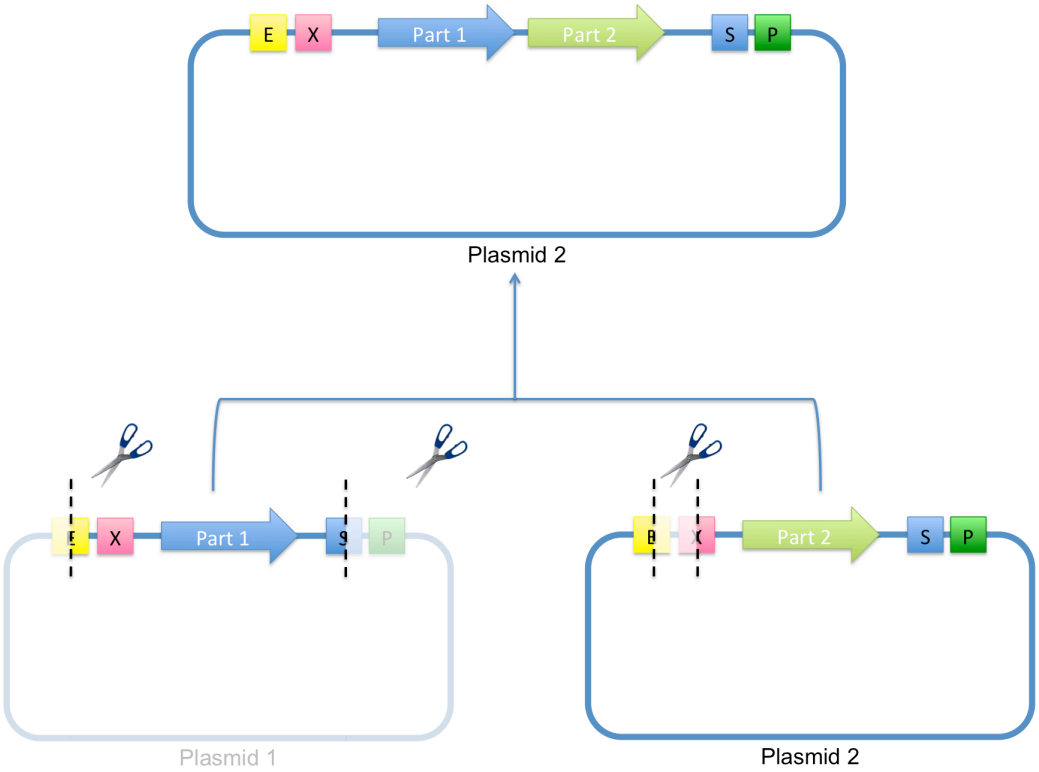


Figure 1.5: **BioBrick™ Standard Assembly** Starting from two parts (green and blue), borne respectively on Plasmid 1 and Plasmid 2, in order to obtain a composite part composed by the blue sequence and the green sequence, borne by the Plasmid 2, the BioBrick™ Standard Assembly procedure requires that Plasmid 1 is digested with EcoRI and SpeI restriction enzymes and Plasmid 2 with the EcoRI and the XbaI restriction enzymes. The portions indicated with a non-transparent color, once purified, can be ligated to obtain the desired final circuit.

involved in the BioBrick™ assembly. The X-S mixed site is often called 'scar sequence'. The standard method to assemble two BioBrick™ parts is reported in the Fig. 1.5.

Briefly, in order to obtain a composite BioBrick™ part composed by the sequence of the blue BioBrick™ and the green BioBrick™, mantaining the plasmid backbone of the green part, the method requires:

- the digestion of the blue BioBrick™ with *EcoRI* and *SpeI* enymes,
- the digestion of the green BioBrick™ with *EcoRI* and *XbaI* enymes,
- the purification of the digested fragments,
- the ligation of the purified fragments, that will generate the desired composite part.

The Standard Assembly procedure of BioBricks™ has contributed to their diffusion among the Synthetic Biology community. However, although the majority of available parts in the Registry conform to this standard and the original Standard Assembly is currently successfully used in many laboratories worldwide, other methodologies and standards have been proposed to overcome some of the limitations that affect the original Standard Assembly. To avoid the need of gel-purification of digested DNA fragments, the Three Antibiotic (3A) Assembly has been proposed to assemble two BioBrick™-compatible parts and select for correct assembles through antibiotics. Because gel-extraction can be avoided, this method is suitable for high-throughput assembly protocols and for automatic construction processes via liquid handling robots. To support the assembly of protein domains, a promising new standard has been proposed by J.C. Anderson [23]. The BglBrick standard supports the assembly of compatible parts via the EcoRI/BglII and BamHI/XhoI restriction sites in prefix and suffix of basic parts, respectively. BglII and BamHI are isocaudamer sites and their mixed site is a 6-nucleotide scar sequence (GGATCT) encoding the glycine-serine aminoacids, demonstrated to be innocuous in most protein fusion applications in a variety of host systems. On the other hand, the original BioBrick™ structure results in a scar containing a stop codon in frame. The described standards and methodologies support the assembly of two parts in each reaction. It is worth mentioning that a procedure called Gibson Assembly has been invented to support the assembly of multiple overlapping DNA molecules by the concerted action of a 5' exonuclease, a DNA polymerase and a DNA ligase in an isothermal, single-reaction. The overlapping regions can be added to the ends of any length of DNA via common molecular biology techniques and no scar is present in the resulting assembled sequence. However, this method requires the synthesis of customized oligonucleotides, specific for the parts to be assembled. For this reason, this is an attractive methodology to construct large genetic systems, but it cannot be defined as a standard procedure, as it requires a customized work for each assembly.

1.3.2 Standardizing measurements in Synthetic Biology

The usefulness of standard biological parts would increase if the behaviour of such parts, both individually and combined in complex systems, were more predictable and if the characterization of their functioning were inexpensive, efficient, predictable and reliable [10]. As previously described, physical standards try to define common procedures to physically link biological modules. Measurement standards, in turn, try to define common units that can express the quantitative behaviour of these modules. A set of quantitatively characterized parts will enable a designer to predict how they will function as a group. For every part, the measured characteristic should be annotated in a compact, prescribed document that collects and summarizes information as a formal set of context-dependent, input-output behaviours, tolerances, requirements, physical interconnected "form factors" (the mechanical requirement for

physical incorporation of the device into a system) and other details about a particular part or subsystem. This compact form allows the simple and error-proof communication of information and enables engineers to rapidly select from a vast list the parts that will meet their design requirements. Adherence to the set of standards ensures that each device and the systems made from them will work as intended [24].
Some features typically reported on datasheets are:

- a definition of the function and interfaces of the device (the accepted inputs and the produced outputs);
- the operating context and the working conditions of the device;
- the measured variables that describe the quantitative behaviour of the device;
- the transfer function, which details the static input/output relationship and allows prediction of the equilibrium behaviour of composed devices.

Biological devices, such as electronic devices, can be grouped in different categories by their logic function: biological sensors [25], actuators [26], logic gates [27], circuits mimicking electronic devices [28] have all been realized and documented. Despite differences in regulation mechanism, nature of the inputs and of the outputs, systems belonging to the same category can be described by the same features (e.g. all sensors are functional devices that convert a specific input signal into a general transcriptional signal, which can drive the expression of a given observable output or can be the input of another network interconnected in series). One of the most popular examples of datasheets in Synthetic Biology is the BBa_F2620 datasheet, described in [9] and reported in Fig. 1.6.

Characterization of the activity of transcriptional promoters

Standard methodologies for the quantitative characterization of biological parts and devices have been proposed in literature. In particular, many efforts have been addressed to the quantitative characterization of promoter activity, since its main role in driving the expression of recombinant circuits.
Polymerase per Second (PoPs) units, defined as the number of times that an RNA polymerase molecule passes a specific point on DNA per time unit [10], are accepted by the scientific community as promoter activity measurement units. Unfortunately, although PoPs can be an interesting common signal for a wide range of transcription-based synthetic biological devices, its *in vivo* measurement is challenging, and it has not yet been reported in literature. In case of promoters, herein considered as PoPs generators, the transcription rate can be a good approximation of PoPs. Transcription rate of promoters could

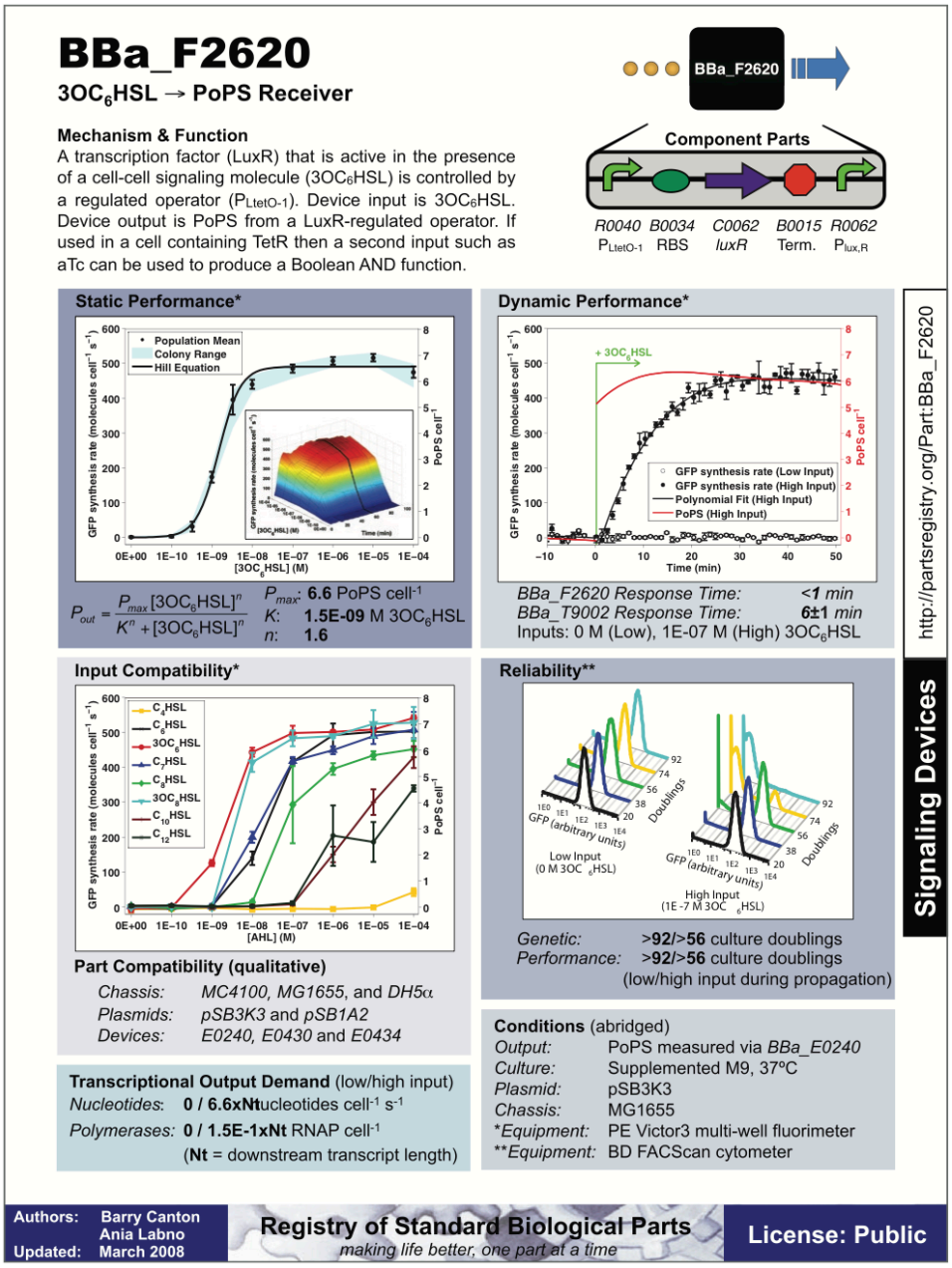


Figure 1.6: **BBa_F2620 datasheet** BBa_F2620 is a regulated synthetic promoter built by exploiting the *quorum sensing* regulation of *Vibrio fischeri* bacteria. This device senses the concentration of a specific molecule (3-oxo-hexanoyl homoserine lactone - $3OC_6$-HSL), which binds the transcription factor LuxR and activates the P_{Lux} promoter. The output intensity depends on the concentration of $3OC_6$-HSL in the environment.

be measured, but such procedure relies on expensive and invasive methodologies (e.g. cell disruption and Real-Time PCR).
Canton et al. [9] defined a model-based procedure for indirect PoPs estimation in a system composed by the BBa_F2620 inducible device with a Green Fluorescent Protein (GFP) reporter device downstream (i.e. the composite part RBS-GFP-Terminator). This method is based on a mathematical model approach. The Ordinary Differential Equations (ODE) system of the synthesis of a fluorescent protein (GFP) is represented by the following equations:

$$\frac{d[M]}{dt} = n \cdot PoPs - \gamma_M[M] \tag{1.1}$$

$$\frac{d[I]}{dt} = \rho[M] - (\alpha + \gamma_I)[I] \tag{1.2}$$

$$\frac{d[G]}{dt} = \alpha[I] - \gamma_G[G] \tag{1.3}$$

where $[M]$ is the mRNA concentration per cell, $[I]$ the immature GFP concentration per cell and $[G]$ the mature GFP concentration per cell. γ_M is the mRNA degradation rate, γ_I and γ_G are, respectively, the immature and mature GFP degradation rates. ρ is the immature GFP synthesis rate and α is the GFP maturation rate, while n is the copy number of the circuit used to quantify the activity of the promoter. The steady-state evaluation[6] of this model permits the determination of the absolute activity of the promoter ($PoPS^{SS}$), that can be explicited as:

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

where S^{SS}_{cell} is the observed GFP synthesis rate per cell at the steady state ($S_{cell} = \alpha \cdot [I]$).
S_{cell} can be experimentally measured as $\frac{d[GFP]}{dt} \cdot \frac{1}{CFU}$ at the steady state, where CFU are the colony forming units. To achieve this, fluorescence (proportional to the number of GFP in the culture) and absorbance (proportional to the number of bacteria in the culture) values, measured in arbitrary units by a machine, have to be converted into absolute units by "fluorescence vs GFP protein concentration" and "absorbance vs colony forming units" calibration lines and then S_{cell} can be numerically computed. Although the performed efforts resulted into a mathematical method for PoPs estimation, this procedure still relies on expensive experiments (e.g. Western Blot) for the identification of the other model parameters. In order to avoid such expensive and hard-to-standardize experiments and to make PoPS measurement independent of the measurement instruments and of the experimental conditions, a relative

[6]Note that the simplifying steady-state hypothesis should be verified when cell cultures are in exponential growth phase, when cell doubling happens at constant time intervals and the cell translational and transcriptional processes are supposed to be at the equilibrium.

approach has been adopted. Kelly et al. [10] proposed the Relative Promoter Units (RPUs) as a standard measurement unit for this biological species. The BioBrick™ constitutive promoter BBa_J23101 has been adopted as a reference standard. The relative activity of a promoter ϕ (RPU_ϕ) can thus be expressed as:

$$RPU_\phi = \frac{PoPS_\phi^{SS}}{PoPS_{J23101}^{SS}} \tag{1.5}$$

By substituting Eq. 1.4 in Eq. 1.5 and introducing the simplifications, under the hypothesis that the experimental conditions and measurement system are identical for ϕ and $J23101$, that $\gamma_{M,\phi} = \gamma_{M,J23101}$, $\gamma_{I,\phi} = \gamma_{I,J23101}$, $\alpha_\phi = \alpha_{J23101}$, $\rho_\phi = \rho_{J23101}$ and $n_\phi = n_{J23101}$, results that:

$$RPU_\phi = \frac{S_{cell,\phi}^{SS}}{S_{cell,J23101}^{SS}} \tag{1.6}$$

under the hypothesis that the growth rates of cultures bearing the circuit with ϕ promoter and $J23101$ promoter are similar.

Using this approach, it is possible to characterize a promoter following a simple, cheap and fast protocol. As reported in Fig. 1.7, it is sufficient to build two genetic circuits, one for the promoter to be characterized and one for the BBa_J23101 reference promoter. The circuits must be identical, except for the different promoter (i.e.: the same RBS, the same fluorescent reporter, the same transcription terminator) and must be assembled in the same plasmid (i.e.: same antibiotic resistance, same origin of replication, same sequence upstream and downstream of the cloning site) or integrated in the same genome *locus* of the same bacterial strain. The cultures bearing the two genetic circuits must be identically assayed in order to measure the culture growth level (i.e.: monitoring the absorbance of a liquid culture) and the produced fluorescence (i. e.: using a fluorimeter to detect the produced fluorescence). As anticipated above, the time series of culture growth (CG) and produced fluorescence (FL) can be processed in order to obtain an estimation of S_{cell}^{SS} according to this formula:

$$S_{cell}^{SS} = \frac{d[FL]}{dt} \cdot \frac{1}{CG} \tag{1.7}$$

evaluated for CG in exponential growth phase. Applying the Eq. 1.6, RPUs can be computed. It is important to note that fluorescence and absorbance conversions from arbitrary units into absolute units is no more required, as the S_{cell} of $J23101$ and ϕ are both expressed in the same arbitrary units.
Relative activity measurements based on an *in vivo* reference standard enable the measurement of promoter activities in shareable units, which do not depend on the experimental condition and provide comparable results among different researchers across many laboratories. RPUs have been used in many scientific works to quantify promoter activity [29, 30, 31, 32] and have provided accurate, robust and comparable results.

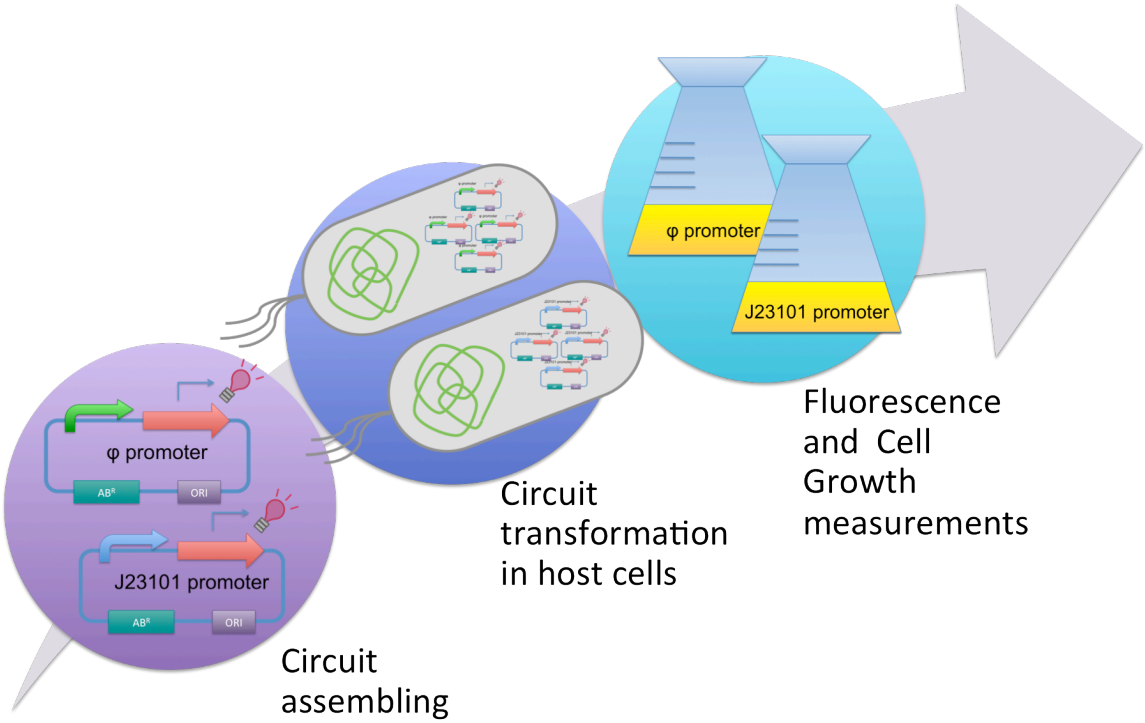


Figure 1.7: **RPU evaluation workflow** The promoter of interest (ϕ) and the reference standard promoter (J23101) are assembled in the same circuit (violet circle) and incorporated in the host cells (blue circle). Cells bearing the two devices are assayed to measure the cell growth and the produced fluorescence (light blue circle). These measurements are then processed in order to obtain the RPU value.

1.4 Mathematical modeling in Synthetic Biology

One of the major challenges of Synthetic Biology concerns the bottom-up engineering of gene networks with desired behaviour by using quantitative models of molecular interactions to predict such behaviour. This prediction problem is extremely complex because of its dimension and complexity, in addition to the scarcity of available models [33] and to the intrinsic noise affecting gene expression [34]. The ability to relate the mathematical models with experimental data is essential to drive the rational design and fine tuning of gene-networks, avoiding expensive, time consuming trial-and-error approaches. Biological engineerings and synthetic biologists have tried to figure out for years what types of models are more appropriate in this field. Gene networks can be studied at the molecular level by implementing dynamic models of all the molecular interactions involved in transcription, translation, regulation and induction [35] or be reduced to a simple and elegant explanation of the observed dynamic behaviour [28], which does not provide useful information on DNA sequences interactions [36].
In Synthetic Biology, model-driven rational engineering of synthetic gene networks can be performed at two different levels [36]:

- the *network topology* definition. By combining simple regulatory interactions,the most various and complex circuits can be realized. Excellent examples of this approach are Gardner's toggle switch [28] and Elowitz's repressilator [37];

- the *molecular* level, which describes the kinetics and strengths of molecular interactions within the system. An example is the study on the lactose operon and of the effects of sequence variations on phenotype performed by Mayo et al. [38].

In this context, the attention will be focused on the first modeling level. Modeling metabolic networks, signaling pathways or gene regulatory networks requires the introduction of many simplifying hypotheses, such as the uniformity of species distribution in cells (to eliminate diffusion rates for concentration gradients), and the equilibrium or steady state for some chemical reactions (i.e.: enzyme kinetic or transcriptional regulation) [39].
Modeling typically requires an accurate knowledge of the system parameters, such as kinetic and binding constants, that can be estimated from experimental data, obtained from *ad-hoc* experiments.

1.4.1 Deterministic and stochastic models

In a deterministic model every species is uniquely determined by model parameters and by the previous states of model variables. Therefore, deterministic models are not affected by random fluctuations and the predictions

from identical initial conditions, with identical parameters, are identical. The most popular deterministic models are the ones based on differential equations. Several algorithms to solve ODE systems with nonlinear dynamics are available, exploiting numerical integration algorithms. In many cases, the systems can be studied at the steady-state, thus introducing simplifying hypothesis in problem solving. The traditional methodologies of ODE analysis can be applied to these systems, such as bifurcation study [39, 40, 41, 42].

In antithesis with this paradigm, stochastic models consider randomness affecting the process and species are not described by deterministic values, but rather by probability distributions. Taking into account the fluctuations that exist in living systems the investigation of the effects of biological noise on the system behaviour is enabled [39]. A commonly used stochastic method is SSA (Stochastic System Algorithm), described by Gillespie in 1976. The SSA describes the system as a collection of randomly moving molecular species with mutual interactions and accurately describes the dynamics of a well-mixed system affected by wide fluctuations due to the small number of molecules. Given the current state of the system, the SSA evaluates the time necessary to the next reaction to occur, executes that reaction, updates the state of the system and increments the simulation time to the new value. This system has proved to be inefficient and computationally onerous in stiff problem solving, when highly different dynamics are present in the same process, since it solves every reaction individually, spending inordinate amounts on fast reactions for very few simulated occurrences of slow reactions [35, 43].

Chapter 2

Preliminary study on nonlinearities and unpredictability of cell behaviour[1]

The central topic of this thesis is the study of nonlinear effects on gene expression introduced by copy number variation. In order to highlight how relevant this aspect is in engineering new gene networks, two case studies are here presented.
First, a study concerning the assessment of modularity in Synthetic Biology will be presented (Sec. 2.1). Here, different transcriptional promoters have been characterized using three different measurement systems and in different copy number conditions, in order to validate the experimental context-dependent variability of their activity. This study has been published in [32].
The second study, presented in Sec. 2.2 discusses the characterization of a synthetic bacterial self-destruction device. Here, a synthetic device expressing a holin and a lysozyme, two proteins involved in bacterial membrane disruption, has been assembled under the control of a regulated promoter and has been fully characterized in terms of lysis entity, delay time and genetic stability. This study has been published in [44].
In both these works, the studied devices have been assembled in plasmids of different copy numbers. Despite the applications are totally dissimilar, in both cases the fine tuning of plasmid copy number has been crucial in order to assess the desired behaviour. Both studies have highlighted issues in the use of high copy number plasmids, which have proven to be unstable and with an unpre-

[1] The content of this chapter has been partially reported in *L. Pasotti, N. Politi, S. Zucca, M. G. Cusella De Angelis, and P. Magni.* PLoS ONE*, 2012.* and in *L. Pasotti, S. Zucca, M. Lupotto, M. G. Cusella De Angelis, and P. Magni.* Journal of Biological Engineering*, 2011.*

dictable output, probably due to the high metabolic burden for the host cell and to the non-linear saturation phenomena occurring in transcription/translation (probably due to the overload of transcriptional/translational cell machinery). At the same time, the stability and the functioning predictability in the low copy number plasmids have been validated.
In this chapter, the main results are presented and the essential information to introduce the issues imputable to copy number variation are provided. The materials and methods sections are reported in App. A. For a detailed and comprehensive discussion, refer to the cited papers.

2.1 Modularity study on basic parts and devices

2.1.1 Promoters characterized via different measurement systems

The activity of a set of promoters was quantified in *Escherichia coli* via different measurement systems (i.e., different plasmids, reporter genes, ribosome binding sites) relative to an *in vivo* reference promoter, as explained in Sec. 1.3.2).

The schematic functional representation of the tested framework is shown in Fig. 2.1 A and B. The activity of the promoters listed in Fig. 2.1 C is computed relative to a standard reference promoter with the RPU approach. Three reporter devices (GFP32, RFP34 and RFP32) and two copy number conditions (LC or HC) are used as different biological measurement systems, as shown in Fig. 2.1 D.

Results are shown in Fig. 2.2 for the low copy condition. Promoters span a >10-fold RPU range, in which P_{LlacO1} is the strongest one, while P_{lacIQ} is the weakest one. Given a measurement system, the quantified activity of each promoter is reasonably reproducible among the technical replicates, giving an average coefficient of variation (CV) of 9%. As expected, promoters characterized via RFP34 give a higher absolute activity than with RFP32 (data not shown), as the BBa_B0032 RBS is weaker than BBa_B0034. Only P_R activity is statistically different among the three tested measurement systems ($P \leq 0.05$, ANOVA), yielding a CV of 22% among the three measured mean activities. The observed variability may be caused by downstream sequence-dependent promoter activity change. The maximum activity variability found in this set of promoters is relatively low and it is smaller than previously reported in other downstream sequence-dependent case studies [29, 30].

Promoters with GFP32 and RFP34 reporter devices were also tested in a high copy number plasmid.

By comparing the activity of promoters characterized via the same reporter device in low copy and high copy vectors (Fig. 2.3), it resulted that the RPUs

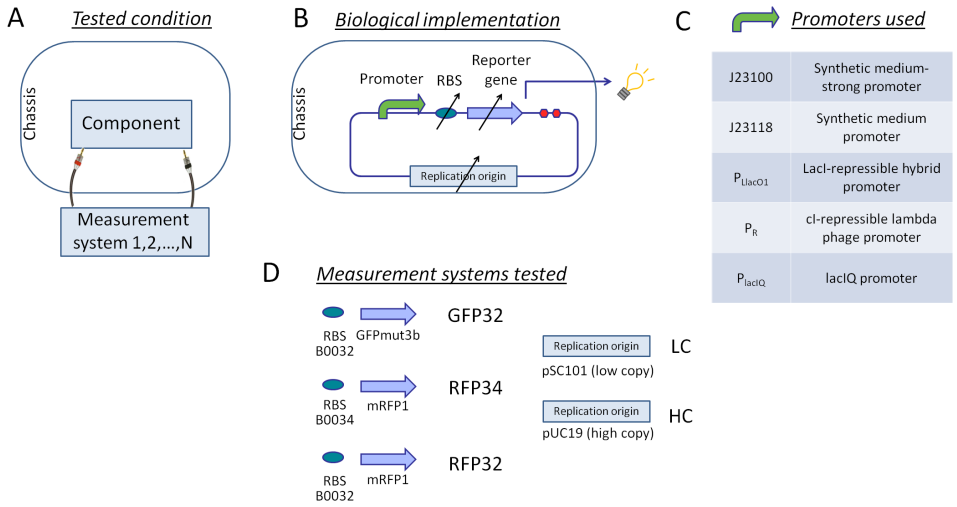


Figure 2.1: **Study of individual promoters activity quantified via different measurement systems** The same component is tested with different measurement systems (A). The promoter of interest is assembled in different contexts and drives the expression of a reporter protein (B). The list of the tested promoters is reported in panel C, while the different measurement systems are reported in panel D.

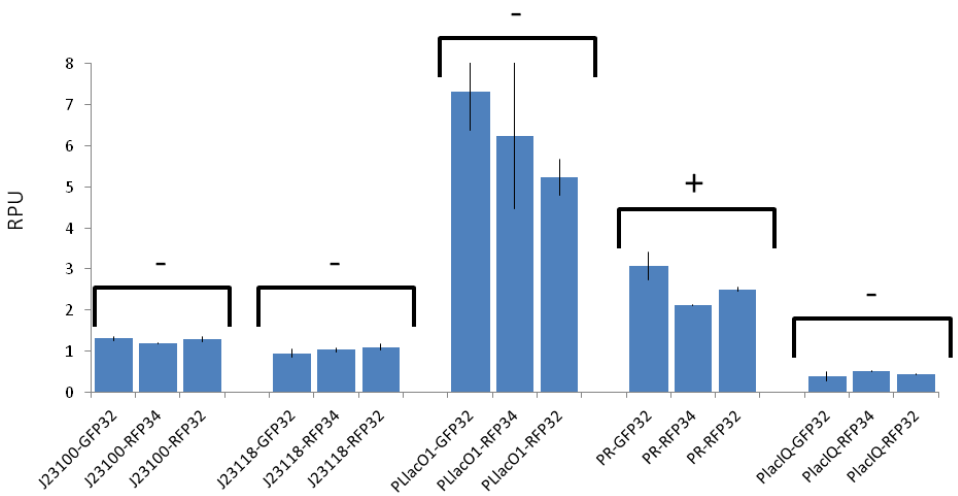


Figure 2.2: **Measured RPU values for the five investigated promoters** Promoters were individually characterized via three reporter devices: GFP32, RFP34 and RFP32. Error bars represent the standard deviation of the mean activity computed on three clones. For each promoter, statistical analysis was performed via ANOVA test to compare the RPU activities measured via the three different reporter devices. Promoters showing a statistical difference ($P<0.05$) in the mean activities among the three conditions are marked with a '+' sign, while promoters not showing any significant difference ($P\geq0.05$) are marked with a '-' sign. Strains with P_{LlacO1} were induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG).

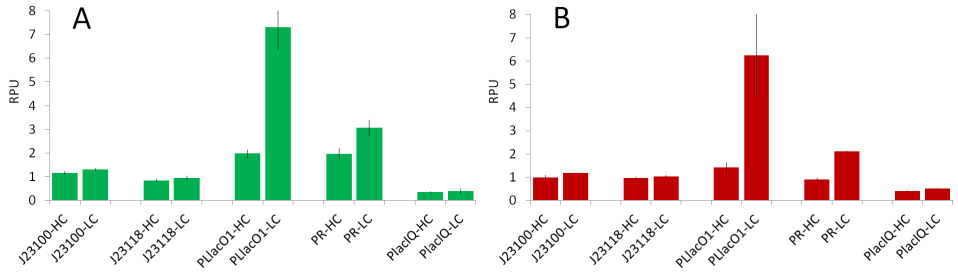


Figure 2.3: **Measured RPU values for individual promoters characterized in a high (HC) or low (LC) copy vector with GFP32 and RFP34 measurement systems** Promoters were characterized via GFP32 (panel A) and RFP34 (panel B) reporter devices. Error bars represent the standard deviation of the mean activity computed on three clones. Strains with P_{LlacO1} were induced with 1 mM of IPTG.

of the two strongest promoters, P_{LlacO1} and P_R, were respectively 4.4- and 2.3-fold lower in high copy when compared to low copy condition. This means that, given a reporter device, the ratio between the activity of the promoter of interest and the reference is lower in high copy when compared to low copy. The other promoters did not show such a large difference ($<$ 1.3-fold). It is unlikely that such huge differences are due to upstream sequence-dependent promoter activity change (caused by the different sequence of pSB1A2 high copy vector when compared to pSB4C5 low copy vector). The observed large-entity variations could be due to saturation effects in transcription/translation processes that occur for the strongest promoters in high copy condition, while such effects were absent for the other (weaker) promoters.

2.2 Characterization of a bacterial self-destruction device

2.2.1 Background

Bacterial cell lysis is a widely studied mechanism that can be achieved through the intracellular expression of phage native lytic proteins. This mechanism can be exploited for programmed cell death and for gentle cell disruption to release recombinant proteins when *in vivo* secretion is not feasible. Several genetic devices for cell lysis have been developed and their quantitative characterization is an essential step to enable the engineering of synthetic lytic systems with predictable behaviour. The BBa_K112808 BioBrick™ lysis device present in the Registry of Standard Biological Parts has been quantitatively characterized. Its activity has been measured in *E. coli* by assembling the device under the control of the well characterized BBa_F2620 HSL-inducible device and the transfer function, lysis dynamics, protein release capability and

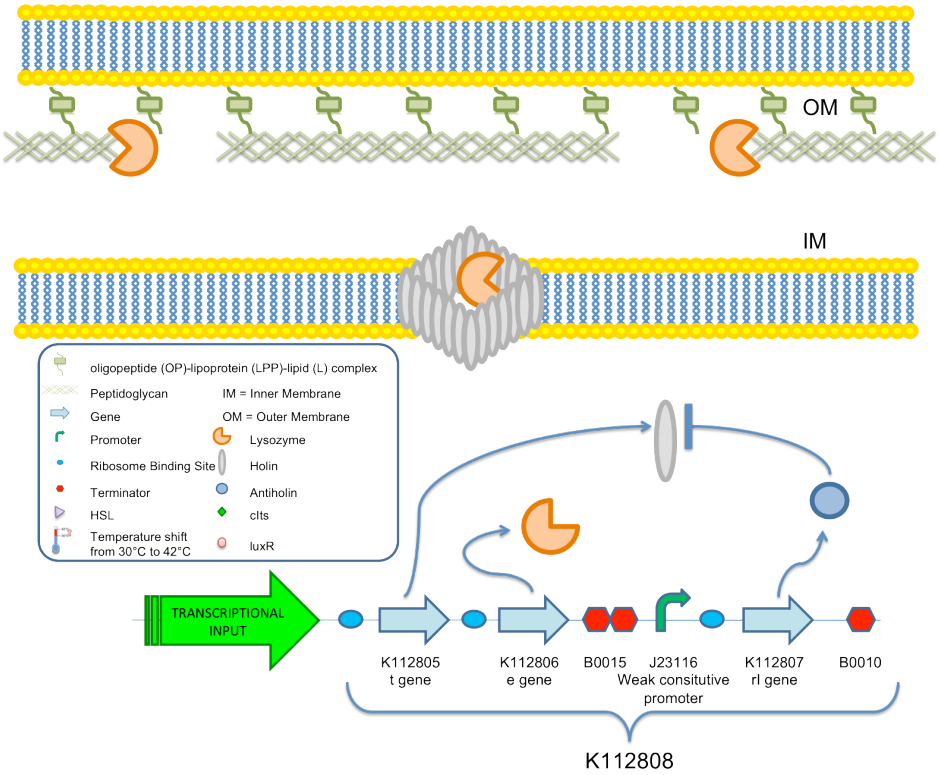


Figure 2.4: **Promoterless lysis device gene network and its functioning scheme in disrupting bacterial membrane.** The gram-negative bacteria envelope consists in an inner membrane, an outer membrane and a peptidoglycan layer, which is linked to the outer membrane via oligopeptide (OP)-lipoprotein (LPP)-lipid (L) complexes [45]. The lysis device consists of a promoterless operon encoding a holin (gene *t*) and a lysozyme (gene *e*). Holin produces lesions in the inner membrane, through which lysozyme can pass, reaching and attacking the peptidoglycan layer, thus causing cell lysis. The weak expression of a antiholin (gene *rI*) driven by promoter BBa_J23116 prevents the cell lysis caused by basal expression of the *t*-*e* operon.

genotypic and phenotypic stability of the device have been evaluated.
In order to understand the biological mechanism underlying the functioning of cell lysis device, the bacterial membrane structure will be briefly described. The gram-negative bacteria envelope consists in an inner membrane, an outer membrane and a peptidoglycan layer, which is linked to the outer membrane via oligopeptide (OP)-lipoprotein (LPP)-lipid (L) complexes and gives structural strength to the cell wall, as shown in Fig. 2.4. The lysis device is composed by a promoter-less operon encoding a holin (gene *t*) and a lysozyme (gene *e*). When the operon is expressed, holin forms lesions in the inner membrane. Lysozyme can pass through these lesions, reaching and attacking the peptidoglycan layer, thus achieving cell lysis. In addition, the weak promoter BBa_J23116 constitutively expresses gene *rI*, encoding an antiholin, which inhibits the holin action caused by basal expression of the *t*-*e* operon (Fig. 2.4).

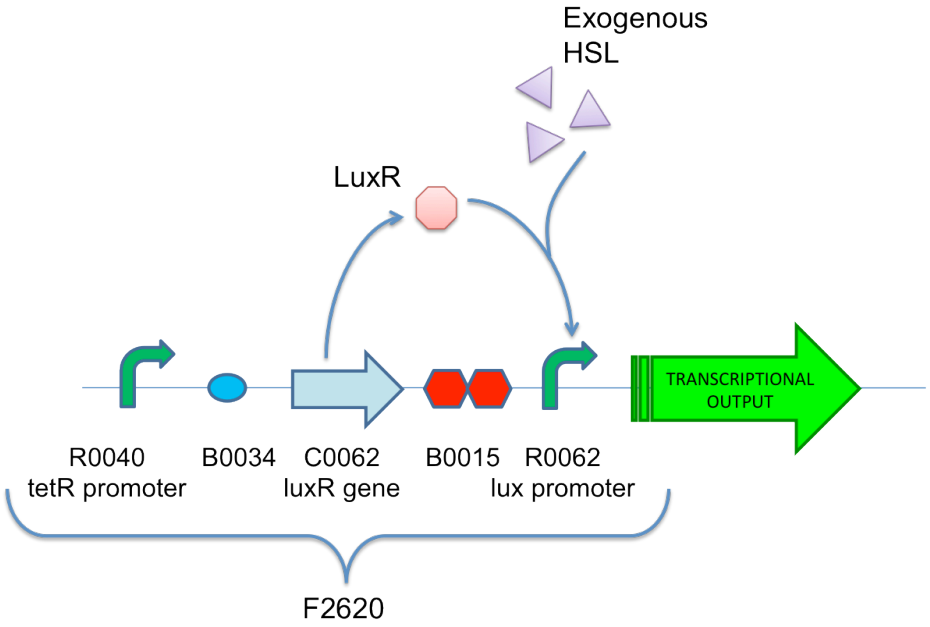


Figure 2.5: **HSL-inducible promoter.** BBa_F2620 HSL-inducible input device consists of a luxR expression cassette driven by the tetR promoter and the lux promoter. When LuxR transcription factor is expressed and HSL is added to the medium, lux promoter expression is triggered.

2.2.2 Design and characterization of regulated lysis devices

This device has been assembled and tested under the regulation of an inducible promoter. BBa_F2620 HSL-inducible input device is composed by a LuxR expression cassette, driven by the tetR promoter, and the lux promoter, which is normally off. Its transcription can be induced by the LuxR transcription factor in presence of exogenously added HSL (Fig. 2.5).

Lysis entity was quantitatively assayed, as described in App. A.2.1, by measuring the turbidity of liquid cultures over time through OD_{600} readings. A typical lysis profile is reported in Fig. 2.6, where *E. coli* TOP10 bearing BBa_K112808 lysis device driven by BBa_F2620 device in a low copy plasmid were induced with HSL 100 nM at t = 0 h (OD_{600} = 0.2, exponential phase), t = 4 h (OD_{600}∼1.3, early stationary phase) or t = 20 h (OD_{600}∼2, late stationary phase). Lysis began after about 15 min from the induction in all the growth phases and its mean entity was about 76.3 ± 0.3%, 75.4 ± 1.1% and 50 ± 2.5 % at t = 0h,t = 4 h and t = 20h, respectively. The mean doubling time of uninduced TOP10 bearing lysis device in low copy was 49.8 ± 1.1 min, while the doubling time of the negative control was 43 ± 1.3 min, thus demonstrating that the HSL-inducible lysis device gives a reasonably low metabolic burden and allows the cells to grow at a rate comparable to their negative control.

Analogous studies were conducted for the HSL-inducible lysis device in high copy plasmid. On 10 experiments carried out in separate days, in half of them bacteria did not lyse upon induction (data not shown). Lysis occurred in exponential and early stationary phases, but never in late stationary phase. Lysis behaviour in high copy plasmid had a high variability between different experiments carried out in separate days, in fact lysis entity in exponential phase and early stationary phase was $58.6 \pm 7.5\%$ and $42.2 \pm 9.9\%$ respectively. *E. coli* TOP10 bearing the HSL-inducible lysis device also showed a very high and variable doubling time (69 ± 6 min), probably due to the toxicity of the lysis proteins expressed by the leakage activity of lux promoter present in high copy number, while the negative control grew faster (doubling time of 45 ± 4.7 min). In cultures bearing the lysis device in high copy that failed to lyse, the doubling time was 47.4 ± 2.1 min, suggesting that mutations caused by selective pressure had occurred and the original culture had been replaced by a mutant culture with higher fitness. Finally, as resulted for the HSL-inducible lysis device in low copy plasmid induced with HSL 100 nM, lysis occurred after 15 min from the induction in all the growth phases and mutants arose after about 2-3 h from the lysis start (data not shown). All the described quantitative results are summarized Tab. 2.1.

Most of the obtained quantitative results are summarized in a specific datasheet for the BBa K112808 device, which is shown in Fig. 2.7. For a detailed description, refer to the paper [44].

2.3 Discussion and conclusions

The success of the bottom-up approach in designing new gene networks strictly depends on the clear definition of the limits in which biological functions can be predictable. Such limits have been investigated in the two case studies presented in this chapter.

It is common knowledge that the activity of genetic parts in high copy number vectors can be nonlinearly affected by the overloading of cell machinery due to the high copy number of the DNA-encoded functions [29]. This was confirmed by comparing the activity of promoters characterized via the same reporter device in low copy and high copy vectors. The results have highlighted the unsuitability of high copy vectors for the characterization of biological parts,

Figure 2.6 *(facing page)*: **Typical lysis profiles for BBa_K112808.** OD_{600} of TOP10 *E. coli* cells with the HSL-inducible lysis device, induced with HSL 100 nM (blue line) and uninduced (red line), negative control (C-) induced with HSL 100 nM (green line) and uninduced (black line). Induction was performed in exponential phase ($OD_{600} = 0.2$) at t = 0 (A), early stationary phase ($OD_{600} \sim 1.3$) at t = 4 h (B) and late stationary phase ($OD_{600} \sim 2$) at t = 20 h (C). Error bars represent the 95% confidence interval of the estimated mean. For clarity of presentation, data points shown here are resampled with a 1-hour sampling time.

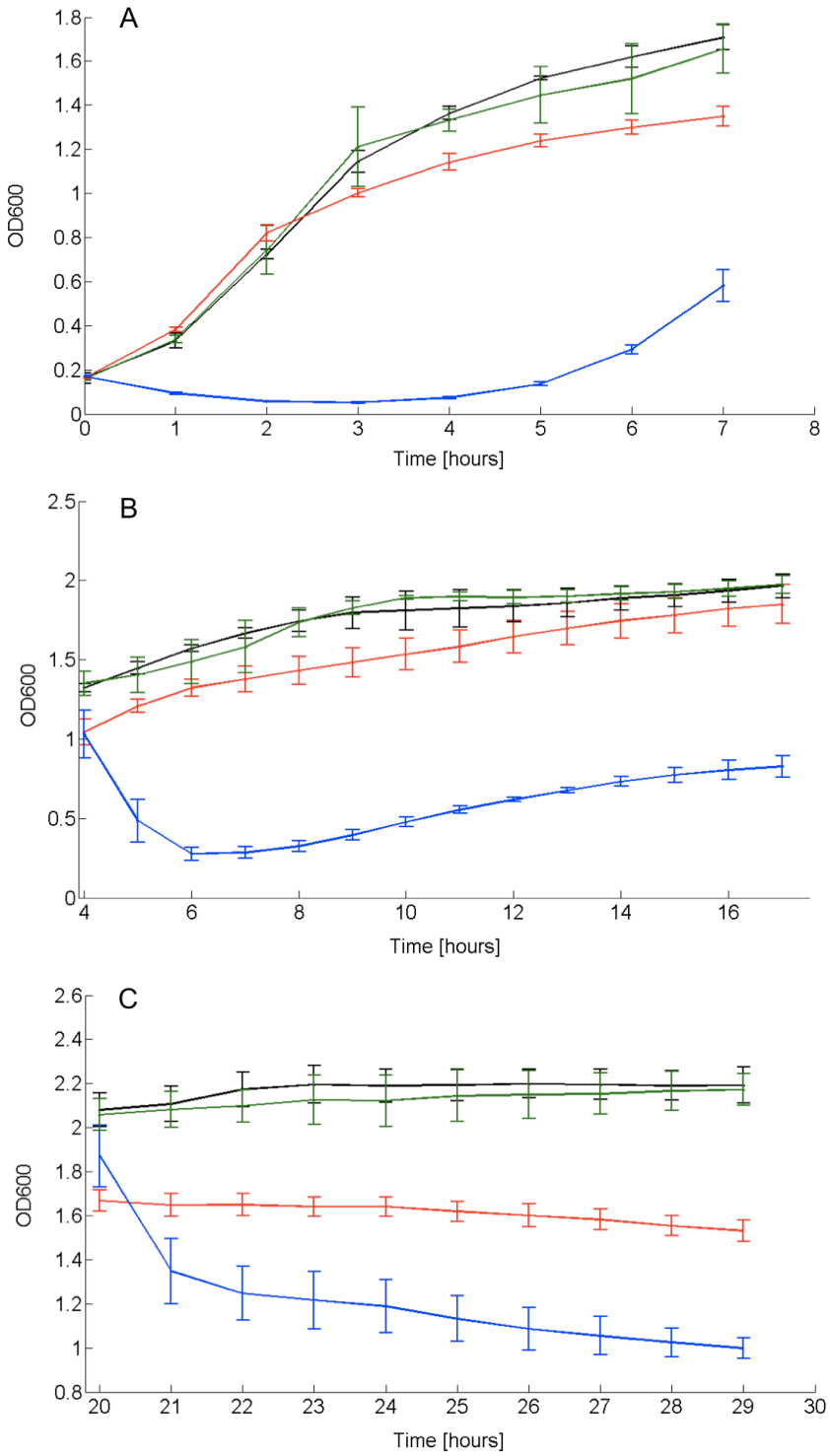
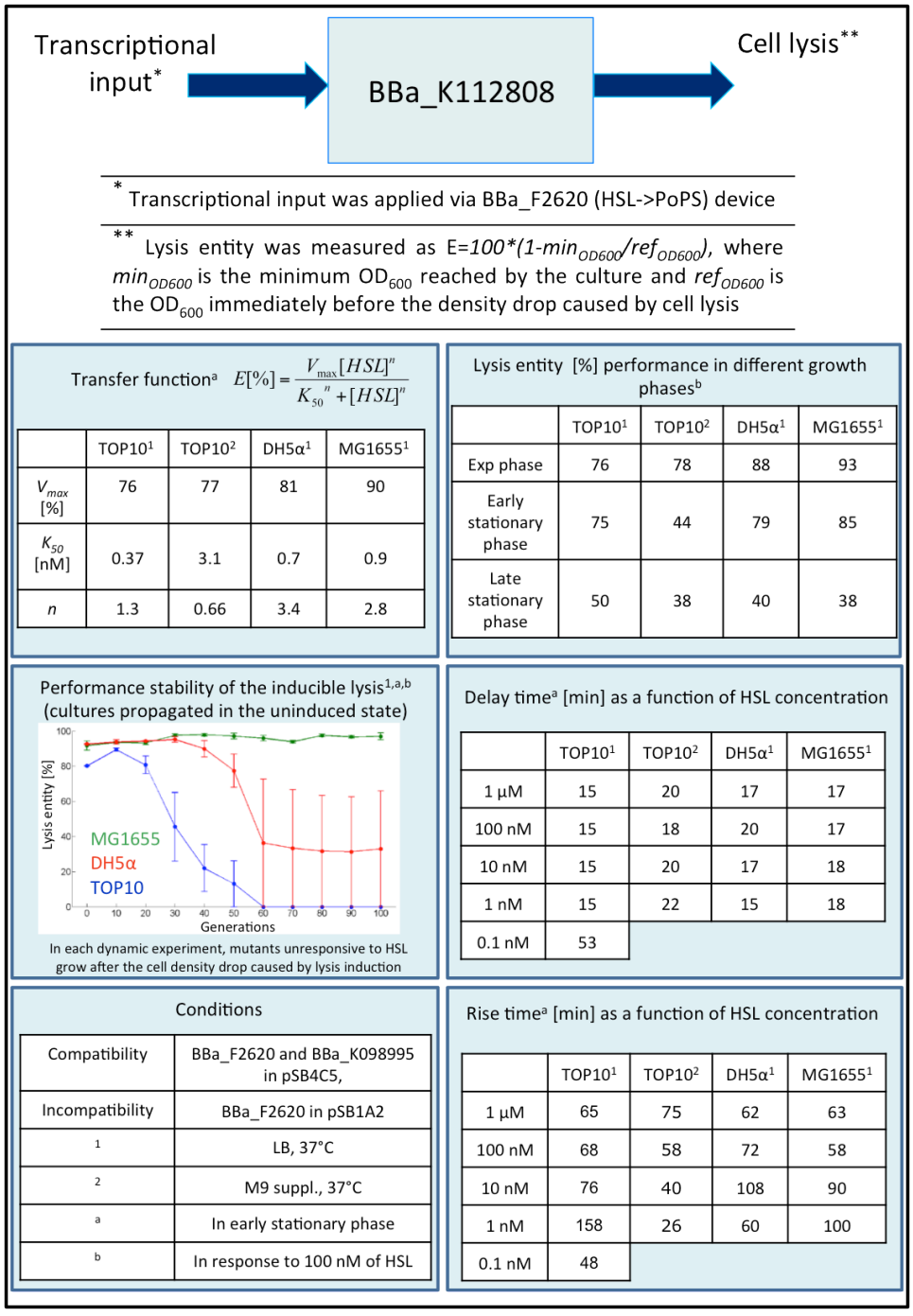

A
OD600
Time [hours]
B
OD600
Time [hours]
C
OD600
Time [hours]

Table 2.1: **Quantitative characterization of lysis parameters.** All the measurements were performed on TOP10 *E. coli* cells bearing the HSL-inducible lysis device in low copy plasmid and high copy plasmid.

Lysis device in low copy plasmid			
	Exponential phase	**Early stationary phase**	**Late stationary phase**
Lysis entity (%)	76.28 ± 0.3	75.43 ± 1.1	50.1 ± 2.5
Delay after induction (min)	15	15	15
Doubling time (min)	49.8 ± 1.1		
Doubling time of negative control (min)	43 ± 1.3		
Lysis device in high copy plasmid			
	Exponential phase	**Early stationary phase**	**Late stationary phase**
Lysis entity (%)	58.6 ± 7.5	42.2 ± 9.9	0
Delay after induction (min)	15	15	-
Doubling time (min)	69 ± 6		
Doubling time when lysis failed (min)	47.4 ± 2.1		
Doubling time of negative control (min)	45 ± 4.7		



	TOP10[1]	TOP10[2]	DH5α[1]	MG1655[1]
V_{max} [%]	76	77	81	90
K_{50} [nM]	0.37	3.1	0.7	0.9
n	1.3	0.66	3.4	2.8

	TOP10[1]	TOP10[2]	DH5α[1]	MG1655[1]
Exp phase	76	78	88	93
Early stationary phase	75	44	79	85
Late stationary phase	50	38	40	38

	TOP10[1]	TOP10[2]	DH5α[1]	MG1655[1]
1 µM	15	20	17	17
100 nM	15	18	20	17
10 nM	15	20	17	18
1 nM	15	22	15	18
0.1 nM	53			

	TOP10[1]	TOP10[2]	DH5α[1]	MG1655[1]
1 µM	65	75	62	63
100 nM	68	58	72	58
10 nM	76	40	108	90
1 nM	158	26	60	100
0.1 nM	48			

Conditions	
Compatibility	BBa_F2620 and BBa_K098995 in pSB4C5,
Incompatibility	BBa_F2620 in pSB1A2
[1]	LB, 37°C
[2]	M9 suppl., 37°C
[a]	In early stationary phase
[b]	In response to 100 nM of HSL

Figure 2.7: **Lysis device datasheet.** The datasheet contains a summary of all the quantitative parameters that characterize the lysis device. It thus enables the sharing and dissemination of information among the scientific community.

because the activity of components with a high energy demand could be underestimated when present in a huge number of copies per cell. Conversely, promoters in low copy plasmids did not show significant activity differences among the different tested reporter devices, yielding a CV of 22% in the worst case, suggesting the suitability of this vector for the quantitative evaluation of transcriptional activity and for a stable and controlled protein expression.
When bearing the lysis device, low copy number plasmids have been found to be the optimal working condition, as lysis could be triggered in all the growth phases of the bacterial culture and the cells grew with a relatively low metabolic burden, according to their doubling time. All the results have been confirmed in different *E. coli* strains and different growth media, thus providing parameters that can be used in models to aid future biological systems design and to facilitate the reusability of this lysis device in low copy plasmid (see data sheet).
The lysis device in high copy number plasmid gave worse performance, in fact lysis entity was lower than in low copy plasmid, the metabolic burden was much higher and the device was strongly unstable, as cell lysis induction usually failed to occur.
The results presented in these two case studies are in accordance and highlight the necessity of investigating the nonlinear effects on gene expression due to copy number variation, when dealing with the design of gene networks. A deep analysis of these nonlinear phenomena will be presented in this thesis.

Chapter 3

A standard vector for the integration of BioBricks™ in *Escherichia coli*[1]

The study of the effects of DNA copy number variation on the behaviour of gene networks requires the use of synthetic devices able to maintain the studied circuits at a given number of copies.
Commonly used plasmids present in the Registry of Standard Biological Parts [20] enable a 40-fold copy number variation (from 5 copies per cell for low copy vectors to $\sim$ 200 copies per cell for high copy vectors) and would seem to qualify as an attractive instrument to study the problem under investigation, thanks to the ease of manipulating and incorporating them in the host cell. Furhermore, their use presents some drawbacks, mainly due to the necessity of applying a selective pressure for their maintenance and to the nonlinear effects connected with high-energy demand devices in high copy plasmids, as described in the previous chapter. To overcome some of these limitations and to prevent copy number artefacts, plasmid-free strains with the desired genetic circuit integrated in single copy in the genome might be used, as discussed in Sec. 3.1.
Synthetic Biology concepts can be a powerful tool for the development of standard and user friendly integrative vectors to easily produce engineered strains and to rapidly characterize the desired genetic parts in single-copy context.
In this chapter, the design of a novel integrative vector that allows the genomic integration of biological parts compatible with the BioBrick™ standard in *Escherichia coli* is reported (Sec. 3.2).
This vector can be specialized by using BioBricks™ to target the desired integration site in the host genome (Sec. 3.2.1). The vector performance has been evaluated in terms of percentage of success in every step of the integration

[1]The contents of this chapter are published in *S. Zucca, L. Pasotti, N. Politi, M. G. Cusella De Angelis, and P. Magni.* Journal of Biological Engineering *Under revision, 2012.*

procedures and the results are reported in Sec. 3.2.2.
The usefulness of this vector has been demonstrated by integrating a set of BioBrick™ promoters in two different *loci* of the *E. coli* chromosome and by characterizing their activity in single copy (Sec. 3.2.3). Construct stability has also been evaluated and compared with plasmid-borne solutions (Sec. 3.2.4).
The results obtained in this study are discussed in Sec. 3.3. The materials and methods relative to the experiments described in this chapter are reported in App. B. Additional information about the base vector design, the phenotype of recombinant strains, the integration efficiency and additional integration protocols are reported in App. B.2.
This study as been published in [46]. In this thesis work, this vector has been used as a tool for the integration of BioBrick™ parts in *E. coli* chromosome in order to build the single-copy circuits, fundamental in the evaluation of the effects of DNA copy number variation (see Ch. 4 as an example of utilization).

3.1 Background

Plasmids are extensively used tools to generate genetically engineered microbes for the expression of recombinant proteins or complex genetic circuits [47, 48, 2]. Even if they are very easy to manipulate and incorporate in the desired host, many disadvantages affect their usage in both industrial applications and research studies. Commonly used plasmids require the selective pressure of an antibiotic to be maintained in cells, which is costly for industrial scale recombinant protein production [49]. Moreover, the spreading of antibiotics and resistance markers is potentially unsafe for the environment [23]. Selection systems without antibiotics are available, but they require mutant host strains, specific growth media or expensive reagents and in some cases they show low efficiency [50, 51]. Commonly used plasmids are replicated in multiple copy in the host cell and this enables the production of a large amount of proteins in industrial processes. However, plasmid-free strains with the desired recombinant genes in single copy are required in many studies, e.g. to investigate the effect of these genes in normal physiological conditions, thus avoiding copy-number artefacts [52]. Genome integration can provide the stable insertion of the desired genes in the host chromosome without the need of any antibiotic or resistance marker. Several tools for *Escherichia coli* have been proposed which exploit homologous or site-specific recombination. In particular, homologous recombination can be used to insert the desired DNA fragment (here called passenger) into a specific genomic *locus* that must show sufficient sequence homology with a second DNA fragment (here called guide) used to target the *locus* [53]. Apart from genomic integrations, this technology can be used for general genome engineering, such as gene knockout and mutation of native genes [54]. Integrative plasmids (that perform single- or double-crossover) [55], linear PCR fragments [56] and also single-stranded DNA [57] can be used to perform the described tasks by means of unspecific

endogenous or heterologous protein machinery (e.g., the λRed system [56, 57]). On the other hand, site-specific recombination uses the mechanism of genome insertion of bacteriophages in the host chromosome through the phage attachment site (attP) and the bacterial attachment site (attB) sequences, which can be found in nature on the phage and bacterial genome respectively [58]. The process is mediated by a specific phage recombinase. A number of attP sites from native phages have been characterized, as well as the specific recombinases that target them into the specific attB site. This mechanism has been exploited for the development of integrative vectors that carry the attP site (guide) and the passenger of interest [59]. The gene expression machinery that mediates homologous or site-specific recombination can be placed on an easily curable helper plasmid transformed in the host strain [53, 59]. As this work is focused on integrative plasmids, Fig. 3.1 illustrates how homologous recombination (with a single-crossover event) and site-specific recombination work.

Generally, integrant clones are selected with an antibiotic resistance marker. It can be removed by exploiting FRT sites: by flanking a sequence with FRT sites, it can be targeted for excision through the yeast Flp recombinase. Helper plasmids which expressed the Flp recombinase have also been constructed [60]. Integrative plasmids must be easily amplified *in vivo* and clones with a successful integration must be easily selected. To this aim, conditional-replication origins are exploited. They support plasmid replication only in specific conditions, like a specific strain or a temperature range, while the plasmid becomes non-replicative otherwise [54].

For example, the R6K replication origin can be used to propagate integrative plasmids only when the pir or pir-116 gene is present in the host strain [59]. Temperature sensitive replication origins are also commonly used to cure helper plasmids. Recent advances in the field of Synthetic Biology include the standardization of biological parts to facilitate the assembly of genetic circuits [59]. BioBricks™ in the Registry of Standard Biological Parts are a rapidly-growing collection of DNA parts that conform to a specific physical standard [20]. BioBrick™-compatible genetic circuits can be easily incorporated in a microbial host through ad-hoc constructed BioBrick™ plasmids with the desired replication origin and antibiotic resistance marker [16]. However, although the construction of integrative systems by using BioBricks™ has already been reported [23], no standard and ready-to-use solution is still available to produce engineered strains with BioBricks™ via chromosomal integration. In this chapter, the design of a BioBrick™-compatible integrative vector is reported. This device allows the genomic integration of BioBricks™ and can also be specialized to target the desired integration site in the host genome by using BioBrick™ parts. The usefulness of the designed tool has also been demonstrated by providing data on modularity of promoters when characterized in a single-copy context and on plasmids.

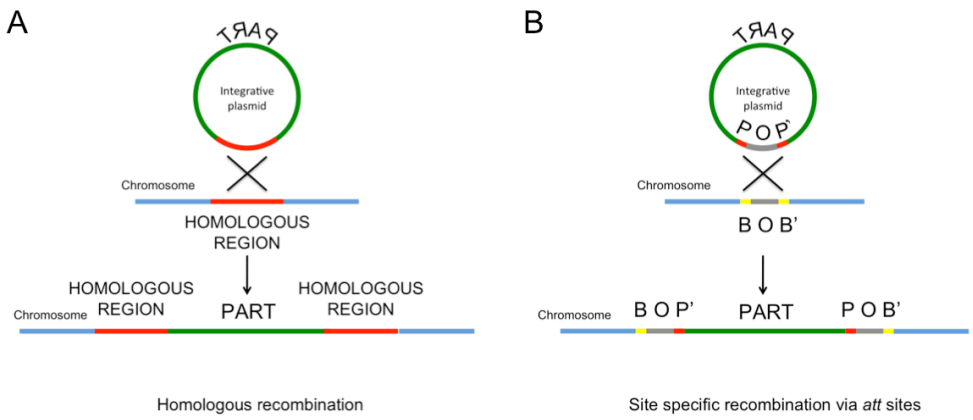


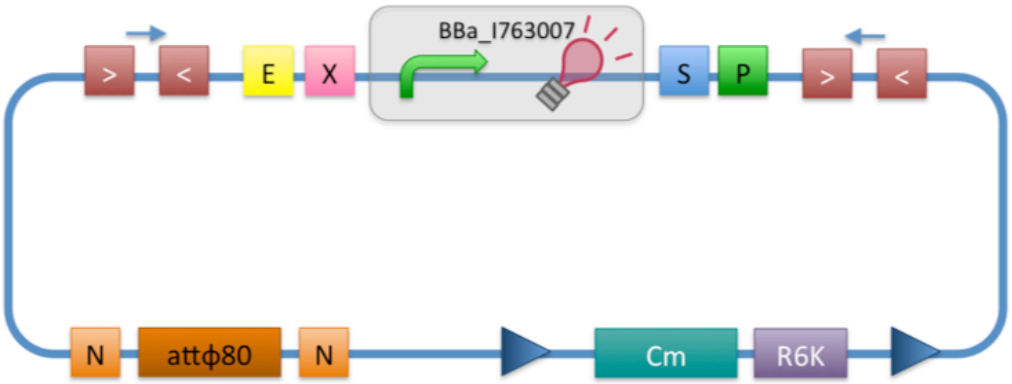
Figure 3.1: **Main plasmid-based methodologies for chromosomal integration of parts in *E. coli*.** A) Homologous recombination. The integrative plasmid (green) carries a sequence (red) that is homologous with a region (red) in the bacterial chromosome (blue). After a single-crossover event, mediated by the endogenous recombination machinery of *E. coli*, the whole plasmid sequence is integrated in the target region of the chromosome. In specific recA-knockout strains the recombination machinery can be expressed via a helper plasmid. B) Site-specific recombination. The integrative plasmid carries a bacteriophage attachment site (attP) that targets the whole plasmid into the specific attachment site (attB) in the host genome. This process is mediated by a specific recombinase that can be expressed via a helper plasmid. The attP and attB sites are composed by the POP' and BOB' sequences, respectively. They share a homologous core sequence (O) and different flanking sequences (P, P', B and B'), so that after site-specific recombination the integrated plasmid is flanked by the BOP' and POB' sequences. All the bacteriophage specific att sequences have this common structure. In both homologous and site-specific recombinations, positive integrants are usually selected via antibiotic resistance, provided that the integrative plasmid sequence contains an antibiotic resistance marker.

3.2 Results and Discussion

3.2.1 Design of the BBa_K300000 integrative base vector

Fig. 3.2 shows the structure of the designed integrative vector, pBBintϕ. It can be considered as a base vector where the default integration guide is the ϕ80 attP and it can be specialized according to user needs by changing guide and passenger. The cloning site is compatible with the original BioBrick™ standard (RFC10) and its related standards RFC23 and RFC12, as it is composed by the original BioBrick™ Prefix and Suffix [11]. The presence of illegal restriction sites (XbaI in FRT and SpeI in the ϕ80 attP) prevents the usage of this backbone in the classic BioBrick™ Standard Assembly process. However, the presence of unique EcoRI and PstI sites in Prefix and Suffix fully supports the assembly of the desired BioBricks™ in the cloning site upon EcoRI-PstI digestion and also supports the 3A Assembly [16, 17] (see Fig. 3.3 A). The two NheI restriction sites flanking the default integration guide enable the engineering of this backbone by assembling new user-defined BioBrick™ integration guides upon XbaI-SpeI digestion, if the desired guide conforms to the RFC10 or a compatible standard (see Fig. 3.3 B). Such assembly is non-directional, however the integration occurs regardless to guide orientation. Like in many other standard vector backbones (e.g., the pSB**5 vector series in the Registry of Standard Biological Parts [20]), the binding sites for standard primers VF2 and VR are present upstream and downstream of the BioBrick™ cloning site respectively. These two sequences are sufficiently distant from the cloning site to enable a good quality sequencing of the insert. The R6K conditional replication origin is necessary to avoid the extra-chromosomal maintenance of the vector during integration and specific strains with the pir or pir-116 genes are required to propagate it during the cloning steps.
The four BioBrick™ transcriptional terminators BBa_B0053, BBa_B0054, BBa_B0055 and BBa_B0062 ensure the transcriptional insulation of the integrated part from its flanking genome sequences, as it is achieved in the pSB**5 vector

Figure 3.2 *(facing page)*: **Integrative base vector pBBintϕ structure** It is a Chloramphenicol-resistant vector with a conditional replication origin (R6K) that impairs its replication in commonly used *E. coli* strains (without pir or pir-116 gene). Two FRT sites flank the resistance gene and the conditional origin, so that they can be excised via Flp-mediated recombination once integrated in the genome, thus leaving a marker-less integrant without R6K. This vector targets the ϕ80 attB site in the chromosome of *E. coli* via the attP integration guide in the vector. The attP sequence is flanked by two NheI restriction sites to enable the engineering of this base vector by easily changing the guide. The cloning site is flanked by BioBrick™ Prefix and Suffix sequences, while four transcriptional terminators implement the insulation of the part when placed in the genome. Primer binding sites are also present. The default insert of the vector is an mRFP1 constitutive expression cassette, driven by the P_R promoter. The glossary explains all the used symbols and lists the BioBrick™ basic parts used to compose this vector.



Part number	Function	Notation
BBa_G00000	BioBrick cloning site prefix	E X
BBa_G00001	BioBrick cloning site suffix	S P
BBa_I763007	PR-mRFP1 expression cassette	BBa_I763007
BBa_J72001	FRT recombination site	
BBa_B0053 BBa_B0054	forward transcriptional terminators	>
BBa_B0055 BBa_B0062	reverse transcriptional terminators	<
BBa_G00100	forward sequencing primer annealing site (VF2)	
BBa_G00102	forward sequencing primer annealing site (VR)	
BBa_B0045	NheI restriction site	N
BBa_K300032	Chloramphenicol resistance marker	Cm
BBa_J61001	R6K conditional replication origin	R6K
BBa_K300991	attϕ80 integration guide	attϕ80

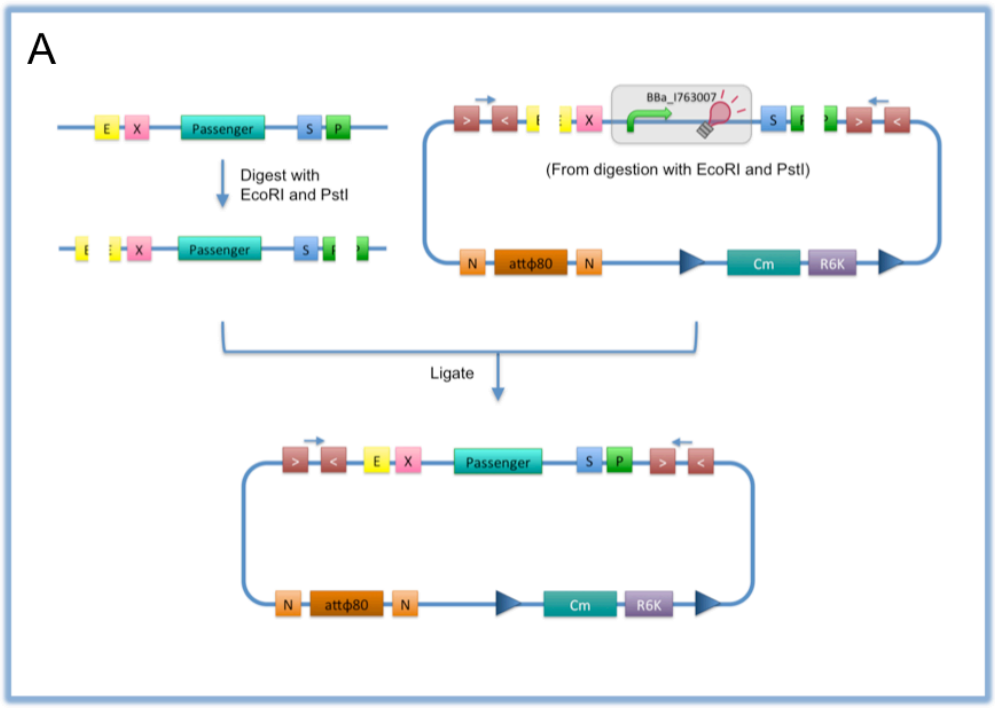
series. The two FRT recombination sites enable the excision of the R6K origin and the Chloramphenicol resistance marker upon Flp recombinase activity. This marker excision allows users to make multiple serial integrations in the same strain in different target *loci*, always using the same antibiotic resistance marker. The same FRT recombination procedure can also be used, provided that essential chromosomal genes lie between the different target *loci* (see [56] and [59] for a detailed description of such potential problem when multiple FRT sites are introduced in the same genome).

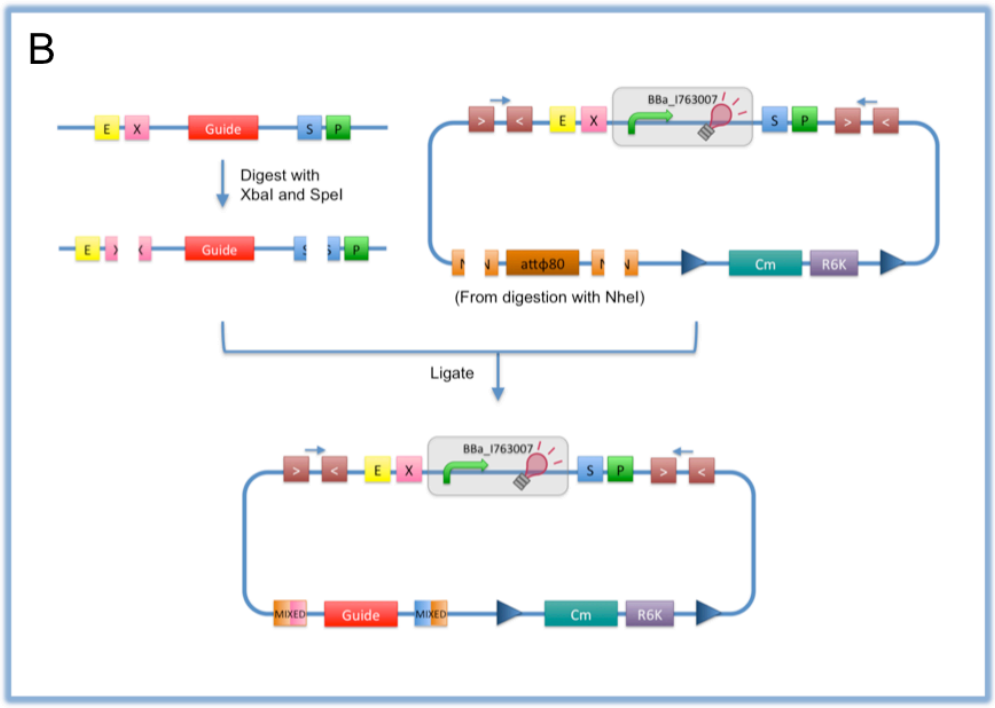
Additional details about vector sequence, features and construction are available in the BBa_K300000 Registry page. The engineering of the integration guide allows the integration of parts in user-defined genome positions and for this reason this vector supports the integration by exploiting bacteriophage attP-mediated integration as well as homologous recombination.
In this study, the default insert of pBBintϕ is BBa_I763007, a constitutive mRFP1 expression cassette driven by the P_R promoter from lambda phage. An early version of the integrative vector included a different default insert that in our experience, as discussed in App. B.2.1, showed problems.

3.2.2 Vector performance validation in the default integration *locus*

The integration and marker excision capabilities of pBBintϕ were tested by integrating a number of BioBrick™ passengers in the chromosome of MG1655 and MC1061 strains. The integrative base vector targets the default integration locus ϕ80 attB.
Integration: 100% of the screened colonies (N=39) on the Chloramphenicol plate lost the helper plasmid and 100% of the screened clones (N=11) showed a correct integration position by PCR (see Fig. B.2). 82% of the screened clones (N=11) had at least two tandem copies of the integrated DNA, as PCR showed (see Fig. B.3).
Marker excision: 100% of the screened colonies (N=100) lost the helper plas-

Figure 3.3 *(facing page)*: **How to engineer the base vector.** A) The desired BioBrick™ insert can be assembled as passenger in the integrative vector: the BioBrick™ passenger of interest must be digested with EcoRI-PstI, while the integrative vector is also digested with EcoRI-PstI to eliminate the default insert BBa_I763007. The two parts can be ligated and the product has to be transformed into a pir/pir-116 strain to enable the propagation of the vector. B) The vector can be specialized with BioBricks™ to target the desired *locus* of the *E. coli* chromosome: the BioBrick™ guide of interest must be digested with XbaI-SpeI, while the integrative vector must be digested with NheI to eliminate the default guide (ϕ80 attP). The open vector must be dephosphorylated to avoid self-ligation, the two parts can be ligated and the product has to be transformed into a pir/pir-116 strain. XbaI, SpeI and NheI all have compatible protruding ends. Note that the ligation is not directional, but the guide can work in both directions. The guide can be another attP site or a part which shows a significant homology with a genomic region of the host.


A
E
X
Passenger
S
P
Digest with
EcoRI and PstI
BBa_I763007
(From digestion with EcoRI and PstI)
N
attϕ80
Cm
R6K
Ligate


B
Guide
Digest with
XbaI and SpeI
BBa_I763007
(From digestion with NheI)
Ligate
MIXED

mid, while 68% of them had a successful marker excision, validated via Chloramphenicol resistance. 94% of the Chloramphenicol-sensitive screened clones (N=33) also showed a correct amplicon (see Supplementary Fig. B.4 for a representative experiment) by PCR with primers P1-P4 (see App. B.1.5), thus validating the presence of the construct of desired size in the correct genomic position. In this case, the multiple tandem copies, previously identified by PCR, became a single copy without antibiotic resistance or R6K origin. This happened because the Flp enzyme excised the entire sequence flanked by the two most distal FRT sites, thus generating a single integrant of the desired construct.
100% of the screened clones with correct P1-P4 amplicon (N=12) showed the expected sequencing results. 83% of them also showed the expected phenotype. App. B.2.2 reports the full list of integrated BioBricks™ (Tab. B.3) and additional information on integration experiments.

3.2.3 Characterization of a set of BioBrick™ promoters in two different genomic positions

A representative set of widely used BioBrick™ constitutive promoters with the RBS-mRFP1-terminator sequence downstream were integrated in the ϕ80 or aspA genomic *locus* and characterized. The considered promoters were J23100, J23101, J23118, P_{lacIQ} and PR (from lambda phage), while the downstream sequence was the BioBrick™ BBa_I13507. These constructs were also characterized in the low copy vector pSB4C5, used as a term of comparison. This set of promoters has been previously characterized in different strains and conditions (see Ch. 2 and [32]).

Fig. 3.4 shows the RPU values of promoters measured in the three different physical contexts. Given a promoter and a physical context, the activity of the J23101 promoter in the same context was used as a reference to compute RPUs. While the activity of J23100 and PR is not statistically different among the three contexts, J23118 and P_{lacIQ} activities show a significant difference in one of the tested contexts with a CV of 20% and 31% respectively. The activity of P_{lacIQ} in plasmid context is different from its activity in the chromosome. This could be due to the sequence upstream of the promoter that is identical for the two genomic contexts, but it is different in the pSB4C5 vector and it could affect promoter activity as shown previously for different promoters [32, 30]. The different upstream sequence is not sufficient to explain the difference observed in the J23118 activity, where the promoter shows the same activity in low copy vector and aspA *locus*, but a slightly lower activity in the ϕ80 locus.

Fig. 3.5 shows the absolute promoter activities ($S_{cell,norm}$ value, see App. B.1.8) in the two genomic *loci*. These results suggest that promoters in the aspA *locus* have a higher expression rate than promoters in the ϕ80 locus. Although statistical analysis highlighted only two promoters where the aspA-integrant had a significantly higher expression than the ϕ80 one, this trend is systematically

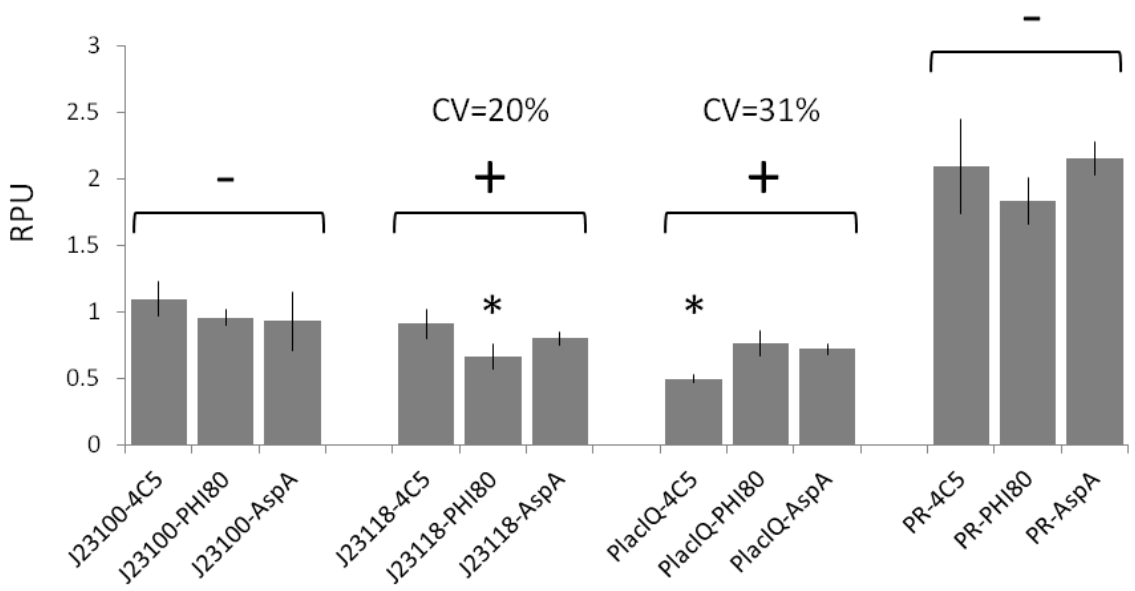


Figure 3.4: **BioBrick™ promoters characterization, expressed in Relative Promoter Units, in single chromosomal copy (ϕ80 and aspA *loci*) and in low-copy context (pSB4C5 vector).** Promoter activity was measured via the BBa_I13507 (RBS BBa_B0034-mRFP1-double terminator) part. Grey bars represent the average values, computed on at least four clones, and the error bars represent the standard deviation. For each promoter, statistical analysis was performed via ANOVA test to compare the RPU activities measured in the three different contexts. Conditions showing a statistical difference ($P<0.05$) in the mean activities among the three conditions are marked with a + sign, while promoters not showing any significant difference ($P \geq 0.05$) are marked with a - sign. Asterisks represent the post-hoc comparison results. They indicate the individual significantly different condition.

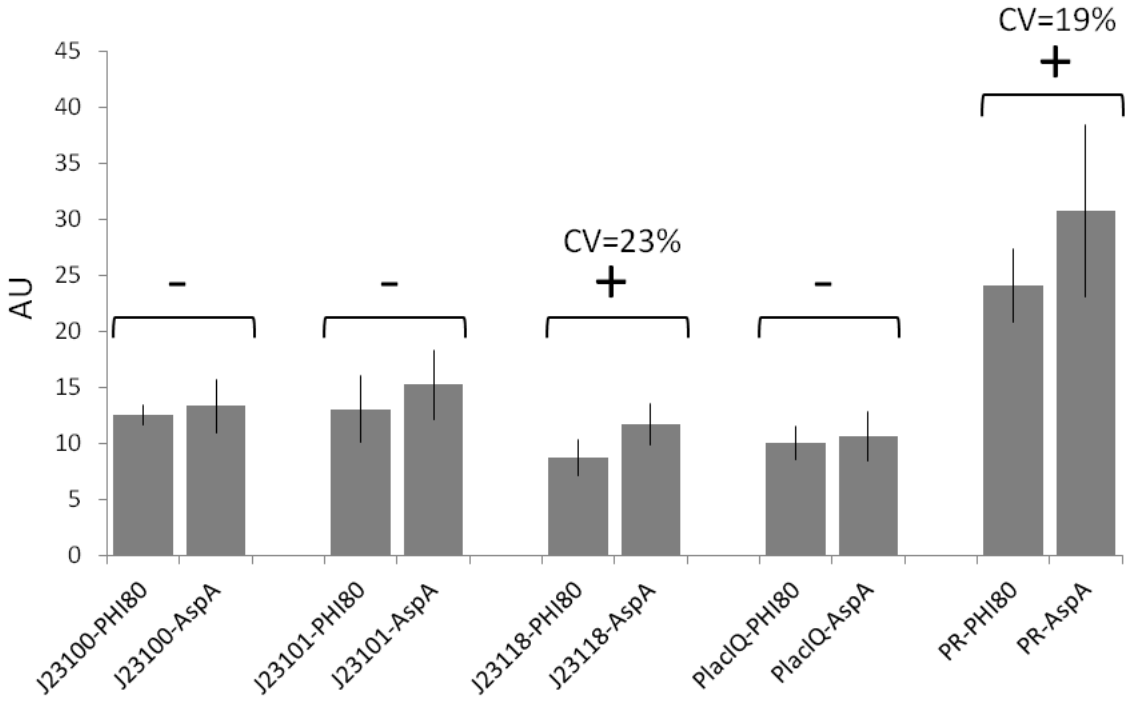


Figure 3.5: **BioBrick™ promoters characterization, expressed in absolute units, in the two single-copy contexts (ϕ80 and aspA *loci*).** Promoter activity was measured via the BBa_I13507 (RBS BBa_B0034-mRFP1-double terminator) part. Grey bars represent the average values, computed on at least four clones, and the error bars represent the standard deviation. For each promoter, statistical analysis was performed via t-test to compare the activities measured in the two different contexts. Comparisons showing a statistical difference ($P<0.05$) in the mean activities among the two conditions are marked with a + sign, while promoters not showing any significant difference ($P \geq 0.05$) are marked with a - sign.

present in all the tested promoters. This was expected, since the ϕ80 *locus* is near the replication termination sequence in the *E. coli* chromosome, while aspA is near the replication origin, thus being present in a higher copy number during DNA replication in growing cells [61]. Finally, the average estimated copy number of the low copy vector is about 7 (measured by dividing the absolute promoter activity in low copy by the one in single copy number), which is consistent with the theoretical copy number of the pSC101 origin [20, 16] and also with previous reports [31, 62].

3.2.4 Evolutionary stability of the integrated BioBricks™

All the recombinant strains tested in the previous section were also studied during continual bacterial growth for 150 generations. Fig. 3.6 shows the average percent activity of promoters over the generations for each investigated condition: devices integrated in the ϕ80 or aspA *locus* or contained in a low

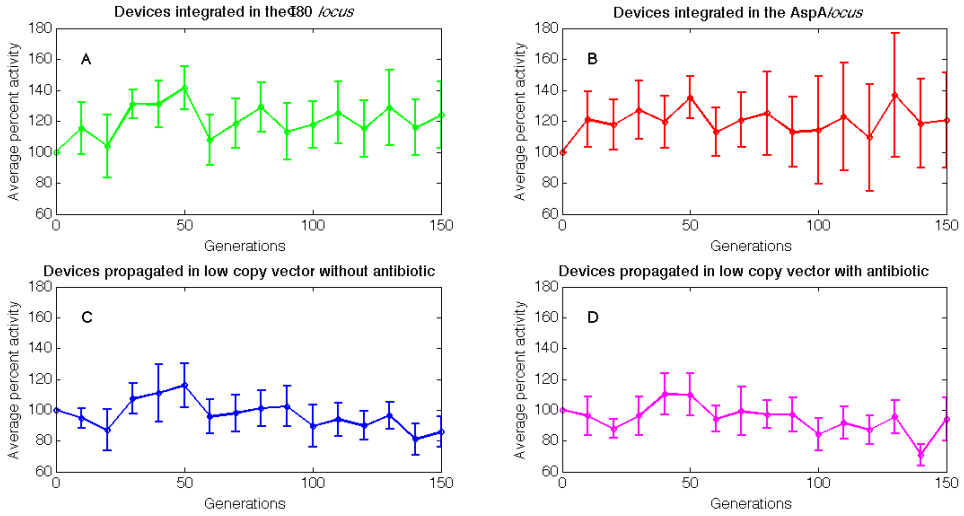


Figure 3.6: **Evolutionary stability of the studied BioBrick™ promoters expressing RFP.** The devices integrated in the ϕ80 (A) and aspA (B) *loci* were propagated without antibiotic, while the devices carried on the low-copy vector pSB4C5 were propagated without (C) or with (D) antibiotic. The four panels show the average percent activity of all the five studied promoters over time in each condition, where 100% represents the average activity at generation 0. For each time point, error bars represent the standard deviations of the measured values.

copy vector propagated with or without antibiotic. Fig. B.5 in App. B.3 shows the results for each individual promoter in the above mentioned conditions. All the recombinant strains are reasonably stable after 150 generations, with only one of the J23100 samples in the aspA *locus* showing a significant activity loss after >80 generations (see Fig. B.5 panel A). Surprisingly, promoter activities are stable after 150 generations even in the low copy condition without antibiotic. These results confirm that integrated constructs are successfully maintained in both *loci* of the bacterial chromosome with stability performance comparable with the low copy plasmid condition.

3.3 Conclusions

Several methods have been proposed to target the desired genetic construct into a specific chromosomal locus [63] or even randomly in the genome [64]. Many studies also focused on single-copy integrants generation [59], on the elimination of antibiotic resistance [56, 64] or on the copy number amplification of the integrated device [65, 66]. Although several genetic tools have been proposed in literature to perform the tasks described above, no plasmid-based standard tools are available to rapidly construct a ready-to-integrate genetic system.

One of the main goals of Synthetic Biology is to apply engineering princi-

ples to facilitate physical construction of biological systems, for example via the definition of proper standards, such as the BioBrick™. Another key concept of Synthetic Biology is to design the desired genetic system from a set of well-characterized parts, whose quantitative behaviour can be included in specific datasheets. In this framework, genetic and computational tools that support the characterization of parts are needed. In particular, the characterization of components in genomic context is an important step that requires a proper insulation from the flanking chromosomal sequences to avoid transcription read-through from/to the integrated part and to avoid position effects caused by flanking DNA which could, for example, unpredictably disturb the activity of the nearby promoter [30]. This work describes a BioBrick™ integrative base vector for *E. coli*. Its standard physical interface allows the assembly of the desired parts as passenger or integration guide, both in BioBrick™ format. This vector is desirable to rapidly disclose, via quantitative characterization, if the parts under investigation are suitable in the genomic context and, if required, in single-copy. The design of this vector was inspired by the works of [16] and [23]. In particular, Shetty et al. [16] conceived a BioBrick™ base vector that could be specialized to construct BioBrick™-compatible vector backbones by using BioBrick™ as well. For example, vectors with different replication origins or antibiotic resistance markers could be easily assembled from a standard ancestor. Before the work of Shetty et al., the construction of vector backbones required specific genetic manipulations whose knowledge is normally owned by experts.

On the other hand, Anderson et al. [23] used BioBrick™ parts to construct different integrative vectors to produce methylase-expressing stable strains without antibiotic resistance marker. Our work merges such concepts by providing a ready-to-engineer base vector containing all the required features for integration (i.e. a conditional replication origin and an antibiotic resistance marker), marker excision (i.e. FRT sites flanking antibiotic marker and conditional origin), assembly of BioBricks™ as passenger (i.e. BioBrick™ Prefix and Suffix) or guide (i.e. NheI sites flanking the default guide). The structure of the vector enables the integration of parts properly isolated from the genomic context. In this way, all the integrated sequences have the same upstream and downstream sequences. Whatever the integration procedure is, the compatibility of the vector with the target chassis always has to be verified: the presence of the integration locus must be evaluated according to the declared genotype, genomic databases or by PCR. In addition, when using site-specific recombination, no prophage has to be present in the desired attB *locus* of the host strain. While our aim was to provide a physical support to standardize the integrative vector construction that can be used with the desired integration procedure (site specific or homologous recombination by single crossover), this work is not intended to contribute novel integration protocols or methodologies that improve its efficiency. In fact, the efficiency depends on the specific guide, passenger and protocol used. For example, in this work the ϕ80 site-

specific recombination had a very high success rate while the aspA homologous recombination, used to construct some of the recombinant strains, showed a lower efficiency (see App. B.2.3). A set of BioBrick™ devices, including widely used constitutive promoters assembled with a proper measurement system, were used as passengers in most of the integration experiments. These devices were quantitatively characterized via RFP measurement to evaluate their behaviour in single copy in two different *loci* and in a low copy plasmid. The measured promoter activities were consistent with previous works (see Ch. 2 and [32, 31]). Given a promoter, its relative activity was comparable among the tested contexts, with a maximum CV of 31% (P_{lacIQ} promoter). Although the different sequence between the upstream regions in the single- and low-copy contexts could be responsible of this activity variation in the case of PlacIQ, the sequence difference could not explain all the observed variability among the contexts. The detected variability entity was also consistent with our previous studies on context-dependent variability of these promoters when tested via different measurement systems or when assembled in different genetic circuits [32]. The absolute activity of each integrated promoter was also compared in the two chromosomal positions. Results highlighted a position-dependent effect where the promoters near the genomic DNA replication origin (aspA *locus*) expressed higher amounts of RFP than promoters near the replication termination sequence (ϕ80 *locus*), as expected. This demonstrated the importance of quantitative characterization of parts in different genomic contexts for the predictable design of genetic functions.

Finally, the stability of these integrated BioBrick™ devices has been evaluated. This experiment demonstrated that integrated solutions had comparable stability with plasmid-based ones. As the used parts were not toxic or hard-to-express parts for the chassis, this experiment only represented a proof-of-concept, while the stability of parts has to be tested for each desired construct in the future and it could be dependent on the integration site, DNA sequence, host organism, metabolic burden and/or specific function implemented by the integrated device.

Chapter 4

Characterization of an inducible promoter in different DNA copy number conditions[1]

Currently, a predominant problem connected with the construction of even simple synthetic biological systems is the unpredictability of the genetic circuitry when assembled and incorporated in living cells, due to, for example, copy number artefacts, transcriptional/translational demand and toxicity of the DNA-encoded functions. It is fundamental to investigate the linearity working boundaries of engineered biological systems when dealing with such phenomena.
For this purpose, the analysis of the behaviour of a simple genetic device varying the DNA copy number has been performed and will be illustrated in this chapter.
An HSL-inducible device (Lux system, see Sec. 2.2.2 and Fig. 2.5) driving the expression of a reporter protein (RFP) has been characterized in different copy number contexts, ranging from 1 copy per cell (integrated in the genome) to hundreds (via multicopy plasmids) (Sec. 4.1).
First, two constitutive promoters (the reference standard J23101 and P_{LacIQ}) have been characterized in the four copy number contexts under investigation. The ratios of the activities in multiple copy relative to the single copy (which has a copy number = 1 by definition) have provided a rough and indirect estimation of the DNA copy number per cell (Sec. 4.2.1).
Subsequently, the HSL-inducible system has been characterized in the same copy number contexts and the input-output transfer function of the device (in terms of RPUs produced for HSL induction) has been determined (Sec. 4.2.2). Since the copy number of the transcription factor (*luxR*) and of the lux promoter varies in concert, a mathematical model has been exploited, in order

[1]The contents of this chapter are published in *S. Zucca, L. Pasotti, G. Mazzini, M. G. Cusella De Angelis, and P. Magni* BMC Bioinformatics*, 2012.*

to predict the not measured amount of the Lux-HSL complex (A), the actual activator of lux promoter. The input-output characteristic has been expressed as a function of A (Sec. 4.2.3).
The materials and methods relative to the experiments described in this chapter are reported in App. C. Additional results relative to the validation of the measurement system and single-cell analysis are reported in App. C.2.
Even in this simple inducible system, nonlinear effects have been observed and non-trivial data processing has been necessary to fully characterize its functioning. The in-depth analysis of model systems like this can contribute to the advances in the Synthetic Biology field, since increasing the knowledge about linearity and working boundaries of biological phenomena (such as copy number tuning) could lead to a more rational design of artificial systems, also through mathematical models, which, for example, have been used here to study hard-to-predict interactions (Sec. 4.3).

4.1 Background

Incorporation of genetic programs in a host, such as a bacterial cell, can be performed either through plasmid vectors, which are able to autonomously replicate in the organism and propagate the DNA-encoded synthetic functions to the progeny [47, 48], or through genomic integration in which the program is stably kept in the cell in single copy [59]. Plasmids are maintained in cells at a copy number ranging from 1 or 2 to hundreds of copies, depending on their replication origin [16].

In the literature, several mathematical models based on the differential equations have been proposed to describe the output of a synthetic circuit, often in terms of a synthesized mRNA or protein amount, as a function of its DNA copy number [67, 10, 9]. Considering the simplest system composed by one promoter and one gene of interest downstream, its dynamic behaviour is governed by the following equations which include transcription and translation processes [8]:

$$\frac{d[M(t)]}{dt} = n \cdot r(t) - d \cdot [M(t)] \tag{4.1}$$

$$\frac{d[P(t)]}{dt} = \rho \cdot [M(t)] - \gamma \cdot [P(t)] \tag{4.2}$$

where squared brackets indicate a per-cell concentration, n is the copy number of the DNA, M is the mRNA, P is the protein, r is the mRNA synthesis rate per DNA copy, ρ is the protein synthesis rate per mRNA and d and γ are the degradation rates of M and P, respectively. Assuming the steady state, it results that $[\bar{M}] = \frac{n\bar{r}}{d}$ and $[\bar{P}] = \frac{n\rho\bar{r}}{d\gamma}$, where bar indicates that the species is at the steady state (constant value). Even if these expressions show that both mRNA and protein concentrations are theoretically linear functions of n, a number of works report that both transcription and translation output

may not change linearly with the DNA copy number and yield hard-to-predict system outputs [68, 52, 69, 44]. In particular, Hajimorad et al. [68] showed that the mRNA level of one or more gene expression devices changes linearly with the device copy number, but only in specific conditions, i.e. limited copy number and number of different devices in the same cell. In general, all the genetic manipulations that cause host overburdening may contribute to nonlinear effects on biological systems. It is also known that highly expressed recombinant genes can lead to saturation effects, caused by the overloading of endogenous transcriptional and translational machinery, partly for the limited availability of RNA polymerases and ribosomes [70, 69].
As shown in Ch. 2, unpredictable outcomes can occur when the expressed proteins are toxic, but also when simple non-toxic gene expression cassettes are incorporated in the chassis.
In this study, an inducible promoter, which can be regulated over a wide range of transcriptional activities, was studied in *Escherichia coli* in single- and multi-copy contexts. The system is based on the widely studied luxR/lux promoter system [9, 71] and it is able to produce a red fluorescent protein (RFP) upon 3-oxo-hexanoyl homoserine lactone (HSL) molecule addition to the bacterial culture in a concentration-dependent fashion (see Fig. 4.1 for the inducible system description).

The DNA encoding the inducible system was placed in the bacterial genome or in plasmids with the pSC101, p15A or the mutated pMB1 replication origins, which yield low, medium or high copy numbers, respectively. The number of DNA copies per cell has been previously reported to be $\sim$5, 20–30 and $>$100 for these three origins [62, 16]. The relative promoter unit (RPU) approach [10] was used to indirectly measure the activity of the lux promoter from RFP fluorescence data for each investigated induction and device copy number context. In such framework, it is possible to investigate if the ratio between the outputs of synthetic devices, in terms of synthesized RFP, is maintained as a function of induced promoter strength and in different copy-number contexts. The results produced in this work can contribute to improve the characterization of a simple but widely used genetic device and such findings can represent a step towards the study of linearity boundaries of Synthetic Biology components when incorporated in living cells, thus enabling a more rational design of biological systems. An in-depth experimental analysis of such nonlinearities may also contribute to the advances in the field of systems biology, as increasing the knowledge about the working conditions of biological systems could lead to the creation of new mathematical models of higher accuracy, useful to computationally study or predict complex biological phenomena through simulations.
The description of all the *Escherichia coli* strains, vector backbones and genetic devices used in this work is shown in Tab. 4.1.

Table 4.1: **Strains, plasmids and biological devices used in this work.**

Strains		
Name	Genotype	Source
DB3.1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm^r) xyl5 Δleu mtl1	Invitrogen
BW23474	F-, Δ(argF-lac)169, ΔuidA4::pir-116, recA1, rpoS396(Am), endA9(del-ins)::FRT, rph-1, hsdR514, rob-1, creC510	CGSC, Yale University, USA
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ-	Invitrogen
MG1655	F- λ- ilvG- rfb-50 rph-1	CGSC Yale University, USA
MG-HSL_{RFP}	MG1655, Φ80(HSL_{RFP})	This study
MG-101_{RFP}	MG1655, Φ80(101_{RFP})	This study
MG-IQ_{RFP}	MG1655, Φ80(IQ_{RFP})	This study

Vector backbones used for RFP expression or genomic integration of the devices			
Name	Replication origin	BioBrick ™ vector code	Antibiotic resistance
pHC	pUC19-derived pMB1 origin (high copy)	pSB1A2	Ampicillin
pMC	pMR101-derived p15A origin (medium copy)	pSB3K3	Kanamycin
pLC	pSC101 origin (low copy)	pSB4C5	Chloramphenicol
pΦ80	R6K conditional origin, used for integration	BBa_K300000	Chloramphenicol
Genetic devices			
Name	Description		
HSL_{RFP}	HSL-inducible mRFP1 expression system expression system		
HSL_{GFP}	HSL-inducible GFPmut3b expression system expression system		
101_{RFP}	Standard reference constitutive promoter with mRFP1 expression device downstream		
101_{GFP}	Standard reference constitutive promoter with GFPmut3b expression device downstream		
IQ_{RFP}	lacIQ constitutive promoter with mRFP1 expression device downstream		

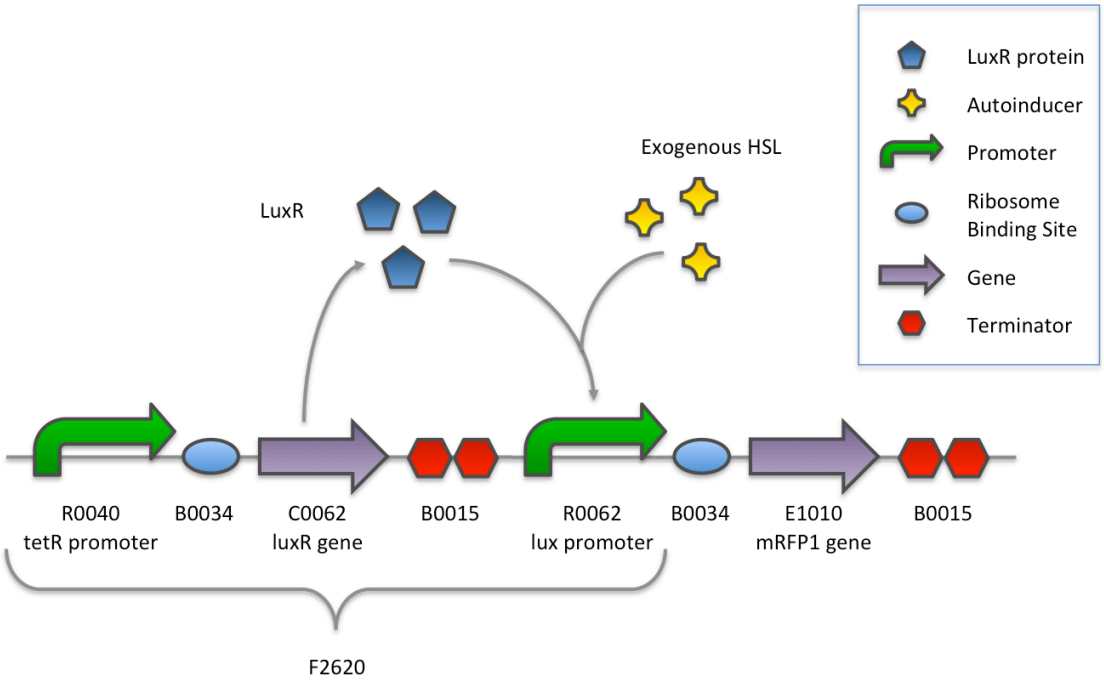


Figure 4.1: **Working diagram of the HSL-inducible system.** The luxR gene is constitutively produced by the tetR promoter and the strong ribosome binding site B0034. The gene encoding for mRFP1 is placed downstream of the lux promoter. LuxR protein is normally unactive, but when HSL is present in the culture it binds LuxR and a complex is formed by two molecules of LuxR and two of HSL. This complex triggers the transcription of the lux promoter in a concentration-dependent fashion.

4.2 Results and Discussion

4.2.1 Per-cell fluorescence of constitutive RFP-producing systems as a function of copy number

In order to show that variations in the copy number in an RFP-producing system cause different fluorescence levels, S_{cell} was measured in the 101_{RFP} and IQ_{RFP} devices. Tab. 4.2 reports the resulting average values, which demonstrate that a wide range of synthesis rate can be achieved in the different conditions. S_{cell} varies more than 100-fold between the two extreme conditions and increases with the DNA copy number as expected. All the S_{cell} values were divided by the corresponding value measured in single copy, thus obtaining a rough and indirect estimation of the DNA copy number per cell [62]. The computed values are reported in Tab. 4.2. They are consistent with previously published copy number results measured in *E. coli* for the same replication origins, except for the medium copy which is about 2-fold higher than expected for a p15A replication origin.

Table 4.2: **Characterization of J23101 and lacIQ promoters in absolute units and indirect copy number estimation** Measurement of the absolute activity of J23101 and lacIQ constitutive promoters via the 101_{RFP} and IQ_{RFP} devices and indirect copy number estimation from average S_{cell} values. AU indicates arbitrary units of RFP.

$\mathbf{S_{cell}}$ **AU** $\cdot$ **cell**$^{-1}$ $\cdot$ **min**$^{-1}$				
	HC	MC	LC	SC
101_{RFP}	88.36	34.2	2.48	0.69
IQ_{RFP}	59.1	22.7	2.14	0.54
Estimated copy number				
	HC	MC	LC	SC
101_{RFP}	128	49	4	1
IQ_{RFP}	110	42	4	1

4.2.2 Induction curves of the HSL inducible system in different copy number conditions as a function of HSL

The HSL-inducible system was characterized in terms of S_{cell} as a function of exogenously added HSL concentrations. Results are reported in Fig. 4.2 for all the copy number amounts. In all the considered situations, induction reaches a steady state value for HSL concentration $>\sim$10 nM. This result is consistent with previously measured induction curves of this HSL-inducible device [9, 44]. Single-cell analysis was also performed to validate if all the cells of an induced culture respond to HSL and the results showed that the population was actually homogeneous for all the tested inducer concentrations (Fig. C.2). Details about this analysis are reported in App. C.2.3.

Tab. 4.3 reports the doubling time of the cultures in the different contexts. The coefficient of variation of the doubling time among different inductions of the HSL_{RFP} device is less than 16%. The average doubling time of the cultures with the HSL_{RFP} device varies only up to 1.4-fold (low copy condition) when compared to the culture bearing the reference device 101_{RFP} or the IQ_{RFP} device. On the other hand, doubling times vary more than 2-fold among different contexts without a specific trend with the copy number. The different antibiotics, plasmids and levels of the expressed heterologous genes may contribute to such unexpected difference in the doubling time values in the exponential phase, even if the chassis is the same. Note that, in our hands, the typical doubling time of the MG1655 strain in the same conditions is 38 (14%) min.

As for the constitutive devices, also for the HSL-inducible system average values of S_{cell}, measured at a full induction, were divided by the value measured in the single copy condition to obtain an indirect estimation of the

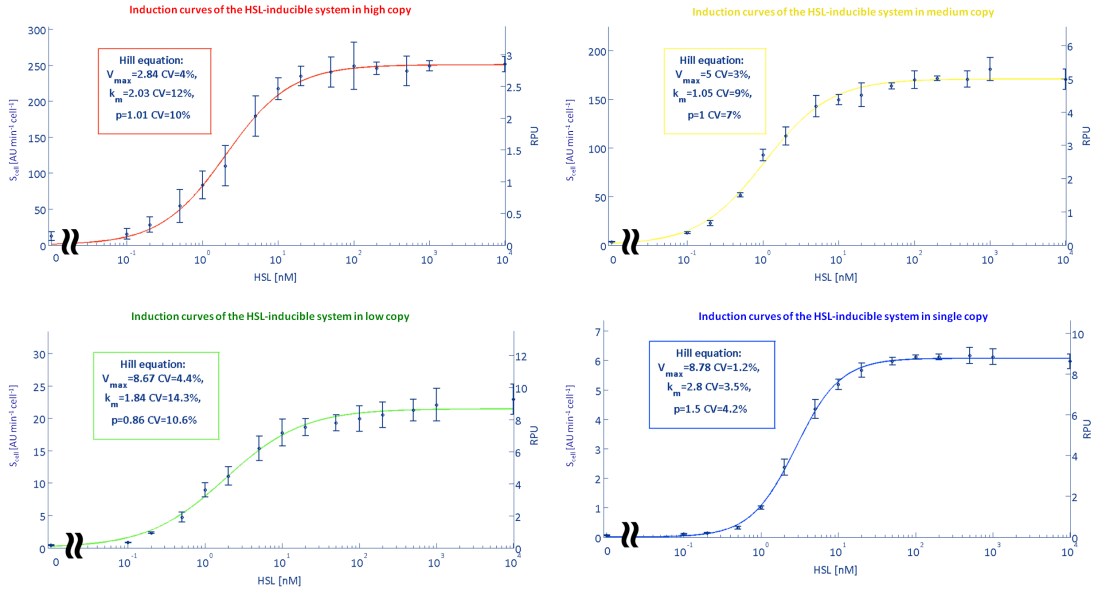


Figure 4.2: **Induction curves for the HSL-inducible system in different copy number contexts.** The HSL-inducible device (HSL_{RFP}) was characterized in the exponential phase after at least 50 minutes from the induction in a microplate reader. For each copy number condition, the curve is expressed in absolute arbitrary units of S_{cell} (left vertical axis) and in RPUs (right vertical axis), computed as the ratio between the S_{cell} of the inducible device and the S_{cell} of the J23101 promoter (via the 101_{RFP} measurement device) in the same growth and copy number condition, considering the same reporter gene (RFP). Experimental data were fitted with a Hill function (continuous line) and the estimated parameters are reported in the boxes with their coefficients of variation. Error bars represent the 95% confidence intervals of the mean value (circles) computed on 3 clones.

plasmid copy number. Resulting ratios are 3.8, 28.5 and 42.1 for low, medium and high copy, respectively. They are quite different from the copy numbers found in the previous section for the two constitutive promoters, apparently yielding fewer DNA copies per cell in medium and high copy number conditions. Moreover, the high copy number value obtained for this system is much lower than previously reported for plasmids with the pUC19-derived pMB1 replication origin in cells grown at 37°C [68]. The growth conditions were the same among the tested cultures and the doubling times were comparable between HSL-inducible system and cultures bearing the constitutive promoters. Therefore, the difference in the estimated copy number is not reliable and more likely the observed effect is due to saturation in transcription and/or translation processes, caused by limited availability of polymerases or ribosomes in the medium and high copy number contexts. In order to enable the comparison of the HSL-inducible system activity among different experimental conditions, the RPU approach was used to express the strength of the lux promoter relative to J23101 constitutive promoter as a reference [10]. RPUs have

Table 4.3: **Doubling times of the studied cultures.** Mean value and coefficient of variation (CV%, between brackets) were computed on 15 differently induced cultures in triplicate for HSL_{RFP} and on 3 clones for 101_{RFP} and IQ_{RFP}.

Doubling time [min]				
	HC	MC	LC	SC
HSL_{RFP}	127 (15%)	63 (7%)	131 (9%)	86 (5%)
101_{RFP}	106 (11%)	59 (8%)	104 (9%)	86 (8%)
IQ_{RFP}	102 (5%)	65 (6%)	92 (5%)	101 (17%)

been reported to produce highly robust activity measurements even in different experimental conditions. Expressing the results in such standard units of measurement also enables the sharing of quantitative characterization results in the Synthetic Biology community. J23101 promoter showed an S_{cell} variation in good agreement with the theoretical copy number of the plasmids used in this work (see Tab. 4.2) and the relative activity of the medium-strength promoter lacIQ could be effectively measured in RPUs in different copy number conditions, yielding highly reproducible results with an average value of 0.744±0.05 (see Fig. 4.3). Taken together, these results suggest that J23101 is a reliable reference for the computation of RPUs in all the copy number contexts. Fig. 4.2 reports the resulting RPUs in the right vertical axis for each curve of the HSL-inducible device. RPU values at full induction (>10 nM) confirm that a saturation trend is present, as single and low copy conditions produce very similar values (RPUs≅8.7), while the activity progressively decreases at medium (RPUs=5) and high (RPUs=2.8) copy contexts. In order to characterize the input-output function of the HSL-inducible device for each specific copy number condition, the curves were fitted with a Hill function, which well described the experimentally measured points. Estimated parameters of the functions are reported in Fig. 4.2. Furthermore, to validate if the observed saturation trend was due to the specific reporter used, the full-induction activity of the HSL-inducible promoter was also measured via GFP (GFPmut3b) instead of RFP (mRFP1) in different copy number conditions. In addition, the GFP reporter device had a different RBS from the RFP device (BioBrick™ BBa_B0032 instead of BBa_B0034). RPU results, shown in Fig. C.3, demonstrate an excellent accordance with the maximum activities reported in Fig. 4.2, thus validating that the saturation trend in medium and high copy number contexts was actually due to the inducible device and not to the used reporter gene.

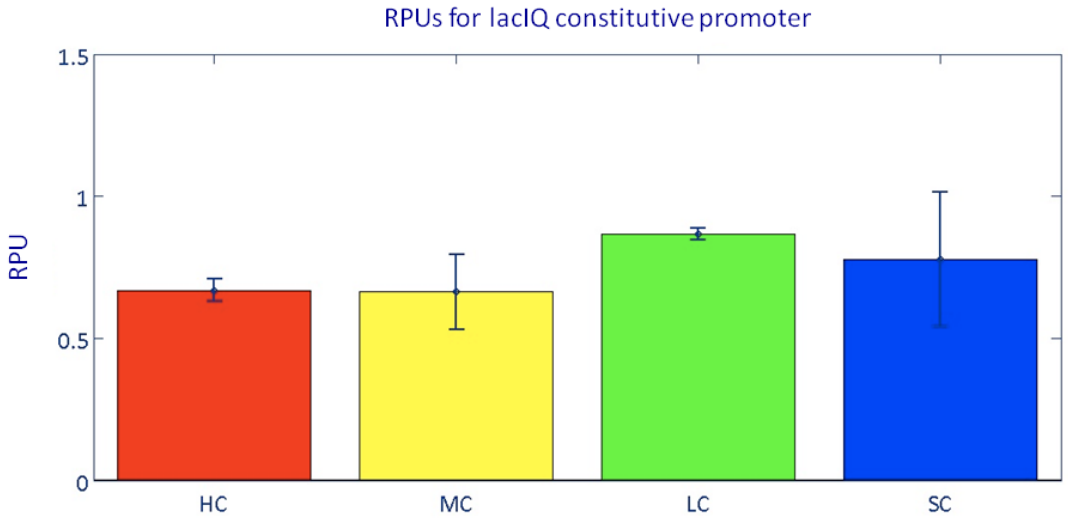


Figure 4.3: **Characterization of the lacIQ constitutive promoter in RPUs.** The ratio between the activities (S_{cell}) of lacIQ and J23101 promoters, measured via IQ_{RFP} and 101_{RFP} devices respectively, was computed to obtain the lacIQ RPUs in the four copy number conditions investigated in this work. Consider that, in this study, IQ_{RFP} and 101_{RFP} have a different DNA scar between promoter and RBS. Unpublished data from our lab showed that the scar present in IQ_{RFP} systematically overestimates promoters activity by 1.43-fold when compared to the scar in 101_{RFP}. Error bars represent the 95% confidence intervals of the mean value computed on 3 clones.

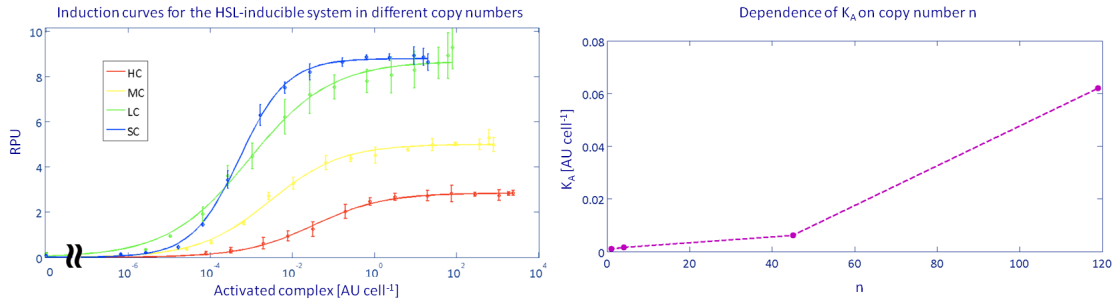


Figure 4.4: **Induction curves as a function of the activated complex and correlation between K_A and n.** Induction curves for the HSL-inducible system as a function of the activated complex HSL-LuxR in different copy number contexts (Panel A) and correlation between the estimated dissociation constant of HSL-LuxR complex and lux promoter (K_A parameter) and copy number (Panel B).The activity of the lux promoter, in terms of RPUs, was studied in each copy number condition as a function of its actual inducer, i.e. the HSL-LuxR activated complex (A), indirectly measured by an *ad-hoc* mathematical model described in App. C.1.7.

4.2.3 Measuring the activity of the lux promoter in different copy number conditions as a function of the activated complex

In all the recombinant strains considered in this study, the DNA copy number of the luxR-expression cassette varies together with the lux promoter-regulated RFP. The different availability of LuxR protein in the investigated conditions makes the HSL-RPU curves described in the previous section not easy to compare among the different copy number conditions, since the amount of the LuxR-HSL activated complex can be different even if the HSL concentration is the same. In order to study the strength of the lux promoter in different copy number conditions, it is necessary to consider the induction curves as a function of the per-cell HSL-LuxR activated complex instead of the HSL concentration. In this way, correct comparisons can be performed to study the strength of the promoter at any induction entity. The mathematical model described in the Methods section was used to evaluate the quantity of the HSL-LuxR complex per-cell (indicated with A) and to express the system output, in terms of RPUs, as a function of the predicted intracellular level of A. The average DNA copy number indirectly measured in this study via RFP measurements in the constitutive devices (see Tab. 4.2) was used to predict the level of available LuxR proteins (X) at the steady state in each copy number condition. K_A, n_A and α_{lux} parameters were estimated from HSL-RPUs experimental data. RPUs are reported in Fig. 4.4 A as a function of A for each copy number context. This figure enables the comparison of the promoter strength as a function of the actual activator level. Fig. 4.2 shows the estimated values of the model parameters. V_{max}, reported in Tab. 4.4, depends on the RPUs

Table 4.4: **Estimated values for the V_{max}, α_{lux}, K_A and n_A parameters.** The coefficients of variation (CV%) are reported between brackets

	V_{max} [**RPU**]	K_A [**AU cell**$^{-}$1]	n_A	α_{lux} [**AU cell**$^{-}$1 **min**$^{-}$1]
HC	2,84 (2%)	3,24e-2 (18%)	0,504 (8%)	2,11
MC	5,00 (1%)	2,87e-3 (13%)	0,502 (6%)	3,73
LC	8,68 (2%)	8,83e-4 (22%)	0,437 (8%)	5,70
SC	8,79 (1%)	5,13e-4 (6%)	0,747 (4%)	6,07

reached for a full induction. α_{lux} is proportional to V_{max} and it decreases as the copy number increases. n_A is similar among the conditions, even though for the single copy it is slightly higher than in the other cases. Moreover, it can be observed that the estimated K_A values are correlated with the copy number. Fig. 4.4 B shows the K_A parameter values as a function of the estimated copy number. Finally, the activated complex-RPUs induction curves show that the lux promoter strength is comparable between single and low copy conditions for all the induction values, while medium and high copy curves show lower strength values than the single and low copy for all the induction levels.

4.3 Conclusions

Because the design of genetic circuits which exhibit a predictable behaviour is the basis of the enormous potential of Synthetic Biology, the investigation of nonlinear phenomena and hard-to-predict engineered cell responses is essential to understand the systems under study and to avoid time- and cost-consuming strategies like trial-and-error approaches. Many variables, such as promoter strength, ribosome binding site efficiency and plasmid copy number, can play a crucial role in systems design. The optimization of a genetic circuitry can be achieved by tuning these elements, but the findings described in literature disclosed nonlinear effects [68, 72]. The copy number of the DNA encoding the engineered genetic network is an important parameter for the regulation of gene dosage and many works reported its tuning to optimize biological circuits such as biosynthetic pathways [73]. However, as it happens with other biological parameters, the linearity of the output of a circuit may not be valid in many copy number ranges, yielding difficult-to-predict saturation effects. Hajimorad et al. studied the mRNA output of simple, independent gene expression cassettes as a function of the copy number, disclosing copy number ranges and conditions in which the superposition of the effects, typical of linear systems, is valid for the studied devices. Here, the aim of our work is to analyze the induction curves produced by an HSL-inducible system as a function of the copy number conditions, considering the fluorescence as the output. The studied system is composed by a luxR constitutive expression cassette and a

RFP expression cassette regulated by the lux promoter, which can be turned on by LuxR in presence of HSL with a concentration-dependent fashion. This device has been previously used in several works [74, 9, 44] and the improvement of its characterization in different copy number contexts can be useful for the rational design of expression systems and cell-cell communication networks involving HSL [40, 75, 76]. In this work, high, medium, low and single copy number conditions were considered. Medium-strength constitutive promoters contained in these plasmids or integrated in the genome exhibit activities that vary more than 100-fold from single to high copy number contexts, which demonstrates that copy number variation can be actually achieved with the used vectors. Moreover, for each constitutive promoter, the fold-change of the fluorescence of plasmid-bearing strains, relative to the fluorescence of the recombinant strain with the device in single copy (assuming a copy number=1 by definition), is in accordance with the theoretical copy number of the used vectors. The strength of the HSL-inducible system was experimentally measured *in vivo* and expressed in RPUs, which are standard units for promoter activity evaluation proposed in literature to lower the variability among experiments. They also enable the sharing of the promoter characterization results in the Synthetic Biology community, as all the activity measurements are performed relative to the activity of a standard reference promoter (J23101), which gives RPUs=1 by definition. Induction curves were obtained in different copy number conditions by exogenously adding HSL to the inducible system under investigation. A saturating trend in the maximum strength was depicted for higher copy number conditions, probably due to limited availability of polymerases or ribosomes which brings transcription/translation processes to saturation, as already reported in literature for other gene expression systems [68, 70]. It is worth noting that, in addition to the lux promoter strength, the presence of the luxR expression cassette may contribute to the saturating trend of the lux promoter maximum relative activity in the higher copy number conditions by increasing the metabolic demand of the device. The same effect was seen for similar recombinant strains with the HSL-inducible device expressing a different fluorescent reporter gene (GFP instead of RFP), thus demonstrating that such phenomenon was not specific for the RFP gene. Characterization of the HSL-RPU transfer function was completed in each copy number context by fitting the curve with a Hill function. As they are expressed in standard units, the results produced here support the re-use of this inducible device, given the specific copy number in which it was characterized. However, these Hill curves do not directly represent the transfer functions of the lux promoter, since it is not simply activated by HSL, but by the activator complex HSL-LuxR. The number of DNA copies of the luxR constitutive cassette varied in concert with the copy number of the rest of the device, so the copy number of this cassette leads to different availability of intracellular LuxR proteins in the different studied contexts. In order to obtain comparable transfer functions among different copy number conditions, an X-axis transformation on the HSL-RPU

induction curves is required to express the promoter strength as a function of the concentration of the actual inducer. A model-based approach has been proposed to indirectly measure the HSL-LuxR activated complex concentration by predicting the intracellular availability of LuxR protein as a function of the expected copy number. This procedure allowed to express the model output (RPUs) as a function of the HSL-LuxR complex (called A). With the help of the mathematical model, the values of the promoter activity curves can be considered as a function of the correct inducer of the promoter, thus allowing comparisons of transcriptional strength in all the studied contexts. The relationship between K_A (i.e. the dissociation constant between A and the lux promoter) and the copy number shows correlation (see Fig.4), while the K_m values found in the input-output induction curves simply as a function of HSL, showed no apparent correlation with the copy number (see Fig. 4.2). However, it is important to note that the used model relies on parameters that have been indirectly estimated in this study (e.g. the plasmid copy number and intracellular concentration of LuxR protein). Moreover, this indirect estimation assumes that the luxR gene expression level per DNA copy (α_{tet}) is constant for each copy number, neglecting any saturation effect in high copy number conditions. Even if more in-depth analysis of the biological systems under study can be performed to measure these parameters and to obtain more accurate relationships, the found behaviour for which the K_A parameter value appears to correlate with the copy number is still valid, since the monotonically increasing trend of K_A was confirmed even after a sensitivity analysis in which the (α_{tet}) parameter was varied 2-fold (data not shown). The same trend was observed even changing the nH parameter from 2 to 1 (data not shown). Finally, a single-cell analysis was performed to improve the characterization of strains bearing the HSL-inducible device. From the literature, it is well known that in some inducible systems not all the cells among the population respond to the inducer, thus causing the simultaneous presence of induced/uninduced subpopulations [77, 78]. On the contrary, results on the HSL-inducible device characterized in this work show that all the cells respond to the induction. This is a very interesting feature for the future usage of such HSL-inducible device in designing more complex circuits.

In conclusion, the characterization in standard, sharable units of a simple synthetic biological device has been performed and the input-output transfer function of the whole device has been reported for different copy number conditions. Even if the device is not significantly complex, nonlinear phenomena were evident in the higher copy number conditions. Moreover, the individual study of the activity of the inducible promoter present in the device was not trivial and it required a model-based approach to indirectly estimate hard-to-measure intracellular species, whose knowledge was necessary to completely characterize the system behaviour and to allow comparisons among different copy number contexts.

Chapter 5

Modeling the effects of promoter and regulator copy number variation on the output of inducible systems in engineered genetic circuits

In this chapter, a mechanistic model to study the effects of promoter and transcription factor copy number variation on a simple HSL-inducible device has been proposed and studied.
First, the pros and cons of empirical and mechanistic models are discussed and the problematic concerning the *in vitro* parameter estimation for *in vivo* studies is introduced (Sec. 5.1). The mechanistic model of the HSL-inducible device based on mass action law is then presented and described and all the initial conditions and parameter values are provided (Sec. 5.2).
The effects of copy number variation of the lux promoter and the luxR transcription factor, individually and in concert, are presented and discussed in Sec. 5.3 and the comparison of the model predictions with experimental data is provided. Supplementary results concerning the analytical study of a simplified portion of the model (App. D.2.1) and the numerical sensitivity analysis on kinetic rates values (App. D.2.2) supported the manual tuning of initial parameter values in the estimation procedure and gave insights on the sensitivity of the output as a function of the specific kinetic rates.
Finally, the discussion of the results and the conclusions will be presented in Sec. 5.4.

5.1 Background

An important step in the design process of gene circuits deals with the use of mathematical models to predict the system output; it avoids the use of time- and resource-consuming trial-and-error approaches for the fine tuning of the network [39].
Although the trend of the dynamics of elementary interactions and simple gene networks can sometimes be inferred without the aid of mathematical models, a simple argument is not enough to reveal the nature of the behaviour of more complex reaction networks. In this context, the mathematical descriptions of molecular-level processes and computational analysis, in combination with experimental results, have proved to be fundamental [79].
Mathematical models can highlight deficiencies in our understanding of biological processes and can sometimes suggest what the missing experimental data are or suggest hypotheses that lead to the imperfect functioning of the model [79].
Despite the true nature of biological processes seems to demand the use of stochastic simulations, often the continuous deterministic approach provides a fairly good approximation which does not require burdensome computational efforts [35, 43].
The most suitable model to describe the investigated phenomenon can be difficult to identify and the use of models of different nature (i. e. empirical [31, 80] or mechanistic [81, 35, 43] models) might be necessary to individuate the best candidate [79].
As a general rule, the expression derived from first principles can be applied, unless the biological process is completely unknown. In this case, generic or empirical rate laws can be applied [79].
As an example, the study illustrated in Ch. 4 and published in [31] reports the utilization of an empirical model (based on Hill-like interactions) to explain the formation of a promoter activator complex by the binding of an inducer molecule and a transcription factor. Following this approach, it is not required to know the detailed mechanism of species interaction in order to simulate the molecule kinetics [80]. Here, few parameters, such as the maximum promoter transcription rate or the induction corresponding to the switch-point of the input-output characteristic, condense the knowledge about the system. These parameters do not have a physiological meaning, since they do not represent the interaction rates between the species involved in the system, but can be easily determined experimentally, lessening the burden for experimental kinetic characterization [80].
The point of force of empiric models is their ease of use, which contrasts with the lack of biological significance of the parameters, which are often expressed in arbitrary units and therefore not comparable among different experiments. Furthermore, the use of empiric models does not permit to deepen the knowledge about every single reaction, making explicit each time the parameters of interest [79] and makes it difficult to verify if the hypotheses underlying the

model are satisfied (i. e. : the equilibrium of the fast reactions when the slow reactions start to occur [82]).
To overcome these issues, mechanistic rate laws derived according to the law of mass action kinetics can be properly used [79, 83]. In these models, reaction rates are directly proportional to the activities of the reactants (e.g. concentrations in dilutions and partial pressures in gas phase) to a power, called the order of the reaction for this reactant, which is equal to its stoichiometry [79]. In other words, these models represent each elementary biomolecular interaction with biochemical reactions [35]. Following this approach, a higher number of parameters and intermediary steps are required. In contrast with what happens for the empirical models, here the parameters have a physiological meaning, representing, for example, association and dissociation rates, dimerization and de-dimerization rates or translation rates [81, 35, 43]. Such parameters might be expensive- and difficult-to-measure [84] or even not experimentally detectable, but only detectable through experimental data fitting with onerous computational costs [80].
Mechanistic models have been widely utilized to study the behaviour of gene regulatory networks, where a transcription factor binds or unbinds the operator site of a promoter, thus controlling its activity [85, 86, 82]. Gene network models give an overall view of gene regulation, which is not limited to the study of the interactions between the promoter and the transcription factor, but implicitly captures the protein and metabolite factors that may influence gene expression [82], offering insight into the structure-function relationship [87]. Gene regulatory networks composed by well-studied basic parts, such as the *luxR* and *luxI* genes and the lux promoter from the *quorum sensing* pathway of *V. fischeri* [81], the *tetR* and *tetA* genes and the tet promoter from the *tet* operon of *E. coli* [35], the *lacI* and *lacY* genes and the lac promoter from the *lac* operon of *E. coli* [43] have been evaluated using mechanistic models.
These models have often been implemented by assigning to the parameters values measured with *in vitro* experiments. A very important issue may be legitimately posed, since *in vitro* reactions occur in an environment that is considerably different from the biological context within the cell. For this reason, often a sensitivity analysis on model parameters is performed, to evaluate the entity of the output variation due to non-authentic parameter values [35].
So far, mathematical models have played a fundamental role in quantifying how much the tuning of specific interactions affects network outcome, identifying key structural parameters for a given response and relating dynamics to network topology and function [87].
An important parameter dramatically affecting the system behaviour is the DNA copy number of transcription factor and promoter [88, 68]. Mileyko et al. [88] have developed a quantitative approach for the assessment of the effect of copy number variation on gene expression in small networks. The study demonstrated that small changes in the copy number of the whole gene circuit can deeply change the network quantitative behavior.

Copy number variation can thus be regarded as a 'control knob' within a non-linear dynamical system.
In this study, a simple gene network composed by a LuxR expression cassette and a lux promoter has been investigated. A mechanistic model has been implemented, in order to assess the effect of the variation of the copy number of transcription factor and/or of the promoter on the system ouptut. Furthermore, sensitivity analysis on model parameters has been implemented, to validate the robustness of the system output referred to the impropriety of *in vitro* estimated parameters to explain *in vivo* observed phenomena.

5.2 Genetic network of lux system

The activity of lux promoter has been described by an Ordinary Differential Equation (ODE) model, derived from the following system, based on the mass action rate law.

$$X + H \underset{k_{-1}}{\overset{k_{+1}}{\longleftrightarrow}} D \tag{5.1}$$

$$2D \underset{k_{-2}}{\overset{k_{+2}}{\longleftrightarrow}} C \tag{5.2}$$

$$P_F + C \underset{k_{-3}}{\overset{k_{+3}}{\longleftrightarrow}} P_C \tag{5.3}$$

$$P_C \xrightarrow{k_{+4}} P_C + M \tag{5.4}$$

$$P_F \xrightarrow{k_{+5}} P_F + M \tag{5.5}$$

X represents the amount of LuxR transcription factor present in the cell. The autoinducer H ($3OC_6$-HSL) is supposed to be freely diffusible and its intracellular and extracellular concentrations are supposed to be identical. X and H binding generates a LuxR-HSL complex, named D (Eq. 5.1), which can dimerize and form the active complex C (Eq. 5.2). Both binding reactions are considered here as reversible reactions [89] although in some species the formed complexes are exceptionally stable [81]. C complex specifically binds a target DNA sequence located in the lux promoter, thus triggering its activation (P_F is the free lux promoter, whose transcription is inactive, while P_C is the lux promoter after the C binding and it is transcription is active) as shown in Eq. 5.3. The activated promoter P_C initiates the transcription of the gene of interest, and the generated mRNA is represented by M (Eq. 5.4). Also the inactive promoter P_F has a leaky transcription activity (Eq. 5.5).
These reactions can be represented by an ODE system, which could be studied analytically and simulated numerically. Kinetic parameters used in this model were taken from the literature. The equilibrium constant for the j reaction can be defined as $K_{eq,j} = \frac{k_{+j}}{k_{-j}}$. The basal values used to simulate the system behaviour are reported in Tab. 5.1.

Table 5.1: **Parameters for lux promoter system** Kinetic parameters and initial conditions used to simulate the lux model. When the source is not indicated, the value of the parameter/initial condition has been hypothesized in this study.
[1] This value is the proportionality coefficient relating the synthesis rate $\frac{dM}{dt}$ to P_C. In this study, P_C has been considered as the system output. See the main text for details.

Name	*Description*	*Value*	*Meas. unit*	*Source*
k_{+1}	$X-H$ association rate constant	$6.02{\cdot}10^{3}$	$M^{-1}\ s^{-1}$	[81]
k_{-1}	$X-H$ dissociation rate constant	$3.3{\cdot}10^{-3}$	s^{-1}	[81]
k_{+2}	D complex dimerization rate constant	$6.02{\cdot}10^{3}$	$M^{-1}\ s^{-1}$	[81]
k_{-2}	C complex de-dimerization rate constant	$3.3{\cdot}10^{-3}$	s^{-1}	[81]
k_{+3}	Association constant rate of C to P_F	$6{\cdot}10^{6}$	$M^{-1}\ s^{-1}$	[81]
k_{-3}	Dissociation constant rate of C from P_C	$4{\cdot}10^{-2}$	s^{-1}	[81]
k_{+4}	Transcription rate of P_C when activated by C complex	-	note used[1]	
k_{+5}	Basal transcription rate of P_F	0	-	
X_0	Initial amount of X	$0.5 \cdot 10^{-6}$	M	[90]
H_0	Initial amount of H	$[0\text{-}10^{-5}]$	M	[31, 9]
C_0	Initial amount of C	0	M	-
D_0	Initial amount of D	0	M	-
P_{F0}	Initial amount of P_F	$N \cdot 1.6 \cdot 10^{-9}$	M	-
P_{C0}	Initial amount of P_C	0	M	-

Note that the association and dissociation constants of the first two reactions (Eq. 5.1 and Eq. 5.2) have identical values. The equilibrium constant of Eq. 5.3 has a higher value than the previous ones, suggesting that the binding reaction of C to P_F occurs faster than the process of C formation.
The transcription rate of lux promoter at full induction (k_{+4}) represents the proportional coefficient relating the rate of transcript formation ($\frac{dM}{dt}$) to the active promoter P_C. The leaky activity of lux promoter was supposed to be negligible ($k_{+5} = 0$). So, the final output of the system, which is the transcription rate $\frac{dM}{dt}$, is directly proportional to the amount of bound promoter P_C.
The initial amount of LuxR present (X_0) was set at 0.5 μM, which corresponds to ∼350 molecules per cell, which is a plausible intracellular concentration of the LuxR protein [90]. At $t = 0$ no dimer D was supposed to be present ($D_0 = 0$), nor complex C ($C_0 = 0$). So, all the promoters were in the free state ($P_{C0} = 0$). P_{F_0} was set equal to $N \cdot 1.6n$M, where N is the DNA copy number of lux promoter. This value has been computed as $N/(n_A \cdot V)$, where n_A is the Avogadro number and V is the volume of an *E. coli* cell, assumed to be 10^{-15} liters [91, 35, 81]. H was considered as an exogenous input and the model was simulated for values of H_0 ranging from 0 M to 10 μM [90].
The dynamics of the species involved have been reported in three representative conditions: in absence of inducer (Fig. 5.1 A), at full induction (Fig. 5.1 B) and with an intermediate amount of inducer (Fig. 5.1 C). In all these simulations, N value was set at 40, corresponding to a medium copy number plasmid. If no inducer is given in input (Fig. 5.1 A), no X is consumed to form the dimer D and then the complex C. All the promoters remain in the free state and no transcript is produced. The system is in its trivial equilibrium state. When a huge amount of H (10 μM)[1] is given as input (Fig. 5.1 B), all X is consumed to form the dimer D. This reaction reaches the steady-state in less than 2 minutes. Since H is in excess with respect to X, all X is consumed in this reaction, while H does not significantly deviate from its initial value. In ∼10 minutes, D dimerization to form C complex reaches the steady-state. Smaller amounts of C than its maximal value reached at the steady state are required to occupy all the free promoters P_F, which are totally bound by C in less than 5 minutes. So, the system output (P_C) reaches the steady-state value in less than 5 minutes and its value is equal to the amount of available promoter (P_{F_0}). If an intermediate amount of H ($H_0 = 10$ nM) is given as input (Fig. 5.1 C), less than the available X_0, the D complex formation reaches a steady state value in ∼100 minutes. All the system dynamics are slower, since the limited availability of reactants ($min(X, H) \ll K_{eq,1}, K_{eq,2}$) reduces the probability of a binding event. The processes of D and C formation are driven by the rarely available species, which in this case is H (since $H \ll X$) and all the formed C is immediately bound by the highly available P_F. The process of P_C formation reaches a steady-state in ∼ 300 minutes.
In order to appreciate the system behaviour in its entirety, a static induc-

[1] with respect to X_0 value

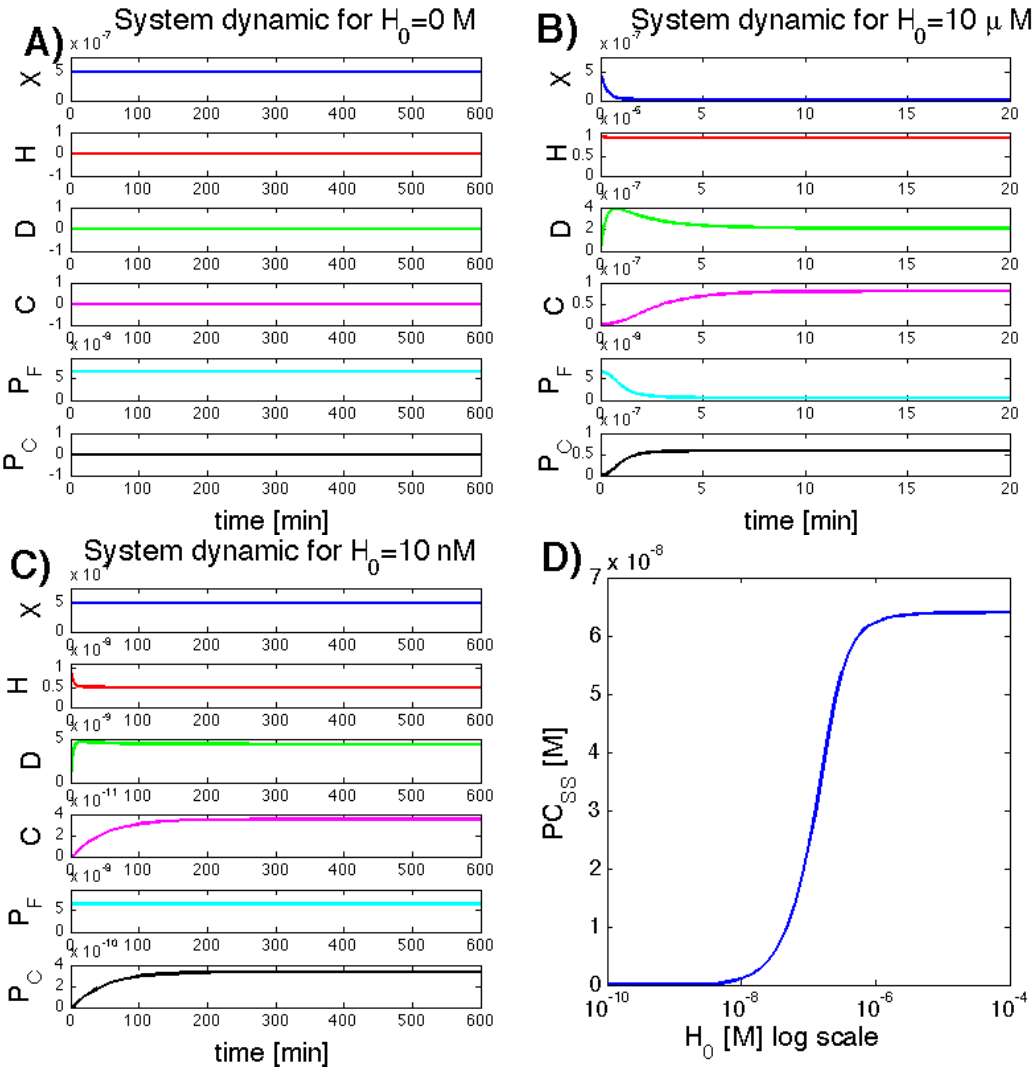


Figure 5.1: **System dynamics and induction curve.** When H_0=0 (panel A), the system does not deviate from the trivial equilibrium state. No dynamic is observed and all the species maintain their initial values. When a huge amount of H_0 is provided in input (panel B), so that $H_0 \gg X_0$, C formation reaches a steady-state value, saturated by the limited availability of X, in ∼10 minutes. C begins to be in excess with respect to P_F since the beginning of its formation process, so P_C formation starts immediately to occur and has a faster dynamic than C formation (driven by k_{+3} and k_{-3} rate constants). The scarce availability of P_F causes a saturation in P_C formation, while C amount is almost not affected. When $H_0 \ll X_0$ (panel C), the processes of D and C formation are driven by the rarely available species, which in this case is H. The free promoter P_F is in excess with respect to the produced C, which saturates the total amount of formed P_C. Almost all the C formed is immediately recruited for P_C formation, while the amount of P_F is almost not affected. Panel D) shows the induction curve.

tion curve has been computed, by considering the steady-state values reached by P_C ($P_{C_{SS}}$) as a function of different H_0 values in the range $[0 - 10^{-5}]$ M (Fig. 5.1 D). The resulting curve has been fitted with a Hill function, with a V_{max} of $6.4 \cdot 10^{-8}$ M (which corresponds to P_{F_0} value), a K_m of $\sim$ 130 nM and η of 1.664. It is important to note that the induction curve obtained with the parameters reported in Tab. 5.1 does not match the curves obtained from experimental data and reported in the literature (see, as an example, Fig. 4.2, where the reported K_m varies from 1.05 to $2.8{\cdot}10^{-9}$ M). Although the output values cannot be directly compared, since they are expressed in different measurement units, a comparison can only be performed on the shape of the curves and the switch point. The K_m obtained from the mass action law model deviates $\sim$60-fold from the K_m estimated by fitting experimental data with a Hill function.

Experimental data from *in vivo* tests are often available as induction curves, measured at the steady state, and they could be theoretically exploited to properly identify model parameters, thus replacing the *in vitro* values used to simulate the system. However, even if the number of reactions involved in this model is limited, mathematical model complexity causes optimization routines for parameter estimation to require appropriate initialization, to ensure convergence to a physiological reasonable minimum. When a kinetic model is as complex as the one considered in this study (see App. D.1.1 for the complete ODE system), optimization routines for parameter estimation require appropriate initialization to ensure convergence to a physiological reasonable minimum. To this aim, an analytical study of a small portion of the model has been performed, in order to determine the analytical relations among the model parameters, species and the system outputs. These results are reported and discussed in App. D.2.1. Unfortunately, even in a simplified context an analytical approach is not feasible for the whole system, since the investigated relations cannot be explicited.

In the studied system, numerical methods need to be applied. To this aim, sensitivity analysis has been performed to pinpoint the parameters of major impact for system quantitative behaviour. The results are discussed in detail in App. D.2.2. Briefly, k_{+1} variation significantly shifts the induction curves without affecting their maximum entity, while k_{+2} has a huge impact on the maximum entity and a reduced influence on curves shift. k_{+3} shows a high influence on both maximum entity and curves shift. This information supported the manual tuning of initial parameter values in the estimation procedure and gave insights on the sensitivity of the output as a function of the specific kinetic rates.

5.3 System behaviour for different copy numbers of promoter and transcription factor

The copy numbers of the promoter and of the transcription factor have been *in silico* varied individually and in concert.

The lux promoter DNA copy number variation has been assessed by tuning the N parameter value, in accordance with the physiological copy number values of commonly used plasmids. LuxR copy number variation has been implemented *in silico* by changing the number of LuxR molecules available to bind the inducer H: the variation of *luxR* expression cassette DNA copy number, in fact, results in a different LuxR availability. So, the variation of the DNA copy number of *luxR* gene or of the produced LuxR molecules by varying the induction of the upstream promoter provide identical results.
The model simulations are here reported and experimental data are presented, to discuss the goodness of the model predictions.
The experimental data relative to the variation of the copy number of the transcription factor only (Sec. 5.3.2) were obtained by changing the *luxR* gene expression level with the aid of an inducible promoter and maintaining the DNA copy number of the whole circuit at 40. So, the LuxR copy number has been tuned *in vivo* by changing the transcriptional strength of the upstream promoter.
Data relative to the variation of copy numbers of both LuxR and lux promoter (Sec. 5.3.3) derive from the experiments described in Ch. 4. In that study the DNA copy number of the whole circuit has been variated.

5.3.1 Copy number variation for lux promoter

The variation of lux promoter copy number has been assessed by assigning to N the desired values, ranging from 1 (a single promoter copy) to 200 (which is a realistic copy number value for some high copy plasmids). The induction curves shown in Fig. 5.2 represent the behaviour of the system in response to these variations. As expected, the P_{max} value is highly affected by copy number variation. In fact, for increasing values of the number of promoters available to drive the expression of the gene of interest, the transcription rate $\frac{dM}{dt}$, in response to the same induction value, will be higher. The relation between P_{max} and N is almost linear, showing a slightly saturating trend for the greatest N values considered (Fig. 5.3 panel B). The H_{50} parameter (see App. D.1.2 for the definition of H_{50} and of the other used statistics) is almost not affected by N variation within the physiological range (Fig. 5.3 panel A), since only a $\sim$2-fold variation is observed (from $1.4 \cdot 10^{-7}$ M to $3.4 \cdot 10^{-7}$ M). This variation entity is in accordance with the data presented in Ch. 4 and discussed in Sec. 5.2, even if the absolute value of H_{50} differs from the experimental data

of more than one order of magnitude. Finally, H* is also stably maintained as a function of N (Fig. 5.3 panel C) and its value corresponds to $\sim$ 1 order of magnitude in H scale (see App. D.1.2 for the explanation of the meaning of this indicator).

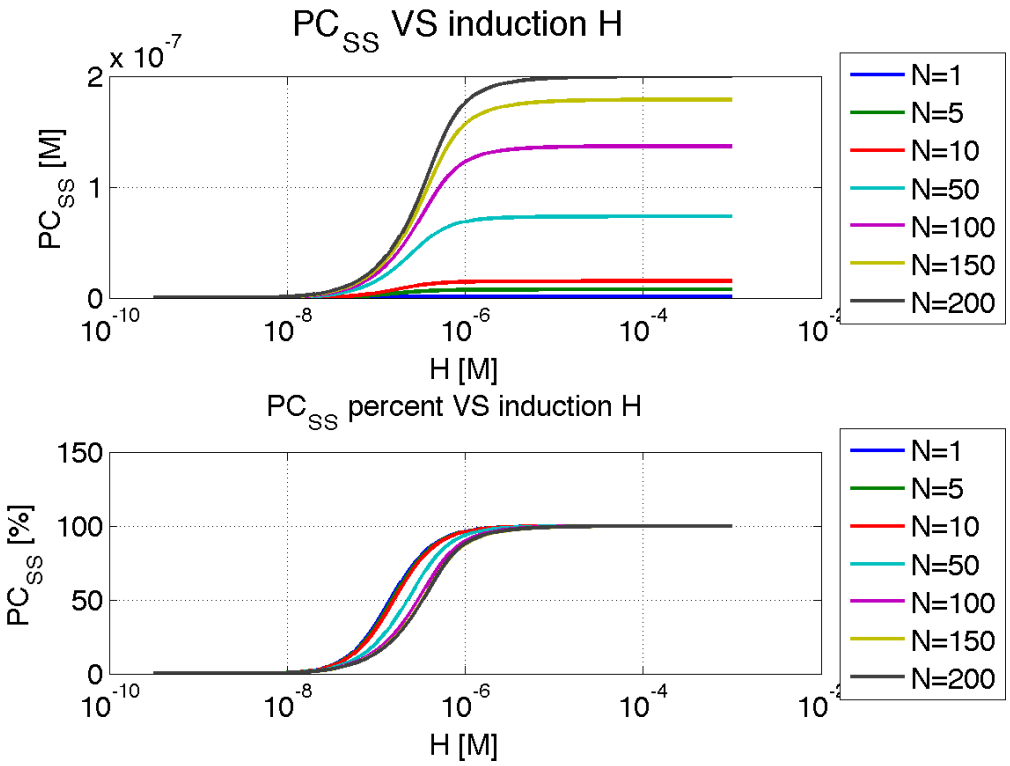


Figure 5.2: **Simulated induction curves for different DNA copy numbers of lux promoter.** The copy number variation for the lux promoter has been achieved by tuning the N parameter within a physiological range (from 1 to 200). For the lowest copy numbers, the induction curves present a lower maximum entity, which increases for increasing values of N. No important differences, in terms of transition range from low to high $P_{C_{SS}}$ values are observable from the curves. In the graph above the induction curves amplitude is in absolute units, while the graph below shows the normalized curves with respect to their maximum. The X-axis of both graphs is in log scale.

5.3.2 Copy number variation for luxR transcription factor

The variation of transcription factor copy number has been modeled as the variation of X available for the binding with H. The nominal X_0 value corresponds to ~350 molecules per cell. So, a 3 order of magnitude variation implies that the number of X molecules changes from less than 1 per cell to more than 300'000 per cell. The promoter copy number N was set at 40,

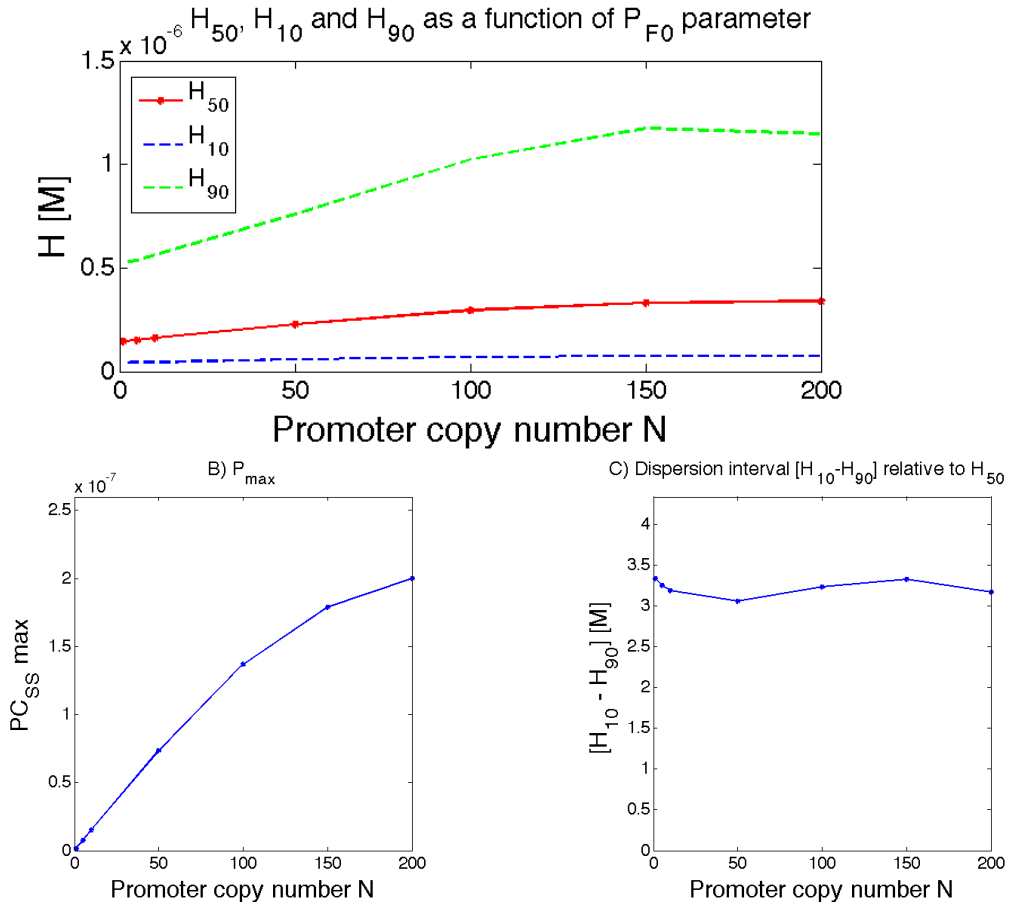


Figure 5.3: **Quantitative parameters to assess the effect of promoter copy number variation on system output.** In panel (A) the H_{50} behaviour against N variation is shown. Its value is maintained almost stably for all the N values. Also H_{10} and H_{90} do not undergo important changes (2-fold variation). Both X- and Y- axes are represented in natural scale. Panel (B) shows the P_{max} behaviour, which is almost linear for increasing N values, showing a slight saturation for the highest values. H* value (panel (C)) is almost not affected by N variation.

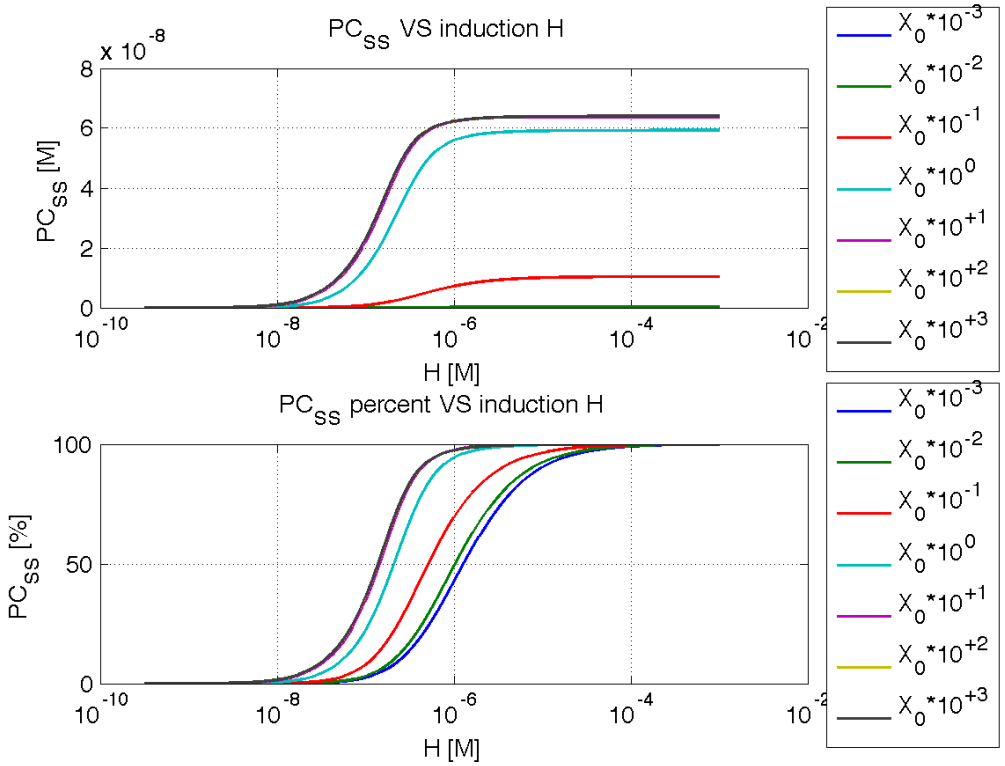


Figure 5.4: **Simulated induction curves for different copy numbers of X transcription factor.** The induction curves reveal how sensitive the system output is in response to X variation. Both maximum entity and curve shift are significantly influenced by X: the former increases for increasing X values, the latter reveals that a saturating trend is present both for the highest and the lowest values. In the graph above the induction curves amplitude is in absolute units, while the graph below shows the normalized curves with respect to their maximum. The X-axis of both graphs is in log scale.

corresponding to a medium copy number plasmid condition. The resulting induction curves, shown in Fig. 5.4, highlight significant effects on system output against X_0 variation. P_{max} strongly varies and the curves are shifted to the left. In particular, H_{50} (Fig. 5.5 panel A) decreases from $\sim 10^{-6}$ to $\sim 10^{-7}$ M with a reverse sigmoidal shape. P_{max} (Fig. 5.5 panel B) changes more than $100,000$-fold, with a sigmoidal shape. This important variation highlights how crucial *luxR* copy number tuning is in designing gene networks, since variations of this parameter dramatically affect the whole system behaviour. H^* halves for increasing values of X copy number (Fig. 5.5 panel C) and most of the variation occurs by varying the X_0 value of only one order of magnitude.

Experimental data obtained by varying the LuxR production and maintaining lux promoter copy number unaltered have been obtained by testing IPTG-HSL$_{RFP}$ device (see App. D.1.3). The resulting induction curves are

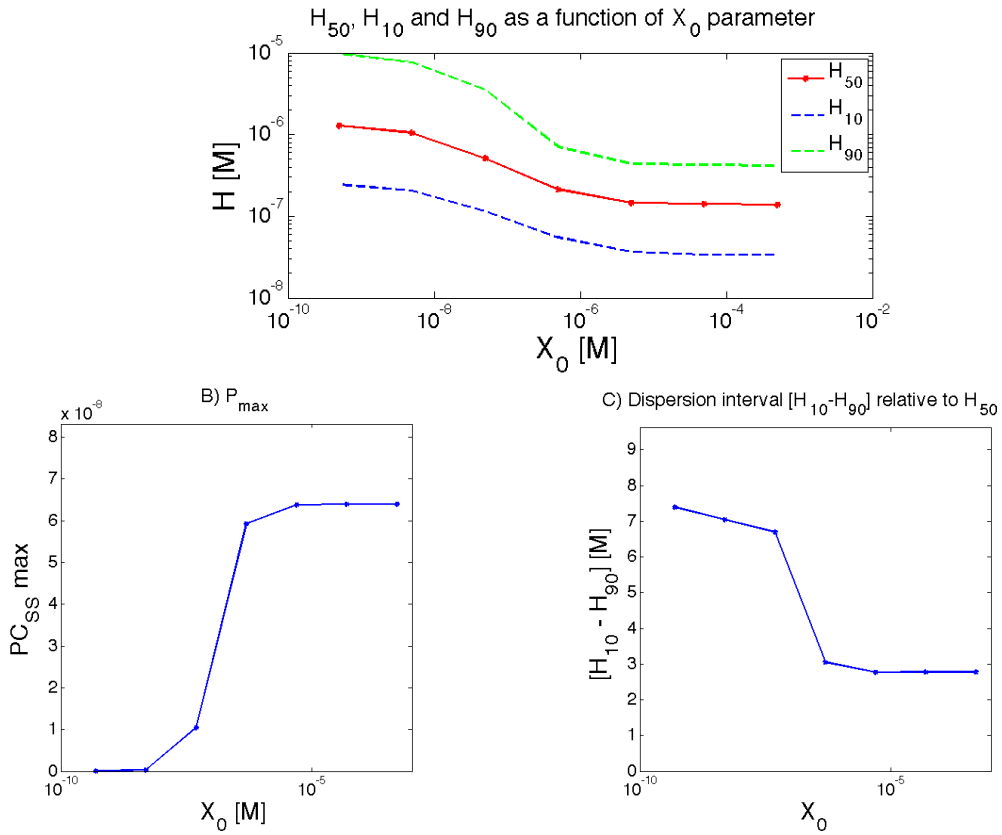


Figure 5.5: **Quantitative parameters to assess the effect of transcription factor copy number variation on system output.** In panel (A) the H_{50} behaviour against X_0 variation is shown. This value strongly depends upon X variation, with a maximum shift of one order of magnitude. Saturating trends are observed both for the lowest and the highest X values. Both X- and Y- axes are represented in log scale. Panel (B) shows the P_{max} behaviour, which has an increasing monotonic behaviour for increasing X values. H* value (panel (C)) decreases for increased X values.

shown in Fig. 5.6. The maximum entity of the induction curves increases for increasing IPTG inductions (corresponding to an augmented LuxR production) from 3.05 to 6.22 RPU. The curves also result > 150-fold shifted to the left on the H_0 axis (K_m varies from 1.66 to 247.09 nM). These results are in accordance with the model predictions, which show both the maximum entity variation and the left-shift (Fig. 5.4). Nevertheless, these predictions are not quantitatively in agreement with experimental data, both for the amplitude fold-change and for the shift entity, which is ∼2 orders of magnitude in experimental data, while the model predicts a variation of ∼1 order of magnitude, only.

Data fitting did not individuate physiological sets of parameters able to quantitatively predict the system behaviour. This may be due to different causes. First, the LuxR variation among different IPTG inductions is unknown. Further experiments for the quantitative assessment of LuxR amount are needed

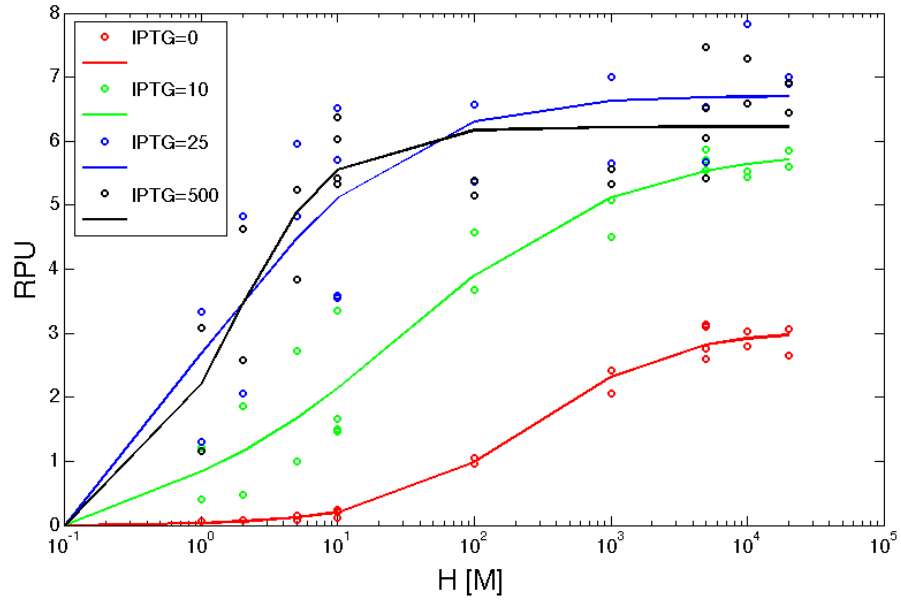


Figure 5.6: **Experimental data obtained by LuxR amount variation.** IPTG-HSL$_{RFP}$ device has been tested as described in App. D.1.3 and four induction curves, corresponding to four different IPTG inductions, have been obtained. The experimental data (dots) have been fitted with a Hill function (line) as explained in App. C.1.6 and the estimated parameters are: (i) IPTG=0; V_{max} =3.05 RPU (CV 3.3%), K_m =247.09 nM (CV 21%), η =0.82 (CV 11%), (ii) IPTG=10 μM; V_{max} =5.88 RPU (CV 6%), K_m =28.4 nM (CV 46%), η =0.53 (CV 20%), (iii) IPTG=25μM; V_{max} =6.7 RPU (CV 5%), K_m =1.8 nM (CV 40%), η =0.68 (CV 36%), (iv) IPTG=500μM; V_{max} =6.22 RPU (CV 4%), K_m =1.66 nM (CV 22%), η =1.18 (CV 27%)

in order to validate the model. Secondly, the uncertainty on kinetic parameter values dramatically affects the estimation process and biologically significant restraints could contribute to overcome this issue. Third, nonlinear phenomena, such as saturation effects, could be present *in vivo* and may explain experimental data better if included in this model, so further biological investigations should be performed to improve the knowledge of the system.

5.3.3 Copy number variation for both lux promoter and luxR transcription factor

The copy numbers of promoter and transcription factor have also been varied in concert. In this case, both P_{F_0} and X_0 have been multiplied by N. This analysis reproduces the experimental conditions described in Ch. 4 and the same data sets presented in Sec. 4.2.2 have been used for the comparison with model predictions and for the parameter estimation. The induction curves are reported in Fig. 5.7. The curves amplitude linearly increases with N, while the curve shift is not monotonic anymore. In fact, for $N < 5$ the curves result

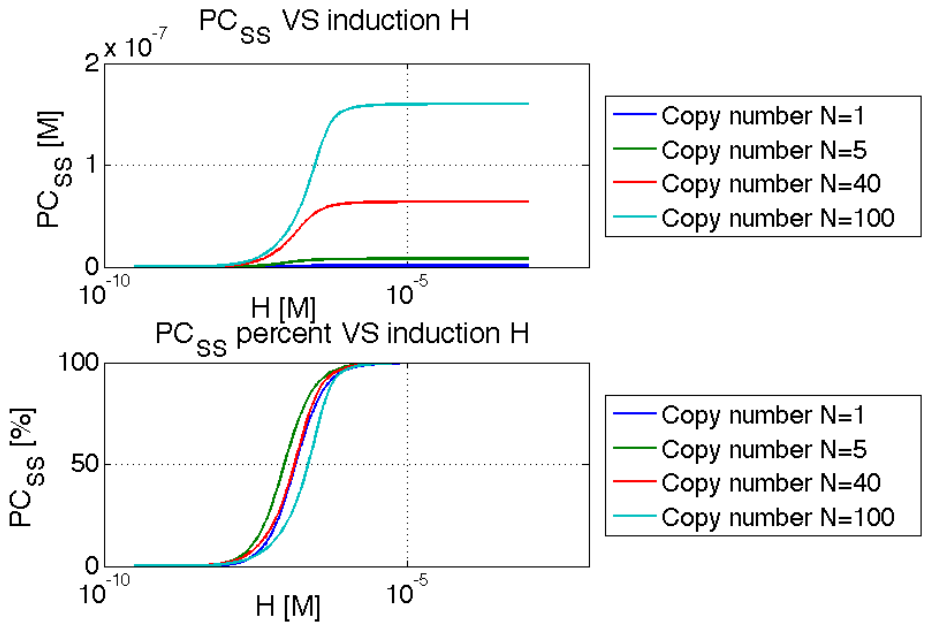


Figure 5.7: **Simulated induction curves for different copy numbers of lux promoter and X transcription factor.** Here, the copy number of both lux promoter and LuxR transcription factor has been varied in concert within a biologically realistic range (1-100). The curves amplitude is proportional to N, while no monotonic reaction subsists between the curves shift and N. In fact, for low N values a left-shift can be observed, while a right-shift is observable for greater N values. In the graph above the induction curves amplitude is in absolute units, while the graph below shows the normalized curves with respect to their maximum. The X-axis of both graphs is in log scale.

shifted to the left, while for $N > 5$ the curves are shifted to the right. This phenomenon is well described by the H_{50} behaviour reported in Fig. 5.8 panel A, in which the two different trends are observable. The P_{max} values linearly increases with the copy number N, while H* slightly decreases for increasing N (Fig. 5.8 panel C). This variation is lower than 1.5-fold. Thus, N variation of both promoter and transcription factor provokes a significant modification ($>$ 100-fold excursion) in the amplitude of the induction curves but does not significantly alter the operative induction range.

The model predictions in response to the variation of the copy number of the whole circuit are qualitatively in accordance with the experimental data presented in Fig. 4.2, Sec. 4.2.2. In fact, the linear relation occurring between P_{max} and N is confirmed in experimental data and in both cases the curves shift is of low entity. As already discussed, the model simulations are not quantitatively in agreement with experimental data, since the predicted switch points diverge of $\sim$ two orders of magnitude from the experimentally measured values.
A physiological set of parameter values able to reproduce the experimental data

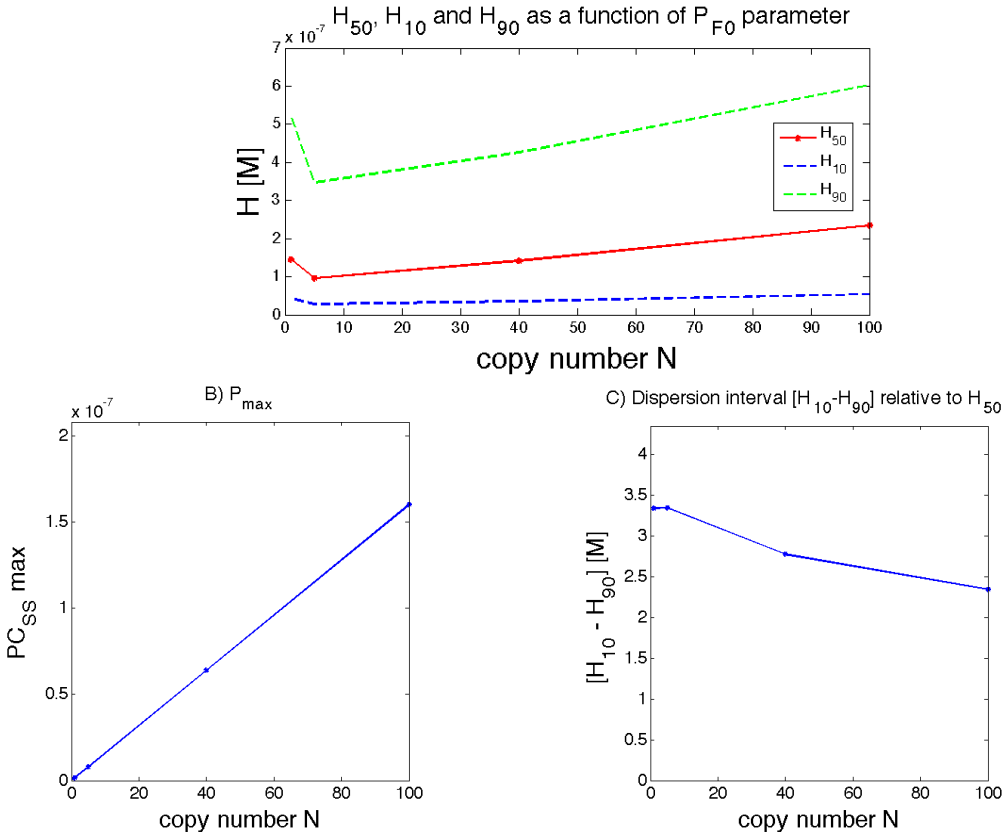


Figure 5.8: **Quantitative parameters to assess the effect of promoter and transcription factor copy number variation on system output.** In panel (A) the H_{50} behaviour against N variation is shown. Despite the variation entity of H_{50} among the different N values is limited ($<$ 2.5-fold), a non monotonic behaviour is observed. This behaviour is observed when the amount of complex C available to bind the promoter P_F is insufficient. When C (that depends on X availability) is scarce, a left-shift is observed. In all the other contexts, a right-shift occurs. Both X- and Y- axes are represented in log scale. Panel (B) shows the P_{max} behaviour, which is linear for increasing X values. H* (panel (C)) decreases for increased X values.

is reported in Tab. 5.2. These values, although not statistically affordable (as suggested by the very high CV values), seem to suggest that the kinetic rates of the first two reactions (Eq. 5.1 and Eq. 5.1) should assume different values, with the second reaction occurring faster than the first. In particular, $K_{eq,1}$ value is confirmed to be $\sim 10^6 M^{-1}$, while $K_{eq,2}$ value dramatically changes, becoming $\sim 6 \cdot 10^{14} M^{-1}$. The binding of C complex to P_F promoter has an equilibrium constant $K_{eq,3} \sim 10^8 M^{-1}$, comparable with the one derived from the nominal values (Tab. 5.1) . The most unexpected predicted value is X_0 ($1.1 \cdot 10^{-10}$), which seems to be in a situation of total scarcity with respect to the free promoter P_F. This scenario is in opposition with the one described by the nominal values, for which LuxR seemed to be in excess with respect to the promoter. The uncertainty on LuxR amount is also motivated by the occurrence of saturation phenomena (as described in Ch. 4) on both LuxR

production and lux promoter activity.

The resulting simulated curves are reported in Fig. 5.9.

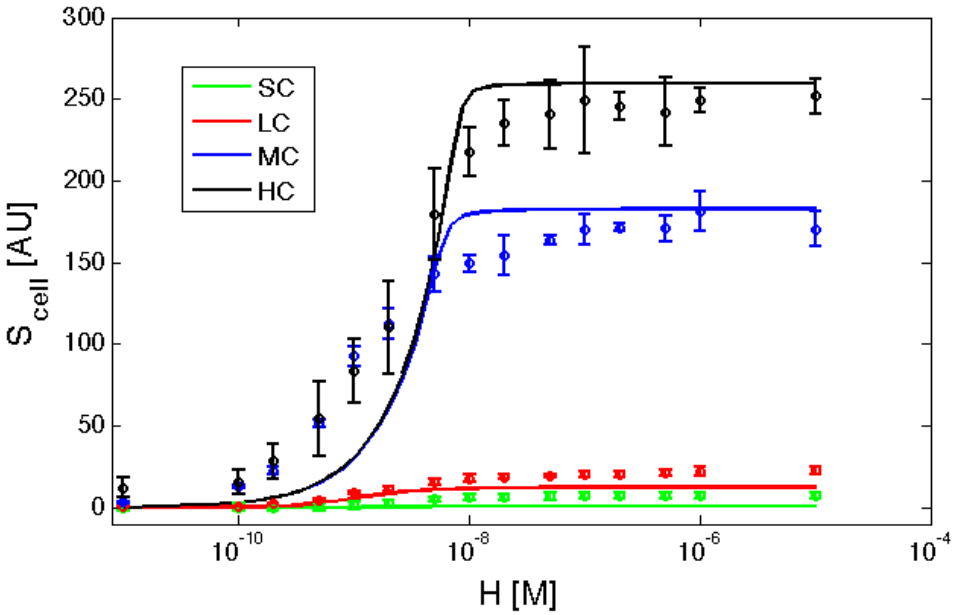


Figure 5.9: **Induction curves for the variation of the copy number of the whole system.** Dots are the average of triplicate measurements, error bars represent the 95% confidence interval. Solid lines represent the fitted curves, with the parameters reported in Tab. 5.2.

5.4 Discussion and conclusions

In Ch. 4 an empiric model based on Hill functions has been proposed to study the effects of DNA copy number variation on a simple gene network, composed by a luxR expression cassette and the lux inducible promoter. This model offers a set of parameters which outline the main features of the system. However, by its nature, such empirical model is not suitable to study the effects of physiological parameters (such as DNA copy number) variation, since they are not explicated in the model equations.

To overcome this issue, the lux system has been investigated with the aid of a mechanistic mathematical model based on the mass action law. This model gives the possibility to tune the system output as a function of copy number and kinetic rates variation. Furthermore, the physiological boundaries of parameter values and the validity conditions for the assumptions of empiric models derived by the simplifications of complex systems can be fixed out.

The mechanistic model has been simulated using parameter values from the literature, estimated via *in vitro* experiments and its quantitative behaviour has been illustrated.

Table 5.2: **Estimated parameters for lux system** Kinetic parameters and initial conditions estimated from experimental data.

[a] parameter estimated from experimental data

[b] assumed values

[c] system input.

Values in brackets are the parameters coefficients of variation.

Name	*Value*	*Meas. unit*	*Source*
k_{+1}	$6.01{\cdot}10^{7}$ ($> 100\%$)	$M^{-1}\ s^{-1}$	[a]
k_{-1}	49.36 ($> 100\%$)	s^{-1}	[a]
k_{+2}	$6.02{\cdot}10^{9}$ (8.3%)	$M^{-1}\ s^{-1}$	[a]
k_{-2}	$1.12\ {\cdot}10^{-5}$ ($> 100\%$)	s^{-1}	[a]
k_{+3}	$6{\cdot}10^{5}$ ($> 100\%$)	$M^{-1}\ s^{-1}$	[a]
k_{-3}	$2.5{\cdot}10^{-3}$ ($> 100\%$)	s^{-1}	[a]
k_{+4}	$6.47{\cdot}10^{10}$ ($> 100\%$)	$AU \cdot M^{-1}\ s^{-1}$	[a]
k_{+5}	0	$AU \cdot M^{-1}\ s^{-1}$	[b]
X_0	$1.1{\cdot}10^{-10}$ ($> 100\%$)	M	[a]
H_0	$[10^{-10} - 10^{-5}]$	M	[c]
C_0	0	M	[b]
D_0	0	M	[b]
P_{F0}	$N \cdot 1.6 \cdot 10^{-9}$	M	[b]
P_{C0}	0	M	[b]

A simplified portion of this model has been analytically studied, but the complexity of the equations did not allow to scale up to the complete system. In order to overcome the issues concerning the fact that parameters can assume different values for reactions that occur in the cell, a sensitivity analysis has been performed and the parameters of major impact on system quantitative behaviour have been individuated.
A study of the behaviour of the system for different copy number conditions has then been performed. First, the *in silico* simulation of the output as a function of the variation of promoter copy number has been provided. Then, the LuxR copy number has been varied and model simulations have been compared with experimental data measured *in vivo* in a genetic circuit where the copy number of lux promoter was constant and the *luxR* expression was varied. The qualitative behaviours of data and simulations are in accordance, despite the absolute quantities do not correspond. Furthermore, no physiological parameter data set has been able to explain the experimental data. The problem of data initialization has a huge impact and additional experiments should be performed, to reduce the uncertainty about the values of the parameters and species.
Finally, the copy number of promoter and transcription factor has been varied in concert, thus simulating the model systems characterized in Ch. 4. The model simulations are totally in accordance with the experimental data and the empiric model has revealed a good approximation of the complex system in the studied conditions (i.e.: the switch point variation is negligible). Also in this case, the absolute quantities do not correspond, but a physiological set of parameters has been individuated, able to reproduce the experimentally observed output. These parameters differ from the *in vitro* measured ones, but the most unexpected result deals with the predicted amount of initial LuxR, which totally differs from the initial hypothesis. In fact, experimental data seem to suggest that the lux promoter is present in excess with respect to the transcription factor and it contrasts which the initially assumed values, for which LuxR was ∼300-fold more concentrated than free promoter. Further experiments for the quantification of LuxR amount as a function of DNA copy number or expression rate variation are required in order to define what are the biologically consistent initial conditions of lux system. Saturation phenomena occurring in medium and high copy number contexts contribute to make this amount difficult to estimate.
Nevertheless, the mechanistic model has proved to be a powerful tool to aid the design and fine tuning of gene network, since the predicted effects on copy number variation are in total accordance with the experimental data. Further efforts, both experimental and computational, are required to improve the model and refine parameter values, in order to obtain quantitatively correct predictions.

Chapter 6

Overall conclusions

The bottom-up programming of living organisms to implement novel, user-defined functions is the main goal of Synthetic Biology. One of the main problems connected with the assembly of genetic circuits from simple genetic parts is the unpredictability of the behaviour of the whole system, given the input/output behaviour of the single modules that compose it. Copy number, transcriptional /translational demand and toxicity of the DNA-encoded functions are some of the major factors which may lead to cell overburdening and thus to nonlinear effects on system output. It is essential to trace the linearity working boundaries of engineered biological systems when dealing with such phenomena. In this thesis work, the nonlinear and saturation effects due to copy number variation have been examined using *in vivo* and *in silico* methodologies.
In order to contextualize the problem of system output variation as a function of DNA copy number, two case studies are presented in Ch. 2. First, a set of five constitutive transcriptional promoters has been characterized in two different copy number conditions. By comparing the activity of promoters characterized via the same reporter device in low copy and high copy vectors, it resulted that the weakest promoters had comparable RPUs (with a variation <1,3-fold), while the two strongest promoters had reduced RPUs in high copy vector relative to low copy vector (with a 2,3- and 4,4-fold variation). This variation could be due to saturation effects in transcription/translation processes that occur for the strongest promoters in high copy condition, while such effects were absent for the other (weaker) promoters.
Secondly, a bacterial inducible self-destruction device has been characterized in *E. coli*. This device is composed by a module for the inducible expression of t and e genes (coding for a holin and a lysozyme, respectively) and a module for the weak constitutive expression of rI gene (encoding for an anti-holin). This device had a highly stable and predictable behaviour when borne on a low copy plasmid, resulting in cell lysis in every bacterial growth phase (lysis entity was 76,28% ± 0,3% in early exponential phase, 75,43% ± 1,1% in late exponential phase and 50,1% ± 2,5% in stationary phase) and without signifi-

cantly affecting cell growth (the doubling time of cells bearing the lysis device in low copy was 49,8 min, while the negative control had a doubling time of 43 minutes). High copy number plasmid gave worse performance, since lysis often failed to occur and, when the device worked, lysis entity was much reduced (58,6% ± 7,5% in early exponential phase, 42,2% ± 9,9% in late exponential phase and 0% in stationary phase) and cell growth was slowed (doubling time of 69 minutes for the cells bearing lysis device in high copy and of 45 minutes for the negative control) when compared to low copy system.
These results evidenced that copy number tuning is crucial both when dealing with very simple circuits, such as the transcriptional promoters driving the expression of a reporter protein, and when characterizing circuits encoding for toxic functions or with high metabolic demand. Therefore, in the design and fine tuning process of gene networks it is necessary to investigate the nonlinear effects on gene expression due to copy number.
A deep analysis on copy number related phenomena has been carried out in this thesis work and biological and mathematical tools have been developed to construct and study devices in different copy number conditions. Such kind of tools and studies can support the design process of biological functions, as copy number effects have a major impact in the quantitative behavior of engineered biological systems. To this aim, in Ch. 3 the design, construction and validation of a standard BioBrick™ compatible integrative vector for *E. coli* are described. This vector can be specialized to target the desired genome *locus* using BioBrick™ parts. In its default version, it is suitable for the integration in the ϕ80 attP site via site specific recombination (experimental data show a high success rate), but it can also be modified to target a different genome *locus* via both site specific and homologous recombination. In this study, the homologous recombination in the AspA *locus* has been validated. *Ad-hoc* experiments to test the evolutionary stability of integrated devices respect to plasmid-borne solutions revealed that the integrated parts could be stably maintained in the chromosome for at least 150 generations of continual cell growth without the need of antibiotic selection, thus confirming that this vector is a powerful tool for the construction of gene circuits stably integrated in *E. coli* chromosome in a single DNA copy.
This vector has been utilized in this thesis work to construct biological systems with genetic circuits in single copy, which have been used to investigate copy number effects. The set of promoters previously characterized in high and low copy vectors has been characterized in single copy, integrated in the ϕ80 attP and AspA *loci*. The RPU values were consistent for the three tested conditions (single copy in ϕ80, single copy in AspA and low copy vector) and the greatest RPU variation yielded a CV of 31%.
In Ch. 4 an HSL-inducible device, composed by a LuxR transcription factor expression cassette and an mRFP expression cassette driven by the HSL-inducible lux promoter, has been characterized in four different copy number contexts. The system output proved to be highly affected by copy number

variation. Maximum activity expressed in RPUs was more than 3-fold reduced from single to high copy context, maybe for saturation phenomena of the transcriptional/translational machinery. The use of an empirical mathematical model allowed the estimation of a difficult-to-measure intracellular species, which is the LuxR-HSL activated complex. The induction curves as a function of the activated complex, which is the real system activator, allowed correct comparisons between the different tested conditions. From these data, the dependence of the switch point on the copy number was evinced.
Even for this simple system, the utilization of mathematical models has been necessary to correctly relate the system output to the copy number variation. Nevertheless, because of its empiric nature, the described model presents some drawbacks when used to predict these effects.
To overcome these issues, in Ch. 5 a mechanistic model based on the mass action law of the HSL-inducible device was defined. This model allows the study of the effects of copy number variation on system output. The model parameters were retrieved from the literature. Since they have been estimated via *in vitro* experiments, they may be not the best parameter set to describe *in vivo* occurring processes: numerical sensitivity analysis has then been implemented to measure the system robustness to parameter values uncertainties. Furthermore, a simple portion of the model was analytically studied in order to express the output dependency on copy number and on the parameters. Unfortunately, the analytical solution is not practicable because of the complexity of the equations.
The copy numbers of the promoter and the transcription factor were variated *in silico*, both individually and in together, and numerical results were compared with experimental data. The model previsions in response to copy number variation were qualitatively in accordance with experimental data, despite the absolute predicted quantities were deviated from the experimental values of one order of magnitude. Further computational and experimental efforts are required to fill the gap between the biological knowledge of the system and the measurement of kinetic rates regulating the system.
Taken together, these results show how fundamental the copy number tuning is in regulating gene networks. Biological and computational tools to study this phenomena have been developed and tested and an important step has been done in the process of defining the linear working boundaries of synthetic biological devices.

Appendix A

Methods and supplementary information for Ch. 2

A.1 Methods and supplementary information for the modularity study on basic parts and devices

A.1.1 Plasmid construction and cloning

BioBrick™ Standard Assembly (Sec. 1.3.1) was used to construct all the plasmids of this study, following a number of conventional molecular biology techniques. As a result, all the DNA junctions between parts had the TACTAGAG sequence (or TACTAG when the downstream part was a coding sequence) [11]. DNA-modifying enzymes were purchased from Roche. DNA purification kits were purchased from Macherey-Nagel. Chemically competent TOP10 (Invitrogen) were cultivated in LB medium and were used as hosts for plasmid propagation. Ampicillin (100 mg/l), Chloramphenicol (12.5 mg/l) or Kanamycin (50 mg/l) were added as required. TOP10 strain was also used as chassis for all quantitative experiments. Enzymes, purification kits and competent cells were used according to manufacturer's instructions. All the plasmids realized in this study were assembled from basic or composite parts from the Registry 2009 or 2010 DNA Distribution [20]. In Tab. A.1 the basic BioBrick™ parts used to design the characterized systems are reported. If not differently stated, all the constructs were tested in the pSB4C5 low copy vector (pSC101 replication origin), maintained in transformants by adding 12.5 mg/l of Chloramphenicol to the growth media or in the pSB1A2 high copy vector (ColE1 replication origin), maintained by adding 100 mg/l of Ampicillin. Long-term stocks, routinely stored at -80°C, were prepared for all the recombinant strains by mixing 250 μl of 80% glycerol with 750 μl of bacterial cells grown in selective LB.

Table A.1: **List of parts used to design the model systems** For each part, the BioBrick™ code is reported and a brief description is provided.

BioBrick™	**Description**
BBa_B0015	double transcriptional terminator
BBa_B0031	weak RBS
BBa_B0032	medium-weak RBS
BBa_B0034	strong RBS
BBa_C0040	TetR repressor coding sequence
BBa_C0062	LuxR activator coding sequence
BBa_E0040	GFPmut3b coding sequence
BBa_E1010	mRFP1 coding sequence
BBa_I14032	P_{lacIQ} constitutive promoter
BBa_J23100	Constitutive promoter family member
BBa_J23101	Standard reference promoter
BBa_J23105	Constitutive promoter family member
BBa_J23106	Constitutive promoter family member
BBa_J23116	Constitutive promoter family member
BBa_J23118	Constitutive promoter family member
BBa_R0011	P_{LlacO1} synthetic promoter
BBa_R0040	P_{tetR} synthetic promoter
BBa_R0051	P_R promoter from lambda phage
BBa_R0062	P_{LuxR} promoter from *Vibrio fischeri*

A.1.2 Promoter characterization

Recombinant bacteria from a long-term glycerol stock were streaked on an LB agar plate and grown for about 20 hours at 37°C. 1 ml of selective M9 supplemented medium (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 2 mM MgSO4, 0.1 mM CaCl2, 0.4% glycerol) was inoculated with a single colony from the streaked plate and incubated at 37°C, 220 rpm shaking for about 20 hours. A 1:500 dilution was then performed and cultures were further incubated for 3 hours at 37°C, 220 rpm. For each culture, a 200-μl aliquot was transferred into a flat-bottom 96-well microplate (Greiner) and assayed for about 6 hours in an Infinite F200 microplate reader (Tecan) with a kinetic cycle programmed with the i-control software (Tecan). Fluorescence (Ex:485 nm, Em:540 nm for GFP; Ex:535 nm, Em:620 nm for RFP) and absorbance (600 nm) were measured every 5 minutes. In every measurement cycle, cultures were shaken (linear shaking, 3 mm amplitude) for 15 seconds and then, after a 5-second wait time, the measurements started. Temperature was kept constant at 37°C during all the experiment. The gain of GFP and RFP fluorescence measurements was set at 50 when assaying reporter genes on high copy vectors, while it was set at 80 for low copy assays, in which the fluorescence signal is weaker. The absorbance of sterile M9 medium and the autofluorescence of the strain without fluorescent proteins were measured in order to estimate the absorbance and fluorescence background, respectively. In each experiment, bacteria bearing the standard reference promoter BBa_J23101 driving a reporter gene (here called reference culture) were also assayed, so that strain, plasmid copy number, antibiotic resistance, reporter gene and RBS were exactly the same as in the cultures of interest.

A.1.3 Data analysis

Data were analyzed as described in [10] to obtain RPUs (see Sec. 1.3.2). Raw absorbance and fluorescence time series were processed with the MATLAB 2007b suite (MathWorks, Natick, MA). Briefly, M9 and strain background values were subtracted to all the absorbance and fluorescence raw measurements respectively to obtain values proportional to the per-well cell count and number of fluorescent proteins. The reporter protein synthesis rate per cell time series ($\overline{S}_{cell}$) of each culture was computed as the numeric time derivative of the fluorescence values, divided by absorbance. This time series was averaged in the exponential growth phase for each well, thus yielding S_{cell}. Exponential growth phase was identified by visual inspection as the linear region of the $ln(OD_{600})$ time series and the slope m of this line was computed with linear regression to yield the cell growth rate. Doubling time was computed as $ln(2)/m$. The RPU value of a promoter was computed as $\frac{\overline{S}_{cell,\phi}}{\overline{S}_{cell,ref}}$, where ϕ is the culture bearing the promoter of interest and ref is the reference culture. When computing the coefficients of variation (CVs), the correction for small

samples was implemented: for N samples, $CV_{corrected} = CV \cdot (1 + \frac{1}{4N})$. Hypothesis tests were performed via MATLAB.
To assess the statistical difference among the mean promoter activity values in a group, ANOVA test was performed. If a difference was detected in the group, individual t-tests were performed to compare the mean values of the group members to identify statistically different sub-groups in order to compute the CV. The p-values (P) were corrected for multiple comparisons with the Bonferroni method. The mean values of the non-significantly different sub-groups were averaged and the final CV within the group was computed on the mean values of the statistically different sub-groups. If ANOVA highlighted a statistical difference, but multiple comparisons showed no evident sub-groups, the CV was computed on all the mean activities among the group.

A.2 Methods and supplementary information for the characterization of a synthetic bacterial self-destruction device

A.2.1 Lysis assays

5 μl of bacteria bearing the lysis device glycerol stocks were inoculated in 5 ml of selective LB medium and grown at 37°C, 220 rpm overnight. The cultures were diluted 1:100 in 5 ml of selective LB medium and grown for additional 4-5 hours under the same conditions as before. After that time, a 200 μl aliquot of each culture was transferred in a flat-bottomed 96-well microplate (Greiner) and the OD_{600} was measured with an Infinite F200 microplate reader (Tecan). Based on this measurement, the cultures were diluted to the same OD_{600} (0.05-0.13) and then six 200 μl aliquots of each culture were transferred in a flat-bottomed 96-well microplate (Greiner). Unless otherwise noted, three wells of each culture were induced with 2 μl of properly diluted HSL (Sigma Aldrich #K3007) and 2 μl of deionized water were added to the other three wells (uninduced wells). If the HSL-inducible lysis device had to be assayed with different HSL concentrations in the same experiment, three 200 μl aliquots of the cultures for each investigated concentration were transferred in the microplate and induced. The microplate was incubated at 37°C in the Infinite F200 microplate reader and assayed every 5 min following this protocol immediately before the measurement: 15 s of linear shaking (amplitude = 3 mm), wait for 5 s, OD_{600} measurement.

A.2.2 Analysis of growth curves

Raw OD_{600} values measured in the Infinite F200 microplate reader were normalized by subtracting for each time point the mean raw absorbance of the media to compute the actual bacterial optical density.
The growth phases of bacterial cultures with the different plasmids used in this study, grown in LB medium in a microplate, have been characterized in each experiment by computing the natural logarithm of the OD_{600} values $\ln(OD_{600}(\mathrm{t}))$ over time. Then, the exponential phase was identified by visual inspection as the linear region of $\ln(OD_{600}(\mathrm{t}))$, the late stationary phase as the constant region and the early stationary phase as the region between the other two.
In all the lysis assays, uninduced bacteria doubling time in exponential growth phase was evaluated by performing linear regression over the $\ln(OD_{600}(\mathrm{t}))$ linear region to estimate the curve slope m, which represents the growth rate of the culture. Then the culture doubling time was computed as $ln(2)/m$. The lysis entity after induction was computed as $100 \cdot (1 - min_{OD_{600}}/ref_{OD_{600}})$, where $min_{OD_{600}}$ is the minimum OD_{600} reached by the culture and $ref_{OD_{600}}$ is the OD_{600} immediately before the density drop caused by cell lysis. The rise

time was computed as the time required for the $100 \cdot (1 - OD_{600}(t)/ref_{OD_{600}})$ signal to rise from 10% to 90% of the lysis entity.

A.2.3 Optical density calibration

OD_{600} measurements performed with the Infinite F200 microplate reader and the NanoDrop ND-1000 were converted to OD_{600} measurements in 1 cm pathlength by calibrating the two instruments with the V-530 spectrophotometer (Jasco), measuring the OD_{600} of serial dilutions of a TOP10 culture grown in LB or M9 supplemented medium.

Appendix B

Methods and Supplementary Information for Ch. 3

B.1 Materials and methods

B.1.1 Strains and plasmids

All the *E. coli* strains used in this study are listed in Tab. B.1. MG1655 [92] and MC1061 [93] were used as integration hosts. BW23474 [59] were used to propagate at high copy number the plasmids with the R6K conditional replication origin. TOP10 were heat-shock transformed according to manufacturer's protocol, while the other strains were made chemically competent as described in [94] and were heat-shock transformed at 42°C for 1 min. Tab. B.2 lists the plasmids used for the construction of the integrative vectors and the used helper plasmids. The genetic devices used for single-copy and low-copy characterization of BioBricks™ have been reported previously (see Fig. 3.1, Tab. B.2 and [32, 31]).

Table B.1: **Strains used in this study.**

Strain	Relevant genotype	Source
MG1655	wild-type K-12, no ϕ80 prophage	CGSC (#7740)
MC1061	no ϕ80 prophage	CGSC (#6649)
BW23474	recA1 endA9(del-ins)::FRT uidA4(del)::pir-116	CGSC (#7838)
TOP10	recA1 endA1	Invitrogen

Table B.2: **Plasmids used for integrative vector construction.** The BioBrick™ code of the plasmids include the vector name out of brackets and insert name in brackets.

[a]: obtained from Mr. Gene DNA synthesis service. pMK-RQ is a Kanamycin-resistant high copy number delivery vector from Mr. Gene.

[b]: obtained from DNA Distributions of the MIT Registry of Standard Biological Parts.

[c]: constructed in this study.

[d]: gift from the Dr. J.C. Anderson lab (UC Berkeley).

[e]: obtained from CGSC (plasmid purified from recombinant strain CGSC#7629).

Name	**BioBrick™ code**	**Description**
pHC-attP-CS	pMK-RQ(BBa_K300983)[a]	DNA fragment including FRT-NheI-attP-NheI-CS(B0033)-FRT, flanked by AvrII sites, in high copy vector
pHC-CmR	pSB1A1(BBa_P1004)[b]	Chloramphenicol resistance gene with promoter in high copy vector
pHC-ter	pSB1AK3(BBa_B0015)[b]	Double terminator in high copy vector
pHC-R6K	pSB1A2(BBa_J61001)[b]	R6K conditional replication origin in high copy vector
pHC-CmRter	pSB1AK3(BBa_K300032)[c]	Chloramphenicol resistance cassette in high copy vector
pHC-P_R -RFP	pSB1A2(BBa_I763007)[b]	RFP constitutive expression cassette, driven by the P_R promoter in high copy vector

pHC-CmRter-R6K	pSB1A2 (BBa_K300008)[c]	Chloramphenicol resistance cassette with R6K origin downstream in high copy vector
pBBintϕ-B0033	BBa_K300982[c]	Integrative base vector, targeting the ϕ80 locus, with RBS BBa_B0033 as insert
pBBintϕ-RFP	BBa_K300000 (BBa_I763007)[c]	Integrative base vector, targeting the ϕ80 locus, with RFP constitutive expression cassette as insert
pHC-aspA	pSB2K3 (BBa_C0083)[b]	aspA coding sequence in high copy vector
pBBintAsp-RFP	BBa_J107058 (BBa_I763007)[c]	Integrative vector, specialized to target the aspA *locus*, with RFP constitutive expression cassette as insert
pInt80-649	BBa_J72008[d]	Ampicillin-resistant low copy helper plasmid, with temperature-sensitive pSC101 origin, containing a constitutively expressed pir-116 gene and the site-specific recombination machinery of the ϕ80 phage under the control of a heat-sensitive promoter.
pCP20	not in the Registry [e]	Ampicillin- and Chloramphenicol-resistant low copy helper plasmid, with temperature-sensitive pSC101 origin, containing the yeast Flp recombinase under the control of a heat-sensitive promoter.

B.1.2 Cloning methods

Plasmids were propagated in recombinant strains grown in selective Luria-broth (LB) [94] at 37°C, except plasmids containing a temperature-sensitive replication origin that were grown at 30°C. Antibiotics were routinely used for plasmid maintenance during DNA propagation: Ampicillin (100 mg/l for high copy plasmids, 50 mg/l for low copy plasmids), Chloramphenicol (12.5 mg/l) and Kanamycin (50 mg/l). Liquid cultures (5 ml) were incubated with shaking at 220 rpm. DNA was extracted with the NucleoSpin Plasmid kit (Macherey-Nagel) from recombinant cultures grown to saturation. Purified DNA was digested as required and gel-extracted with the NucleoSpin Extract II kit (Macherey-Nagel). DNA parts were ligated to construct all the plasmids described in this work. 1 μl of ligation mix was transformed into a proper *E. coli* strain, DNA was purified and screened via restriction digest and electrophoresis. Sequencing was performed to validate all the plasmids with primers VF2 (TGCCACCTGACGTCTAAGAA) and VR (ATTACCGCCTTTGAGTGAGC). Long-term bacterial stocks were prepared by mixing 750 μl of a saturated culture with 250 μl of 80% glycerol and stored at -80°C. NheI restriction enzyme is from Fermentas. Antarctic Phosphatase is from New England Biolabs. Platinum Taq DNA polymerase (Invitrogen) or Phusion Hot Start Flex DNA Polymerase (New England Biolabs) were used to perform PCR. All the other DNA-modifying enzymes are from Roche Diagnostics. Enzymes have been used according to manufacturer's protocols. Oligonucleotides were purchased from Primm (San Raffaele Biomedical Science Park, Milan, Italy) or from Sigma Aldrich. Sequencing was performed by BMR Genomics (Padua, Italy) DNA analysis service.

B.1.3 Integrative base vector construction procedure

Fig. B.1 shows the assembly scheme followed to construct the pBBintϕ vector backbone. Briefly, pHC-CmRter was constructed via BioBrick™ Standard Assembly by assembling the insert of pHC-CmR, digested with EcoRI-SpeI, to pHC-ter digested with EcoRI-XbaI. Analogously, pHC-CmRter-R6K was obtained by assembling the insert of pHC-CmRter, digested with EcoRI-SpeI, to pHC-R6K digested with EcoRI-XbaI. The insert of pHC-attP-CS is composed by, in this order, FRT-NheI-attP-NheI-CS(BBa_B0033)-FRT, where CS is a cloning site composed by EcoRI and PstI with four transcriptional terminators flanking them and two standard primer binding sites for VF2 and VR (see Fig. 3.2). In this plasmid, the CS contains the small BioBrick™ RBS BBa_B0033 between EcoRI and PstI. The insert of pHC-attP-CS was excised through digestion with AvrII and it was assembled with the insert of pHC-CmRter-R6K, excised from its vector through digestion with XbaI-SpeI and dephosphorylated, to obtain pBBintϕ-B0033. Finally, pBBintϕ-RFP was obtained by assembling the insert of pHC-P_R -RFP, digested with EcoRI-PstI, to the vector backbone of pBBintϕ-B0033, digested with EcoRI-PstI in order

to cut out and eliminate the B0033 RBS.

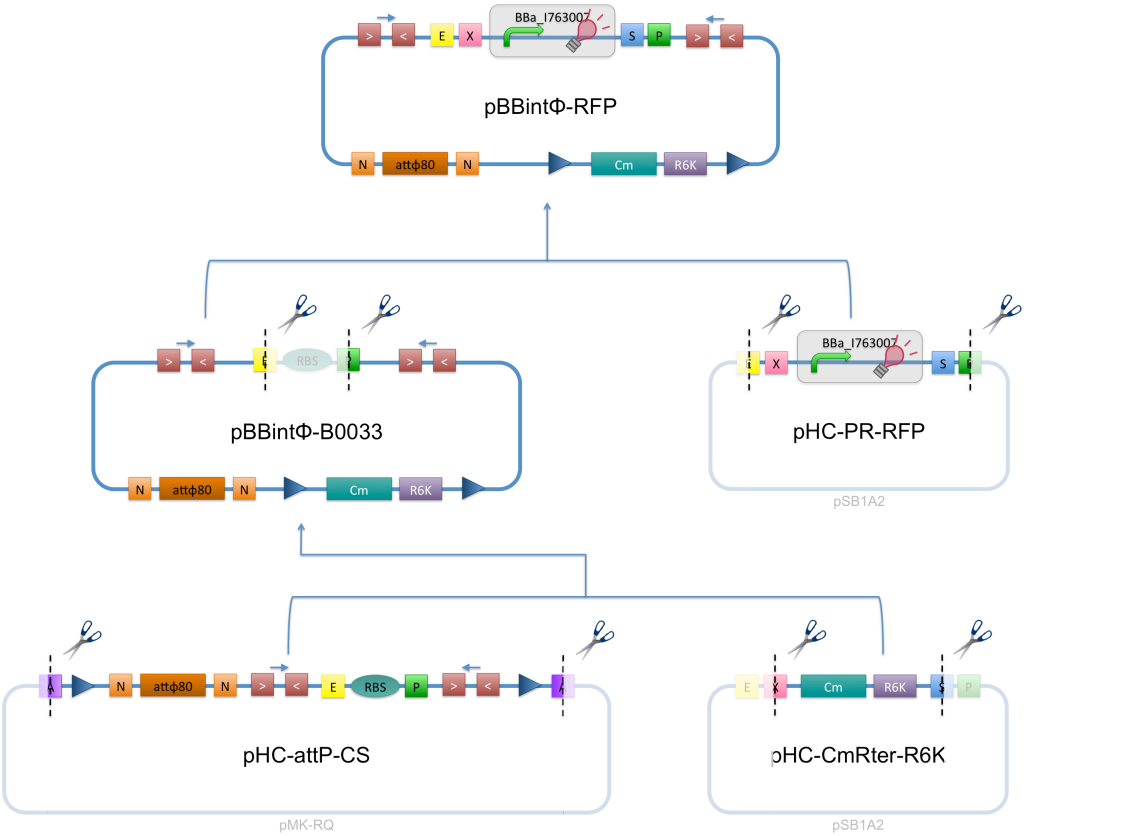
B.1.4 Assembly of the desired passenger and guide

The source or assembly procedure of all the BioBrick™ constructs used as passengers has been reported previously [32, 31]. These constructs have been excised from their original vector backbone upon EcoRI-PstI double digest and they have been cloned into the pBBintϕ or pBBintAsp vector as described in Fig. 3.3 panel A. The aspA coding sequence, used as DNA guide for homologous recombination, was excised from its original vector backbone upon XbaI-SpeI double digest and it has been cloned into the pBBintϕ-RFP plasmid, digested with NheI, as described in Fig. 3.3 panel B, and dephosphorylated.

B.1.5 Chromosomal integration

The protocol described in [23] was used for site-specific recombination. Briefly, competent cells were transformed with the pInt80-649 helper plasmid [23]. Transformants were grown at 30°C and were made competent again. They were transformed with the pBBintϕ integrative vector containing the part of interest as passenger and were grown on Chloramphenicol plates at 30°C until colonies appeared. A single colony was propagated in LB with Chloramphenicol at 37°C, 220 rpm overnight. The culture was streaked on a Chloramphenicol plate and incubated overnight at 43°C. A single colony was grown in LB+Chloramphenicol to yield the integrant strain. This colony was also streaked on an Ampicillin plate to validate the loss of pInt80-649. Colony PCR was occasionally performed with primers P1 (CTGCTTGTGGTGGTGAAT) [59] and P2 (CTCTTACGTGCCCGATCA), which anneal upstream of the ϕ80 chromosomal attB site and in the R6K origin respectively. The P1-P2 amplicon (452 bp) indicates the correct integration position, as described

Figure B.1 *(facing page)*: **Assembly scheme of the integrative base vector.** The Chloramphenicol resistance gene, including its promoter, was assembled to a transcriptional terminator and then the resulting part was assembled to the R6K conditional replication origin. All of them were existing parts from the Registry of Standard Biological Parts. The resulting CmRter-R6K sequence was digested with XbaI and SpeI, it was dephosphorylated to prevent self-ligation and it was assembled to a *de-novo* synthesized sequence (insert of pHC-attP-CS) composed by (in this order): AvrII restriction site, FRT recombination site, NheI restriction site, ϕ80 attP, NheI restriction site, forward transcriptional terminator (BBa_B0053), VF2 primer binding site, reverse transcriptional terminator (BBa_B0055), EcoRI restriction site, BioBrick™ RBS BBa_B0033, PstI restriction site, forward transcriptional terminator (BBa_B0054), VR primer binding site, reverse transcriptional terminator (BBa_B0062), FRT recombination site, AvrII restriction site. AvrII, XbaI and SpeI all have compatible sticky ends. This assembly is non-directional, but the integrative vector can work with the parts ligated in both directions. The ligation orientation of the selected clone was identified by EcoRI-HindIII (Roche) digestion screening. Finally, the BioBrick™ BBa_I763007 was ligated in place of the RBS, thus generating a standard BioBrick™ cloning site containing a default insert.



Part number	Function	Notation
BBa_G00000	BioBrick cloning site prefix	
BBa_G00001	BioBrick cloning site suffix	
BBa_I763007	PR-mRFP1 expression cassette	
BBa_J72001	FRT recombination site	
BBa_B0053 BBa_B0054	forward transcriptional terminators	
BBa_B0055 BBa_B0062	reverse transcriptional terminators	
BBa_G00100	forward sequencing primer annealing site (VF2)	
BBa_G00102	forward sequencing primer annealing site (VR)	
BBa_B0045	NheI restriction site	
BBa_K300999	Cloramphenicol resistance marker (reverse orientation)	
BBa_K300998	R6K conditional replication origin	
BBa_B0048	AvrII restriction site	
BBa_B0033	B0033 RBS	
BBa_P1004	Cloramphenicol resistance cassette without terminator	
BBa_B0015	Double transcriptional terminator	
BBa_M31983	EcoRI restriction site	
BBa_K300981	PstI restriction site	

in [59], while negative clones show no amplicon. It is also possible to identify clones with multiple tandem copies of the integrated part via colony PCR with primers P2 (described above) and P3 (AGACGTCAGGTGGCAAAC), which anneal in opposite directions in the R6K origin and in the upstream region of the cloning site. PCR yields a 572-bp amplicon when at least two tandem copies are present in the genome, while no amplicon is produced otherwise. If not differently stated, the Chloramphenicol concentration used in all the steps of the site-specific recombination was 12.5 mg/l.
An analogous protocol was used for homologous recombination, with the exception that the passenger was present in the pBBintAsp vector and the Chloramphenicol concentration used was 8 mg/l. In this case, the pInt80-649 plasmid was used to replicate the integrative plasmid via the pir-116 gene, while the ϕ80 site-specific recombination machinery was not exploited, although it was expressed during the protocol. Other integration protocols were tested (see Sec. B.2.4).

B.1.6 Excision of antibiotic resistance and R6K origin from an integrant strain

The protocol described in [23] was used. Briefly, the integrant strain was grown in LB and made competent. This strain was transformed with the pCP20 helper plasmid [60], plated on Ampicillin plates and incubated at 30°C until colonies appeared. A single colony was propagated in non-selective LB at 37°C, 220 rpm overnight. The culture was streaked on a non-selective plate and incubated at 43°C. A single colony was propagated in non-selective LB to obtain the final single-copy integrant strain without the Chloramphenicol resistance marker and the R6K origin. The loss of the antibiotic marker and the pCP20 helper were validated by streaking the obtained strain on Chloramphenicol and Ampicillin plates respectively. The obtained integrant strain was also assayed through colony PCR with primer pair P1 (reported above) and P4 (CTCTTACGTGCCCGATCA) for ϕ80 attB targeting and with primer pair AspAFw (TGCGAGGATCGTGATGTATTTCGG) and AspARv (ATGATCTCGGGTATTCGGTCGATG) for aspA targeting. PCR products were also purified with the NucleoSpin Extract II kit and sequenced with the previously reported primer pairs or with VF2 and VR.

B.1.7 Characterization of promoters

MG1655 was used as chassis for the quantitative experiments. All the promoter characterization experiments were performed as described in App. A.1.2. Sterile medium and a non-fluorescent culture of MG1655 were always present in order to measure the background absorbance and fluorescence. Integrant strains were tested without antibiotic, while the plasmid-bearing strains were tested in presence of 12.5 mg/l of Chloramphenicol. When required, analogous

procedures were performed to test recombinant MC1061 strains.

B.1.8 Data Analysis

Data were analyzed with Microsoft Excel and the MATLAB 2007b suite (MathWorks, Natick, MA). Raw absorbance and fluorescence time series were processed as previously reported [32, 31, 10] to obtain a value proportional to the average RFP synthesis rate per cell in exponential phase (called S_{cell}). As values obtained by dividing the activity of a promoter of interest by the activity of a reference culture can enhance the reproducibility of processed data between experiments [10], the S_{cell} of promoters in each condition were divided by the S_{cell} of J23101 promoter in the low copy condition, measured in the same experiment, thus obtaining $S_{cell,norm}$. By definition, the $S_{cell,norm}$ value is 1 AU (Arbitrary Unit of RFP) per cell per time unit for the J23101 promoter in low copy. Relative Promoter Units (RPUs) were obtained as described previously [32, 31, 10], by dividing the $S_{cell,norm}$ of the promoter of interest, in a given copy number condition, by the $S_{cell,norm}$ of the J23101 standard reference promoter in the same copy number condition. By definition, the RPU of the J23101 promoter is 1 in all the copy number conditions.
The CV computation and the ANOVA test were performed as described in App. A.1.3.

B.1.9 Evolutionary stability experiments

Recombinant bacteria were grown as described above in 1 ml of medium in 2-ml tubes for 24 hours, starting from a single colony of a freshly streaked LB agar plate. Every 24 hours, the cultures grown to saturation were 1000-fold diluted in 1 ml of sterile medium and incubated under the same conditions as before, in order to achieve 10 generations per day [44]. Every day, 1 μl of each saturated culture was also added to 200 μl of sterile medium, incubated in the Infinite F200 reader and assayed as described above. Data were analyzed as before to obtain $S_{cell,norm}$. The activity of a promoter over time was expressed as percentage of the activity at generation 0 (i.e. the first point of the time series). Integrant strains were propagated and tested without antibiotic. The plasmid-bearing strains were propagated and tested in presence of 12.5 mg/l of Chloramphenicol at generation 0, while they were tested with and without antibiotic for the following generations. A freshly-inoculated non-fluorescent culture of MG1655 and a culture bearing the J23101 promoter with RFP in low copy vector were always assayed in order to perform the fluorescence background subtraction and the S_{cell} normalization with non-propagated cultures.

B.2 Supplementary results

B.2.1 Additional information about integrative base vector design

An early design of the pBBintϕ base vector included the default insert BBa_I52002, containing a pUC19-derived pMB1 replication origin and ccdB toxin constitutive expression cassette. It enabled the propagation of the plasmid in ccdB-tolerant strains, such as DB3.1, but not in other strains. The replication origin in this insert allowed the high-copy propagation of the plasmid in DB3.1 or similar strains even if they are not pir or pir-116. When the default insert had to be replaced with the BioBrick™ passenger of interest, the ligation product had to be transformed into a pir or pir-116 strain because this plasmid did not contain a standard replication origin anymore. Transformants with the uncut plasmid contaminant DNA could not grow because of the ccdB toxin in BBa_I52002.
Even if the BBa_I52002 had attractive features as a default insert, we decided to use another default insert (BBa_I763007, containing an mRFP1-expression cassette driven by the P_R promoter) because pBBintϕ with BBa_I52002 was not stably propagated by the DB3.1 host strain. In fact, this plasmid occasionally gave unexpected bands on agarose gel in digestion screenings, while pBBintϕ with the RFP cassette could be stably propagated by BW23474 strain, showing the correct bands every time.

B.2.2 Additional information about integrated BioBricks™ and phenotypes of recombinant strains

The pBBintϕ integrative vector was exploited to integrate a number of BioBricks™ in the ϕ80 *locus*. Supplementary Table 1 reports the full list of tested BioBricks™, successes and failures. In two cases, ϕ80-integrant clones of MG1655 did not behave as expected, although their sequencing was correct. In particular, after integration and marker excision one clone of the J23118 promoter with RFP measurement system downstream (BBa_I13507) showed an unexpectedly higher fluorescence (>3-fold) than the other clones with the same passenger in the genome. This could be due to additional integration events in the chromosome that were not included in the P1-P4 fragment and thus could not be detected by PCR screening. The other unexpected phenotype involved the PLlacO1 promoter (BBa_R0011) with RFP measurement system downstream (BBa_I13507). After integration and marker excision, the recombinant strain constitutively produced RFP independently from the IPTG concentration added, while the PLlacO1 was expected to show an IPTG-inducible behaviour. This could be due to lacI gene deletion that impaired the IPTG-inducible phenotype. In both cases, the obtained phenotypes were not further studied.

B.2.3 Additional information about integrative vector efficiency

In this study, the pBBintϕ was also specialized to target the aspA *locus*, thus obtaining pBBintAsp, in the genome of *E. coli* MG1655 to construct the recombinant strains used to study the context-dependent promoter activities. The pInt80-649 helper plasmid was used to maintain the integrative plasmid after transformation via the pir-116 gene, while the ϕ80 recombination machinery of the helper was not used. Although positive aspA-integrant clones could be obtained for all the BioBrick™ passengers under study, the efficiency of the integration process was significantly lower than measured for site-specific recombination in ϕ80. In fact, the transformation plate of MG1655 (with pInt80-649 and pBBintAsp containing the BioBrick™ device of interest) incubated at 30°C showed a small set of colonies (<1%) with red phenotype, probably due to the correct RFP expression cassette in the integrative plasmid, and a large set of colonies without detectable colour. Only the red colonies could be successfully integrated in single copy, with an efficiency close to 100% in all the required steps (i.e., helper plasmids loss, marker excision, correct phenotype and sequencing) for all the five promoters used in this study. The other colonies did not show a correct phenotype (validated fluorimetrically) and, when picked after the first incubation at 42°C on agar plate they did not show a VF2-VR PCR product, demonstrating that the desired insert was not present in the genome. On the other hand, when pBBintAsp containing the BioBrick™ device of interest was transformed into BW23474 (a pir-116 strain) all the colonies showed the red phenotype.

B.2.4 Additional integration protocols and results

All the integration experiments described so far in this thesis involved the protocols reported in Sec. B.1.5. However, other protocols were tested and they are described below.

- *ϕ80-targeting by using pAH123 as helper plasmid.* Instead of using pInt80-649 as helper plasmid (containing a heat-inducible ϕ80 phage integrase, a low copy temperature sensitive replication origin and a constitutive cassette of pir-116), the pAH123 helper plasmid [59] was also used (prepared from the CGSC#7861 strain), which is identical to pInt80-649, but lacks the pir-116 gene. In this case, the conditional-replication integrative vector could not be propagated in the host strains for integration carrying this helper. pBBintϕ with the default insert BBa_I763007 was used in this experiment, thus targeting the BBa_I763007 device into the ϕ80 attB of MG1655 and MC1061. Competent host strains were heat-shock transformed with pAH123, propagated in selective media and made competent again. About 3 μg of miniprepped pBBintϕ-BBa_I763007 were heat-shock transformed. After heat shock, 1 ml of pre-warmed LB

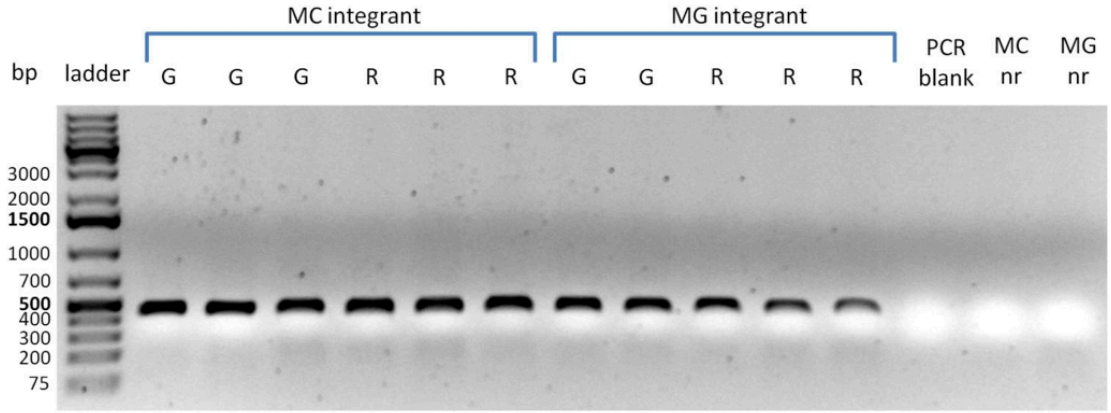


Figure B.2: **Colony PCR on ϕ80-integrant strains with primers P1-P2 in a representative experiment.** The primers anneal in opposite directions in the genome and in the integrative plasmid respectively (see Sec. B.1.5). If the integration position is correct, a 452-bp amplicon is produced. MC = MC1061 strain; MG = MG1655 strain; G = GFP cassette as passenger (BioBrick™ BBa_K173001); R = RFP cassette as passenger (EcoRI-PstI fragment of BBa_J23101 in the BBa_J61002 vector). Control reactions on non-recombinant (nr) strains are also reported.

was added and cells were incubated at 42°C with shaking for 2 hours. Cells were plated on LB + Chloramphenicol at 12.5 mg/l and incubated overnight at 42°C. Both MG1655 and MC1061 showed colonies (85 and 2000 respectively). For each strain, 3 colonies were screened and they resulted correct integrants (positive 452-bp amplicon in P1-P2 PCR) with at least 2 tandem copies of the desired part (positive 572-bp amplicon in P2-P3 PCR). All the clones lost the pAH123 helper plasmid (verified by Ampicillin counterselection). The phenotype was validated fluorimetrically (data not shown), thus demonstrating that the part of interest was functional.

- *aspA-targeting in direct suicide mode [54].* Homologous recombination was attempted by directly transforming the non-replicating plasmid pBBintAsp (10 μg of DNA), with BBa_I763007 as passenger, into competent MG1655 and MC1061 and by plating the transformed cells on Chloramphenicol (8 or 12.5 mg/l) plates. Unfortunately, no colonies could be obtained, probably because of the too low transformation efficiency of the used strains with heat-shock procedure (it was about 10^5 CFU per μg of DNA for MG1655 and 10^6 CFU per μg of DNA for MC1061, estimated from transformation experiments with control plasmids).

B.3 Supplementary figures and tables

Table B.3: **BioBricks™ used as passengers in the integration experiments** If not differently stated, integration/marker excision succeeded and phenotype was correct. Passenger names are expressed as BioBrick™ codes, where the '+' sign means that the passenger is the result of a Standard Assembly of BioBricks™.

Passenger	**Host strain and integration *locus***	**Notes**
BBa_J107028 (BBa_J23100+BBa_I13507)	MG1655, ϕ80 and MG1655, aspA	-
BBa_J107029 (BBa_J23101+BBa_I13507)	MG1655, ϕ80 and MG1655, aspA	-
BBa_J107031 (BBa_J23118+BBa_I13507)	MG1655, ϕ80 and MG1655, aspA	-
BBa_J107012 (BBa_I14032+BBa_I13507)	MG1655, ϕ80 and MG1655, aspA	-
BBa_I763007 (BBa_R0051+BBa_I13507)	MG1655, ϕ80; MG1655, aspA and MC1061, ϕ80	-
BBa_J107010 (BBa_R0011+BBa_I13507)	MG1655, ϕ80	No integrant clones with correct phenotype were obtained (see Sec. B.2.2)
BBa_J107011 (BBa_I14032+BBa_E0240)	MG1655, ϕ80	Phenotype could not be validated, as our instrument gave a too high background fluorescence when using the GFP filters (Excitation: 485 nm, Emission: 540 nm) on strains with a single copy of a GFP-expressing device

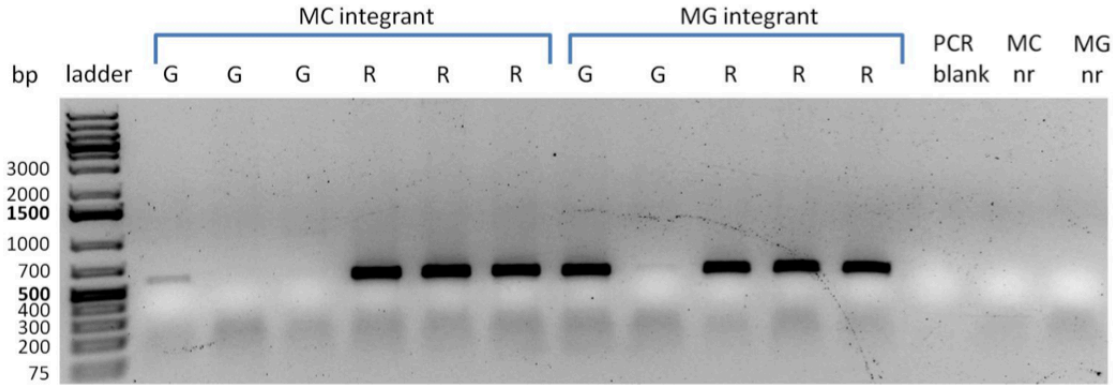


Figure B.3: **Colony PCR on ϕ80-integrant strains with primers P2-P3 in a representative experiment.** Both primers anneal in the integrative plasmid. Reaction produces a 572-bp amplicon if two or more tandem integrants are present in the genome (see Sec. B.1.5). MC = MC1061 strain; MG = MG1655 strain; G = GFP cassette as passenger (BioBrick™ BBa_K173001); R = RFP cassette as passenger (EcoRI-PstI fragment of BBa_J23101 in the BBa_J61002 vector). Control reactions on non-recombinant (nr) strains are also reported.

EcoRI-PstI fragment of BBa_J23100 in the BBa_J61002 vector	MG1655, ϕ80	-
EcoRI-PstI fragment of BBa_J23101 in the BBa_J61002 vector	MG1655, ϕ80 and MC1061, ϕ80	This part was also characterized in another work where the integrative vector reported here was used to construct the clones [31]
EcoRI-PstI fragment of BBa_J23118 in the BBa_J61002 vector	MG1655, ϕ80	-
BBa_K173001 (BBa_J23101+BBa_E0240)	MG1655, ϕ80 and MC1061, ϕ80	Phenotype could not be validated, as our instrument gave a too high background fluorescence when using the GFP filters (Excitation: 485 nm, Emission: 540 nm) on strains with a single copy of a GFP-expressing device
EcoRI-PstI fragment of BBa_F2620 in the BBa_J61002 vector	MG1655, ϕ80	This part was also characterized in another work where the integrative vector reported here was used to construct the clones [31]

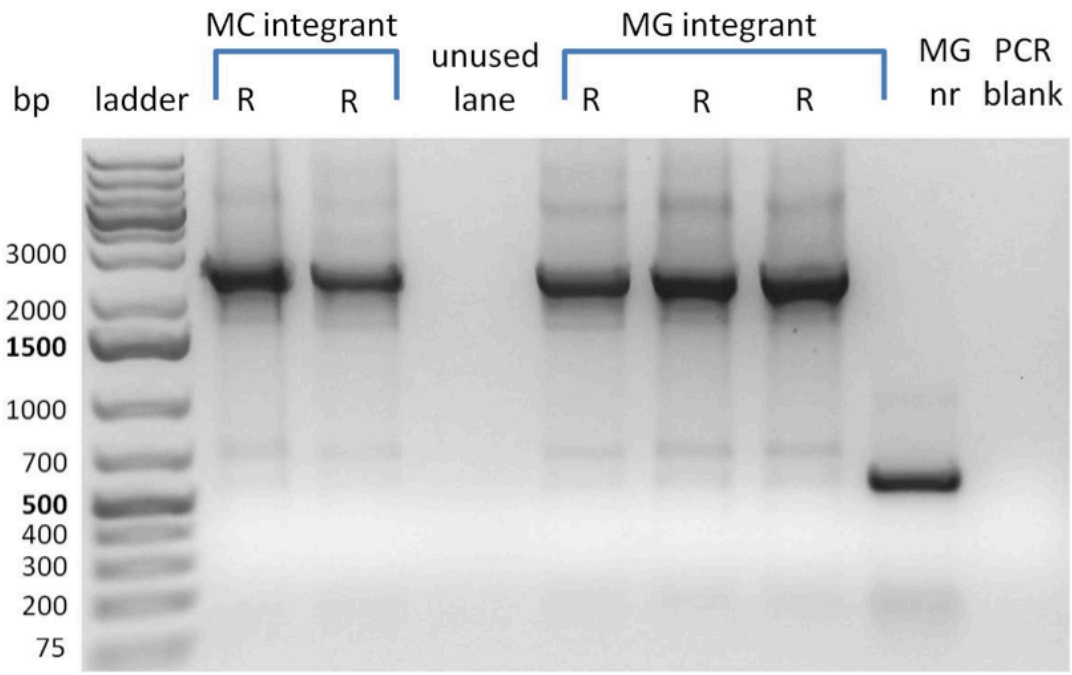


Figure B.4: **Colony PCR on ϕ80-integrant strains, after the FRT/Flp-mediated marker excision, with primers P1-P4 in a representative experiment.** Both primers anneal in the genome in opposite directions. Reaction produces a 2.3-Kbp amplicon if a single integrant of the passenger, without Chloramphenicol resistance or R6K origin, is present in the ϕ80 genomic *locus* (see Sec. B.1.5). MC = MC1061 strain; MG = MG1655 strain; R = RFP cassette as passenger (EcoRI-PstI fragment of BBa_J23101 in the BBa_J61002 vector). A control reaction on a non-recombinant (nr) strain is also reported.

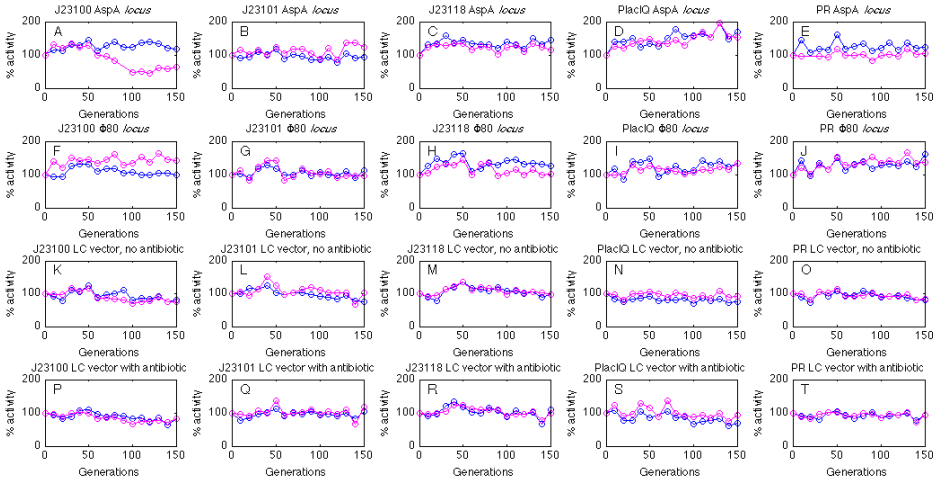


Figure B.5: **Evolutionary stability of the individual clones of integrant strains and low copy plasmid-bearing strains.** The studied BioBrick™ promoters expressing RFP were integrated in the ϕ80 and the aspA *loci*, propagated without antibiotic, or they were carried on the low-copy vector pSB4C5, propagated with or without antibiotic. The panels show the percent activity of all the five studied promoters in two experimental replicates over time in each condition, where 100% represents the activity measured at generation 0.

Appendix C

Methods and Supplementary Information for Ch. 4

C.1 Methods

C.1.1 Strains and plasmids

The description of all the *Escherichia coli* strains, vector backbones and genetic devices used in this work is shown in Tab. 4.1. The description of plasmids assembly steps is reported in the App. C.2.1. All the plasmids containing ccdB toxin were cloned in the ccdB-tolerant strain DB3.1. All the conditional-replication plasmids (pϕ80 series) were cloned in the pir-116 strain BW23474. TOP10 strain was used to clone all the other plasmids. MG1655 strain was used as a host for all the quantitative experiments of this study.

C.1.2 Cloning methods

Chemically competent TOP10 and DB3.1 were used according to manufacturer's instructions. Chemically competent MG1655 and BW23474 were prepared according to [94]. All the cell and DNA manipulation procedures were performed as described in App. B.1.2.

C.1.3 Integrant strains

MG-HSL$_{RFP}$, MG-101$_{RFP}$ and MG-IQ$_{RFP}$ were obtained by integrating HSL$_{RFP}$, 101$_{RFP}$ and IQ$_{RFP}$ into the MG1655 host (through pϕ80-HSL$_{RFP}$, pϕ80-101$_{RFP}$ and pϕ80-IQ$_{RFP}$ integrative plasmids, respectively), as described in App. B.1.5 and App. B.1.6 and in the BBa_K300000 integrative vector page in the Registry of Standard Biological Parts. After genomic integration (mediated by the pInt80-649 helper plasmid [23]) and marker excision (mediated by the pCP20 helper plasmid [56]), only one copy of the expres-

sion device of interest remains in the *E. coli* chromosome and it is flanked by four transcriptional terminators to achieve isolation from genomic context. No antibiotic resistance is present at the end of the procedure, so the resulting recombinant strains are always grown in nonselective medium. Integrants were PCR-verified with primers P1 (5'-CTGCTTGTGGTGGTGAAT-3') and P4 (5'-TAAGGCAAGACGATCAGG-3'), annealing in the genome (see Sec. B.1.5 and [59]). Platinum Taq DNA polymerase (Invitrogen) was used for PCR. When required, amplified fragments were gel-extracted as described in App. B.1.2 and sequenced.

C.1.4 Population-based fluorescence analysis

Recombinant strains were tested to measure RPUs as described in App. A.1.2. Here, the 1:500 dilution was replaced by a dilution at the optical density at 600 nm (OD_{600}) of 0.03. In this appendix, OD_{600} values refer to the pathlength constituted by 200 μl of bacterial culture in a 96-well microplate, measured by the F200 reader. They have not been converted to 1-cm pathlength absorbance values. The cultures were then incubated for 45 minutes before the transferring in the microplate reader. For the cultures bearing the HSL-inducible system, 200 μl for each investigated HSL concentration were transferred and 2 μl of properly diluted HSL (Sigma Aldrich) were added to the culture before incubation to yield the desired concentration.
A preliminary evaluation of the measurement system was carried out and results are reported in the App. C.2.2.

C.1.5 Single-cell experiments

Recombinant strains expressing RFP or GFP were grown as described above and induced with HSL when required. After 2/3 hours from the induction, cells were either spread on a glass slide (for microscopy analysis) or properly diluted in sterile PBS (for flow cytometric analysis).

RFP-expressing cells were analyzed with an Olympus BX51 microscope with standard fluorescence equipment (HBO100W/2 lamp). Microphotographs were taken using an Olympus Camedia C-4040 digital camera. Fluorescence was detected through the green excitation performed with a 530-560nm band pass excitation filter and a 590nm dichroic mirror combined with a long pass barrier filter at 620nm. UPlanFl 40X objective was used for all the acquisitions.

GFP-expressing cells were analyzed through a Partec PAS II flow cytometer equipped with an argon ion laser using the 488nm blue line for excitation. Fluorescence emission was collected in FL1 by means of a 515-545nm band pass filter. 100,000 events were collected and stored for each sample.

C.1.6 Data analysis

RPUs were computed as described in App. A.1.3.

Induction curves were fitted with the Hill equation $Y = V_{max} * \frac{I^q}{{K_m}^q + I^q}$, where V_{max} is the maximum activity, Y can be either the S_{cell} or the RPU of a system and I can be either the HSL concentration or the intracellular level of the HSL-LuxR activator complex, depending on the specific application. Basal activity of the lux promoter was omitted from the Hill equation because its entity was much lower than V_{max} in all the curves. The least squares fitting was performed through the MATLAB *lsqnonlin* routine. Among the presented results, S_{cell} were shown for a set of cultures expressing RFP. Because a wide range of fluorescence intensities is produced by the studied cultures and then to have a good sensitivity in the acquisition process the fluorescence acquisition gain was tuned accordingly, S_{cell} results had to be reported to the same gain (in this work a gain=50 was chosen) to enable comparisons among values. To this aim, the fluorescence of a MG1655 culture bearing pHC-HSL$_{RFP}$, grown for 6 hours in presence of 100 nM of HSL and diluted to OD_{600}=0.01, was measured with different gain factors (50, 60, 70, 80, 90 and 100), set via i-control™software. All the collected measurements were divided by the RFP raw value at gain=50, to compute RFP$_{norm,50}$ and a calibration curve of gain vs RFP$_{norm,50}$ was obtained (data not shown). Thus, S_{cell} values computed from data acquired at gain=β were reported to gain=50 by dividing them by the right conversion factor between β and 50, i.e. the RFP$_{norm,50}$ value of the calibration curve corresponding to gain β.

Bright-field pictures acquired from the microscope were processed through the ImageJ software (Wayne Rasband, NIH) to enhance the contrast of photographed cells when required. Fluorescence images were not processed in any way.

Flow cytometric data analysis was carried out with the FloMax (Partec, Munster, Germany) software.

C.1.7 Mathematical modeling

The following set of differential equations was used to describe the dynamics of LuxR (X), HSL-LuxR activated complex (A) and RFP (R) as a function of HSL (H) in the HSL-inducible system:

$$\frac{dX}{dt} = n \cdot \alpha_{tet} - \gamma_X \cdot X \tag{C.1}$$

$$A = \frac{X}{1 + \left(\frac{K_H}{H}\right)^{n_H}} \tag{C.2}$$

$$\frac{dR}{dt} = \frac{n \cdot \alpha_{lux}}{1 + \left(\frac{K_A}{A}\right)^{n_A}} - \mu \cdot R \tag{C.3}$$

The description of the species and parameters is reported in Tab. C.1. Parameter values that have not been estimated in this study are described in [91]. The degradation rate of RFP is much lower than the cell growth rate, so only the dilution caused by cell division contributes to the intracellular extinction of RFP (results not shown). On the other hand, dilution rate was neglected for LuxR when compared to the protein degradation rate [91]. The observable (measured) variable in each experiment is the RFP synthesis rate per cell, $S_{cell} = \frac{n\alpha_{lux}}{1+(\frac{K_A}{A})^{n_A}}$. This is measured at the steady state and the maturation dynamics of the fluorescent protein [8] was not taken into account. Considering the steady state, $\overline{X} = \frac{n\alpha_{tet}}{\gamma_X}$, $\overline{A} = \frac{\overline{X}}{1+(\frac{K_H}{H})^{n_H}}$ and $\overline{S}_{cell} = \frac{n\alpha_{lux}}{1+(\frac{K_A}{A})^{n_A}}$, where the bar indicates that the species is at the steady state. Model fitting was performed from RPU data points instead of S_{cell} to improve the reliability of the measured data. In particular, the α_{lux} parameter was computed by multiplying the RPUs at full induction (indicated with V_{max}) by the $S_{cell,101_{RFP}}/n$ term, where $S_{cell,101_{RFP}}$ is referred to the 101_{RFP} device at the copy number n.

C.2 Additional results

C.2.1 Plasmid construction

Tab. C.2 lists all the plasmids used and/or assembled during this study. Plasmids containing the five final genetic systems used in quantitative experiments (HSL_{RFP}, 101_{RFP} and IQ_{RFP}, HSL_{GFP} and 101_{GFP}) are in bold-type. pSCANSD is a pSCANS-derived vector (pSB2K3 in the Registry of Standard Biological Parts) which is propagated at high-copy number in TOP10 and recombinant strains bearing such plasmid vector were selected with Kanamycin at 50 mg/l during cloning [16].

pHC-HSL_{RFP} was obtained by assembling the promoterless RFP-containing insert of pHC-101_{RFP} digested with SpeI-PstI to pHC-HSL digested with SpeI-PstI. The insert of pHC-$_{RFP}$ was digested with XbaI-PstI and assembled to pSCANSD-IQ, digested with SpeI-PstI, to obtain pSCANSD-IQ_{RFP}. pHC-IQ_{RFP} was obtained by assembling the insert of pSCANSD-IQ_{RFP} digested with EcoRI-PstI to pHC-101_{RFP} vector backbone, digested with EcoRI-PstI, thus eliminating its original insert. All the other plasmids were obtained by assembling the inserts of pHC-HSL_{RFP}, pHC-101_{RFP}, pHC-IQ_{RFP}, pHC-HSL_{GFP} or pHC-101_{GFP}, digested with EcoRI-PstI, to the appropriate vector backbones (from pMC-LAC_{RFP}, pLC-ccdB or pϕ80-ccdB plasmids) digested with EcoRI-PstI to eliminate their original inserts.

C.2.2 Validation of the measurement system

The main purpose of this work is to study nonlinear effects in a biological system, measuring suitable outputs. It is then crucial to preliminarily verify

Table C.1: **Species and parameters included in the mathematical model.** Parameter values that have not been estimated in this study are described in [91]. AU indicates arbitrary units of RFP.

Parameter or species	Description	Value	Units
X	LuxR protein concentration per cell	Measured	AU
A	LuxR-HSL activated complex concentration per cell	Variable	AU
R	RFP per cell	Variable	AU
H	Inducer concentration	Variable	nM
n	DNA copy number	Measured	-
α_{tet}	LuxR synthesis rate per cell per DNA copy	1.2	AU min^{-1} $cell^{-1}$
α_{lux}	Maximum RFP synthesis rate per cell per DNA copy	Measured	AU min^{-1} $cell^{-1}$
γ_X	LuxR protein degradation rate	$6*10^{-2}$	min^{-1}
μ	Cell growth rate	Measured	min^{-1}
K_H	Dissociation constant of HSL-LuxR	553	nM
K_X	Dissociation constant of HSL-LuxR complex and lux promoter	Measured	AU
n_H	Hill cooperativity constant of HSL-LuxR	2	-
n_X	Hill cooperativity constant of of HSL-LuxR complex and lux promoter	Measured	-

Table C.2: **Plasmids used and/or assembled during this study**

[a] taken from the Registry DNA Distribution 2009

[b] previously constructed in our laboratory

[c] given by the iGEM Headquarters, Massachusetts Institute of Technology, Cambridge, USA [d] constructed in this study

The BioBrick™ code of the plasmids includes vector name with insert name in brackets. BBa_J61002, BBa_J107055, BBa_J107056 and BBa_J107057 are identical to pSB1A2, pSB3K3, pSB4C5 and BBa_K300000, respectively, but they have a promoterless mRFP1 expression device between SpeI and PstI in the BioBrick™ cloning site suffix. pSB1A3 is identical to pSB1A2, but it has an additional transcriptional terminator downstream of the cloning site.

The prefix pHC- indicates a high copy vector backbone (pSB1A2, pSB1A3 or BBa_J61002), pMC- indicates a medium copy vector backbone (pSB3K3 or BBa_J107055), pLC- indicates a low copy vector backbone (pSB4C5 or BBa_J107056), pϕ80- indicates an integrative conditional replication vector (BBa_K300000 or BBa_J107057) and pSCANSD- indicates a variable copy number vector backbone (pSB2K3).

	Plasmid	
Name	Description	BioBrick™ code
pMC-LAC$_{RFP}$	Hybrid promoter with lac operator sites with mRFP1 downstream in medium copy plasmid	pSB3K3 (BBa_J04450)[a]
pLC-ccdB	pUC19-derived pMB1 replication origin and ccdB toxin constitutive expression cassette in low copy vector	pSB4C5 (BBa_I52002)[a]
pϕ80-ccdB	pUC19-derived pMB1 replication origin and ccdB toxin constitutive expression cassette in the conditional replication integrative vector	BBa_K300000 (BBa_I52002)[b]
pHC-HSL	HSL-inducible device in high copy plasmid	pSB1A2 (BBa_F2620[c])

pHC-101$_{RFP}$	Standard reference constitutive promoter BBa_J23101 with mRFP1 expression device downstream in high copy plasmid	BBa_J61002 (BBa_J23101)[a]
pSCANSD-IQ	lacIQ constitutive promoter in pSCANS-derived plasmid	pSB2K3 (BBa_I14032)[a]
pSCANSD-IQ$_{RFP}$	lacIQ constitutive promoter with mRFP1 expression device downstream in pSCANS-derived plasmid	BBa_J107058 (BBa_I14032)[d]
pHC-IQ$_{RFP}$	lacIQ constitutive promoter with mRFP1 expression device downstream in high copy plasmid	BBa_J61002 (BBa_I14032)[d]
pHC-HSL$_{RFP}$	HSL-inducible mRFP1 expression system in high copy plasmid	BBa_J61002 (BBa_F2620)[d]
pMC-HSL$_{RFP}$	HSL-inducible mRFP1 expression system in medium copy plasmid	BBa_J107055 (BBa_F2620)[d]
pMC-101$_{RFP}$	Standard reference constitutive promoter BBa_J23101 with mRFP1 expression device downstream in medium copy plasmid	BBa_J107055 (BBa_J23101)[d]
pMC-IQ$_{RFP}$	lacIQ constitutive promoter with mRFP1 expression device downstream in medium copy plasmid	BBa_J107055 (BBa_I14032)[d]

pLC-HSL$_{RFP}$	HSL-inducible mRFP1 expression system in low copy plasmid	BBa_J107056 (BBa_F2620)[d]
pLC-IQ$_{RFP}$	lacIQ constitutive promoter with mRFP1 expression device downstream in low copy plasmid	BBa_J107056 (BBa_I14032)[d]
pϕ80-HSL$_{RFP}$	HSL-inducible mRFP1 expression system in the conditional replication integrative vector	BBa_J107057 (BBa_F2620)[d]
pϕ80-101$_{RFP}$	Standard reference constitutive promoter BBa_J23101 with mRFP1 expression device downstream in the conditional replication integrative vector	BBa_J107057 (BBa_J23101)[d]
pϕ80-IQ$_{RFP}$	lacIQ constitutive promoter with mRFP1 expression device downstream in the conditional replication integrative vector	BBa_K300000 (BBa_J107012)[d]
pHC-101$_{GFP}$	Standard reference constitutive promoter BBa_J23101 with GFPmut3b expression device downstream in high copy plasmid	pSB1A2 (BBa_K173001)[b]
pHC-HSL$_{GFP}$	HSL-inducible GFPmut3b expression system in high copy plasmid	pSB1A3 (BBa_T9002)[a]
pMC-HSL$_{GFP}$	HSL-inducible GFPmut3b expression system in medium copy plasmid	pSB3K3 (BBa_T9002)[d]
pMC-101$_{GFP}$	Standard reference constitutive promoter BBa_J23101 with GFPmut3b expression device downstream in medium copy plasmid	pSB3K3 (BBa_K173001)[d]

pLC-HSL$_{GFP}$	HSL-inducible GFPmut3b expression system in low copy plasmid	pSB4C5 (BBa_T9002)[d]
pLC-101$_{GFP}$	Standard reference constitutive promoter BBa_J23101 with GFPmut3b expression device downstream in low copy plasmid	pSB4C5 (BBa_K173001)[d]

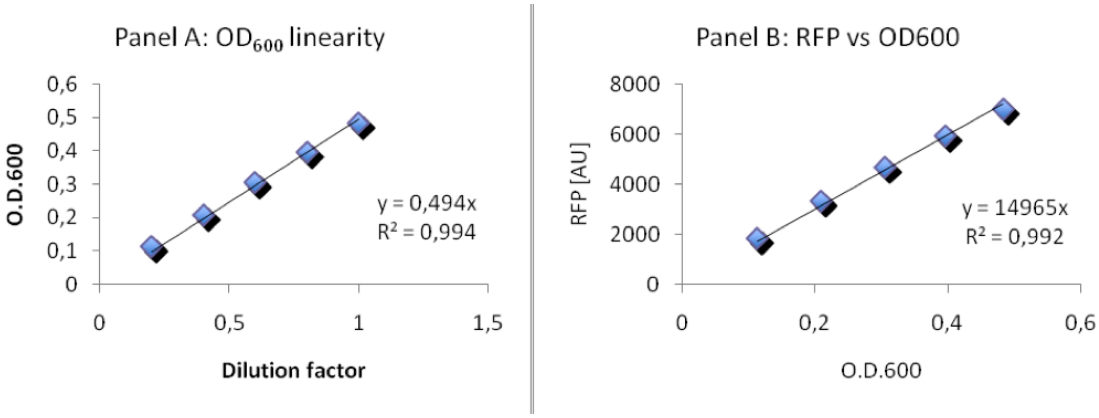


Figure C.1: **Linearity between OD_{600} and dilution factor (panel A) and between RFP and OD_{600} measurements (panel B).** Experiments were performed on a culture of MG1655 bearing pHC-HSL_{RFP} induced at 100 nM. Diamonds represent experimental data points and continuous line represents the linear regression line. AU indicates arbitrary units of RFP.

that the measurement system itself, i.e. the TECAN Infinite F200, does not introduce nonlinearities in the measurements, at least in the range of the working conditions of interest. To this aim, the following *ad-hoc* experiment was performed: a fluorescent culture of MG1655 bearing pHC-HSL_{RFP}, grown to OD_{600}=0.5 in selective M9 supplemented medium and induced with 100 nM of HSL, was serially diluted in fresh medium and absorbance (OD_{600}) and fluorescence (RFP) were statically measured (see the main text, Methods section, for technical details). After a background subtraction (see Ch. 4 for details), OD_{600} measurements were plotted against the dilution factor and RFP was plotted against the OD_{600}. The resulting curves are shown in Fig. C.1.

Both absorbance and fluorescence measurements show a linear dependence from the dilution (R2>0.99 for both of them). Note that in the quantitative experiments performed in this work measurements were taken only in the exponential growth phase, which ends up at OD_{600} <0.2 for all the tested cultures. Therefore, in this range, the measurement system does not introduce nonlinear effects during data acquisition. Similar experiments were done with GFP-expressing cells and they demonstrated the absence of nonlinearities also for GFP fluorescence detection (data not shown). Moreover, in order to verify that changes in the fluorescence gain do not introduce nonlinear effects, the above procedure was repeated at different gain factors. Results confirm the linear relationships for all the investigated fluorescence acquisition gains (data not shown).

C.2.3 Single-cell analysis

Homogeneity of reporter expression level throughout the population was studied in order to assess if all the bacterial cells can respond to induction by

HSL. Recombinant strains bearing the RFP- or GFP-expressing HSL-inducible systems in medium copy number were induced via three different HSL concentrations and analyzed through fluorescence microscopy or flow cytometry. Fig. C.2 reports representative microscopy images and fluorescence histograms. Fig. C.2 A shows that all the cells detected in the bright-field express red fluorescence upon induction for the HSL-inducible device with 1 nM, 10 nM and 10 μM of HSL. Uninduced bacteria show a much lower per-cell fluorescence. Such low fluorescence intensity is not seen in any cell when cultures are induced. Fluorescence histograms in Fig. C.2 B show that GFP distributions are all unimodal, with their average value increasing with the HSL concentration, as expected from the induction curves (see Fig. 4.2). As observed in the microscopy images, in the histograms of the three induced cultures no significant amount of the cell population expresses the typical fluorescence intensity seen in the GFP histogram of the uninduced culture. The overall results of microscopy and flow cytometry are in accordance and demonstrate that induced bacterial populations homogeneously respond to HSL.

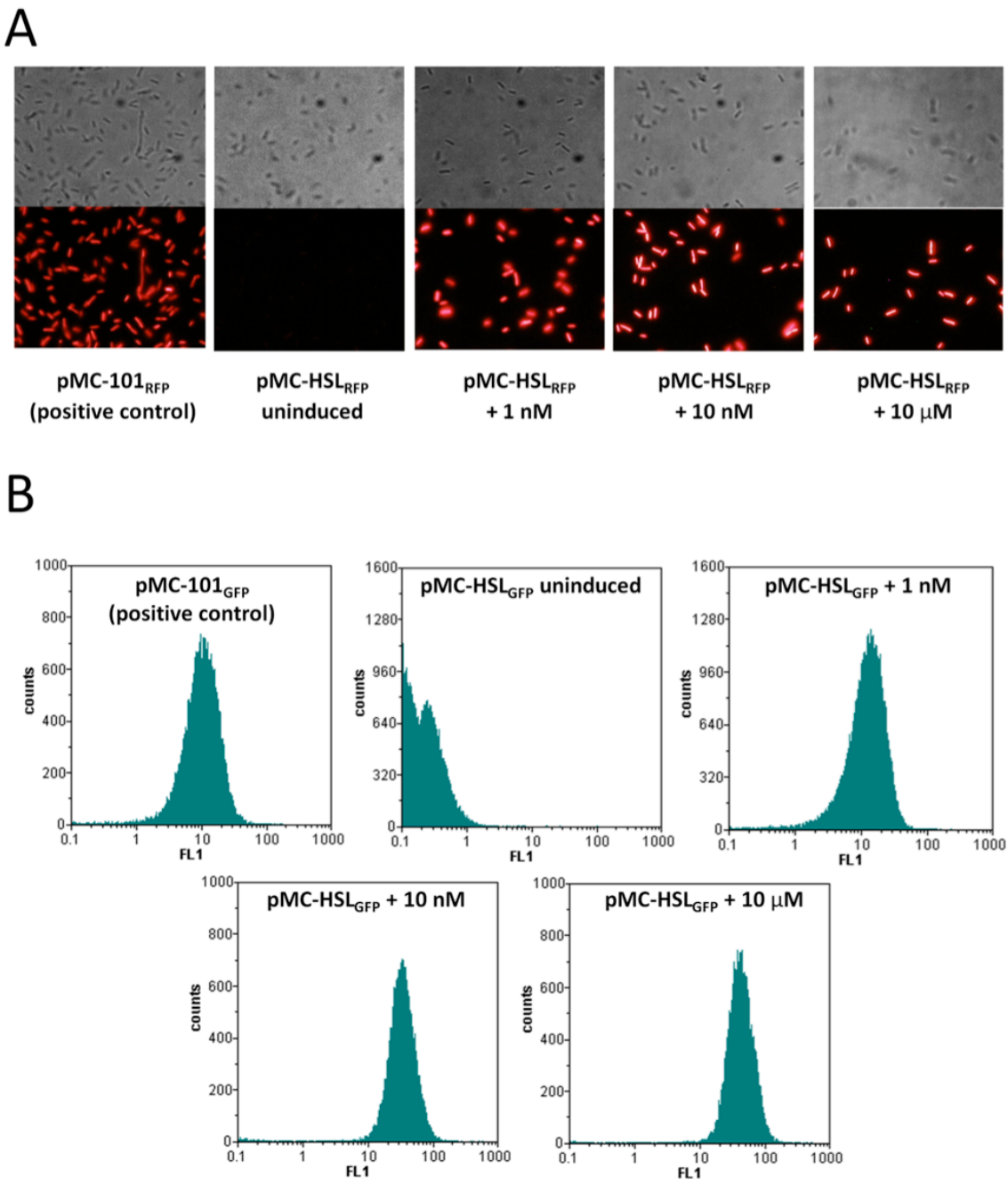


Figure C.2: **Bright-field and fluorescence images and fluorescence histograms resulting by flow cytometry measurements** A) Bright-field and fluorescence images acquired in the microscopy analysis for the RFP-expressing HSL-inducible device in medium copy. All the fluorescence images have been acquired with the same exposition time of 1.3 s. B) Fluorescence histograms acquired in the flow cytometric analysis for the GFP-expressing HSL-inducible device in medium copy. In all the plots, the x-axis represents the fluorescence intensity (in arbitrary units of GFP) and the y-axis represents the cell count.

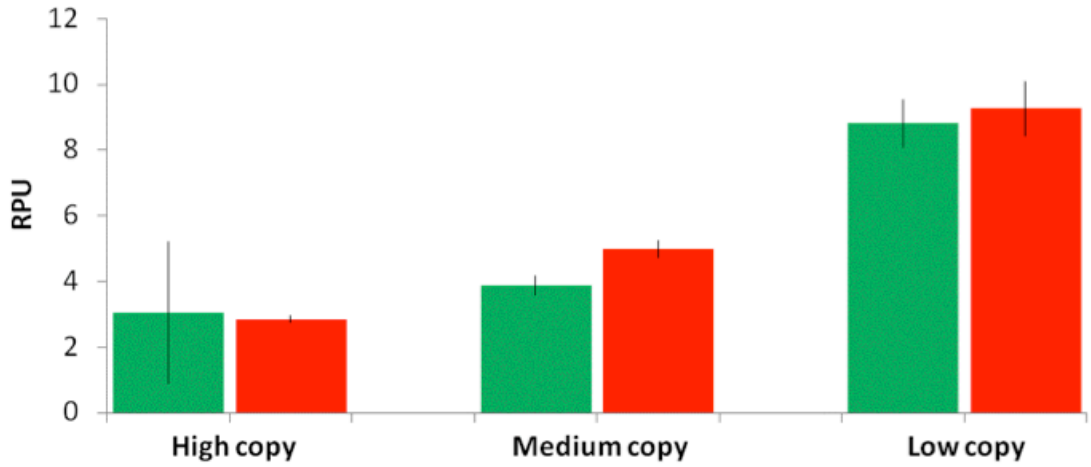


Figure C.3: **Comparison of the relative activity of the HSL-inducible device at full induction expressing GFP or RFP.** Full induction = induced with 10 μM of HSL. Activity measured via GFP (green bars) or RFP (red bars) reporter devices. Error bars represent the standard deviation of three independent experiments.

Appendix D

Methods and Supplementary Information for Ch. 5

D.1 Materials and methods

D.1.1 Model simulation and induction curves computation

The ODE system used in the simulations is here reported. All the simulations were performed via Matlab R2009b software (MathWorks, Natick, MA), using the 'ode23s' routine for the numerical system solving ('AbsTol' property was set to 10^{-12}). The parameters and initial conditions are listed in Tab. 5.1. The simulation time was set to 1000 seconds and increased when required to ensure that all the species dynamics reached the steady state values.
The model has been simulated for different H_0 input values within the range $[10^{-10} - 10^{-5}]$ M. For each considered input, the $P_{C_{SS}}$ output was computed as the steady state value reached by P_C species. The $P_{C_{SS}}$-H_0 characteristic corresponds to the induction curve of lux promoter.

$$\frac{dX}{dt} = -k_{+1} \cdot X \cdot H + k_{-1} \cdot D \quad \text{(D.1)}$$

$$\frac{dH}{dt} = -k_{+1} \cdot X \cdot H + k_{-1} \cdot D \quad \text{(D.2)}$$

$$\frac{dD}{dt} = k_{+1} \cdot X \cdot H - k_{-1} \cdot D - 2 \cdot k_{+2} \cdot D^2 + 2 \cdot k_{-2} \cdot C \quad \text{(D.3)}$$

$$\frac{dC}{dt} = k_{+2} \cdot D^2 - k_{-2} \cdot C - k_{+3} \cdot P_F \cdot C + k_{-3} \cdot P_C \quad \text{(D.4)}$$

$$\frac{dP_F}{dt} = -k_{+3} \cdot P_F \cdot C + k_{-3} \cdot P_C \quad \text{(D.5)}$$

$$\frac{dP_C}{dt} = k_{+3} \cdot P_F \cdot C - k_{-3} \cdot P_C \quad \text{(D.6)}$$

$$\frac{dM}{dt} = k_{+4} \cdot P_C + k_{+5} \cdot P_F \quad \text{(D.7)}$$

D.1.2 Numerical sensitivity analysis

Sensitivity analysis on model parameters was performed by varying them one at a time of -3, -2, -1, 1, 2 and 3 orders of magnitude from the nominal value reported in Tab. 5.1. For each parameter variation, the corresponding induction curve was computed.
In order to extract summary statistics to relate the system behaviour to the parameter variation, the following indicators have been computed.
P_{max} is the maximum value of $P_{C_{SS}}$. It indicates how much affected the availability of activated promoter, and thus the synthesis rate of the gene of interest, is by the parameter variation.
H_{50} is the induction required to obtain a $P_{C_{SS}}$ value equal to 50% of P_{max}. H_{10} and H_{90} values are the H_0 values for which the output is 10% and 90% of P_{max}, respectively. H_{50} indicates where the switch-point of the induction curve and where the range of working inductions [H_{10}-H_{90}] are located. This interval represents the set of inductions that can be given in input to the lux system to effectively regulate its transcription rate. The broader this range is, the more a robust behaviour in response to noisy inputs can be achieved.
In order to define a measurement for the amplitude of this interval comparable among different contexts, the relative measurement $\mathrm{H}^* = \frac{H_{90}-H_{10}}{H_{50}}$ was computed. This value represents the average ratio between the interval width and its central point, provided that the interval may be not symmetrically distributed around H_{50}. As an example, two hypothetical cases can be considered, for which the indicators assume the values:
case 1: H_{50}=100, H_{10}=10, H_{90}=1000
case 2: H_{50}=1000, H_{10}=100, H_{90}=10000
The amplitude of $H_{10} - H_{90}$ range is 900 in case 1 and 9000 in case 2. H^* assumes the same value of 9,9 in both cases, since the interval boundaries are one order of magnitude distant from the H_{50} values.

D.1.3 Experimental data of promoter activity at different DNA copy numbers

The model has been validated using two different datasets, obtained by two different experimental layouts, as shown in Fig. D.1.

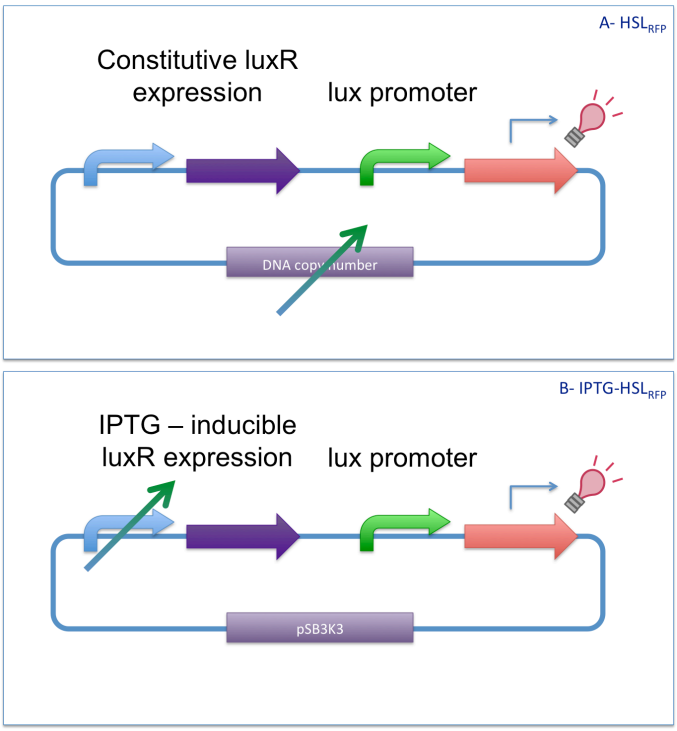


Figure D.1: **Two different experimental layouts to obtain copy-number variation data.** HSL$_{RFP}$ device (panel A) has been used in experiments where the copy number of the whole device (both luxR expression cassette and lux promoter driving the expression of RFP) has been varied. IPTG-HSL$_{RFP}$ device in pSB3K3 medium copy plasmid (panel B) has been used to tune luxR production maintaining lux promoter copy number constant. Different IPTG inductions cause the copy number variation of the LuxR transcription factor. In this work, LuxR protein concentrations as a function of IPTG have not been measured.

The HSL inducible device expressing RFP (HSL$_{RFP}$, see Tab. 4.1) has been tested as described in Ch. 4 in MG1655 strain and the experimental data have been presented in Sec. 4.2.2 for four different copy number contexts (high copy, medium copy, low copy and single copy).
A different HSL inducible device expressing RFP (IPTG-HSL$_{RFP}$, BioBrick™ code BBa_J107053) has been tested in *E. coli* MG1655-Z1 strain (MG1655 strain that constitutively overexpresses *lacI* from its chromosome [74]). It is composed by BBa_F2622 HSL-inducible device driving the expression of mRFP and has been tested in the medium copy number plasmid pSB3K3. In this device, LuxR expression is regulated by P_{LlacO1} promoter, whose activity can be triggered by isopropyl β-D-1-thiogalactopyranoside (IPTG).

Experiments to quantify IPTG-HSL$_{RFP}$ activity and to measure its induction curve have been performed as described in App. A.1.2 and data analysis was carried out as described in App. A.1.3, using BBa_J23101 in pSB3K3 plasmid in MG1655-Z1 strain as a reference standard promoter and the MG1655-Z1 strain as a negative control. All the experiments were performed in M9 supplemented medium with the addition of Kanamycin (20 mg/l), with the exception of the negative control (no antibiotic added). Here, four IPTG concentrations have been used to trigger the LuxR production, spanning the operative working range of P_{LlacO1} promoter (unpublished data): 0 M, 10 μM, 25 μM, 500 μM. For each IPTG concentration, different HSL inductions have been performed, in order to obtain four HSL induction curves, one for each tested IPTG induction (corresponding to a LuxR level).
After 2 hours from the 1:500 dilution, 1 μl of properly diluted IPTG and 1 μl of properly diluted HSL (Sigma Aldrich, #K3007) have been added to the cultures (1 ml). Cultures were further incubated at 37 °C 220 rpm for 45 minutes before being transferred in the microplate reader.

D.1.4 Parameters estimation from experimental data

Given an experimental data set, composed by four induction curves, one for each considered copy number, parameter estimation procedures have been implemented as follows. All the dynamic parameters (with the exception of k_{+5}, set at 0) were simultaneously estimated from all the induction curves. N values were set according to the copy numbers estimated from experimental data. The initial amount of X_0 was simultaneously estimated from data. When the copy number of the transcription factor was supposed to variate (condition described in Fig. D.1 B), X_0 was multiplied by N, while in experimental condition of Fig. D.1 A, also P_{F_0} was multiplied by N.
Experimental data were fitted using the nonlinear least squares approach, by the Levenberg-Marquardt optimization iterative algorithm as implemented in the Matlab (MathWorks, Naitck, MA) *lsqnonlin* routine (optimization toolbox). Each residual was weighted proportionally to the inverse of the related measurement.

D.2 Supplementary results

D.2.1 Analytical study of the simple system: $X+H \leftrightarrow D$

In order to investigate the analytical behaviour of a simple portion of the model, a single reaction system has been considered.

$$X + H \underset{k_{-1}}{\overset{k_{+1}}{\longleftrightarrow}} D \tag{D.8}$$

Here, the two species X and H can react to form D complex and the system dynamics are governed by k_{+1} and k_{-1} kinetic rate constants. An ODE model has been derived by the reaction described in Eq. D.8. This model has been simulated with the parameters and initial conditions reported in Tab. 5.1 and its behavior is shown in Fig. D.2 A in response to three different initial amounts of H_0.

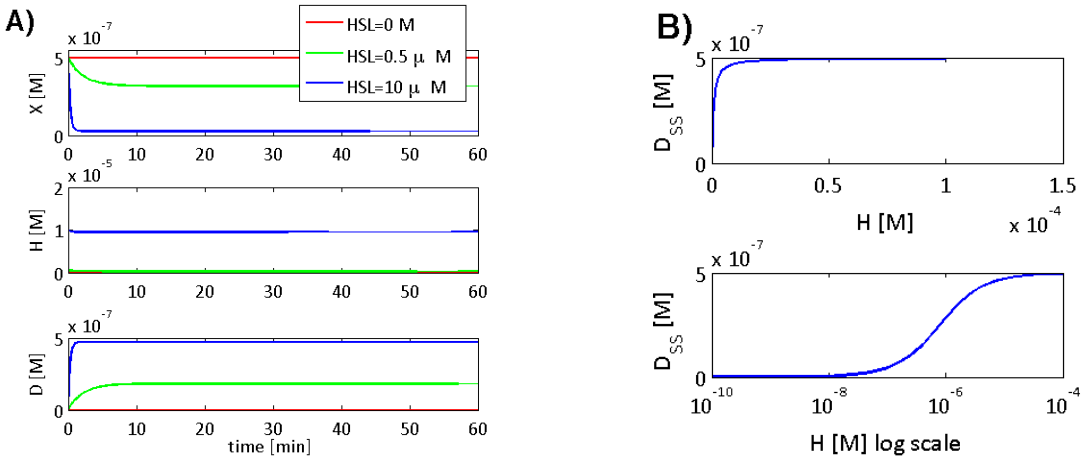


Figure D.2: **Dynamic and induction curve of the simplified system** $X + H \leftrightarrow D$ In panel A, the three plots show the behaviour of the three species X, H and D. When no inducer is given in input (H_0=0, red lines), the system remains in the trivial equilibrium state and the species do not deviate from their initial values. For similar amounts of inducer ($H_0 = 0.5$ μM) and X_0, X is consumed to form the D complex and the steady state is reached in $\sim$ 15 minutes. When the inducer is in excess ($H_0 = 10$ μM), the totality of X is consumed for the production of D and the steady state is reached faster (<2 minutes). In panel B, the top graph represents the static H_0-D_{SS} characteristic described by Eq. D.9 in natural X-scale. A Michaelis-Menten like behaviour can be observed. The bottom graph represents the same curve in logarithmic X-scale. These simulations were performed with the following parameter and initial condition values: N=40 for P_F, $X_0 = 0.5 \cdot 10^{-6}$M, $k_{+1} = 6.02 \cdot 10^3 M^{-1}s^{-1}$, $k_{-1} = 3.3 \cdot 10^{-3}s^{-1}$.

If no H is provided as input ($H_0 = 0$), no dimer D is produced. For higher amounts of H (H_0=0.5 μM and H_0=10 μM), the amount of produced D increases. A saturation in the amount of D occurs when all the X present has been consumed and no more is available to bind H.

The steady state analytical solution for this system is:

$$\overline{D} = \frac{X_0 + H_0 + \frac{k_{-1}}{k+1} - \sqrt{(X_0 + H_0 + \frac{k_{-1}}{k+1})^2 - 4 \cdot X_0 \cdot H_0}}{2} \qquad \text{(D.9)}$$

The static induction curve of this system is reported in Fig. D.2 B and presents a Michaelis-Mentel like characteristic. The induction curve is represented both in natural (Fig. D.2B, top panel) and logarithmic (Fig. D.2 B, bottom panel) X-scale.

When $H_0 \ll X_0$ (first portion of the curve), the steady state solution can

be approximated by the relation:

$$D_{SS} = \frac{X_{SS} \cdot H_0}{X_{SS} + \frac{1}{K_{eq,1}}}, \qquad X_{SS} \simeq X_0 \tag{D.10}$$

This approximation is represented in Fig. D.3 (red dots). When the condition $H_0 \gg X_0$ is satisfied (last portion of the curve), the approximated solution is:

$$D_{SS} = \frac{X_0 \cdot H_{SS}}{H_{SS} + \frac{1}{K_{eq,1}}}, \qquad H_{SS} \simeq H_0 \tag{D.11}$$

This approximated solution is represented in Fig. D.3 (green dots).

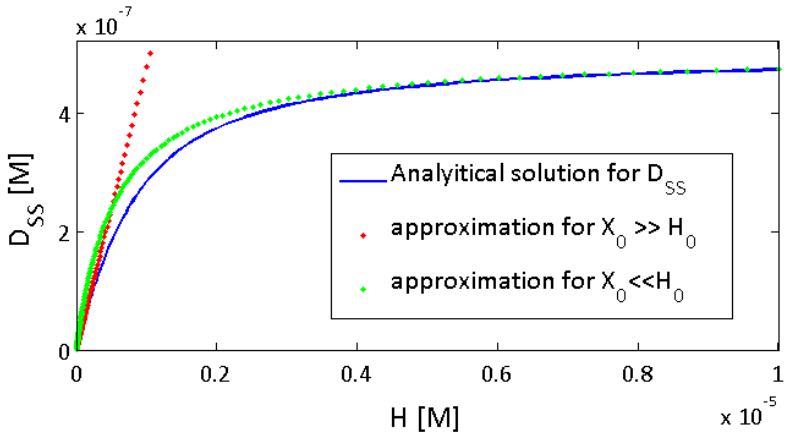


Figure D.3: **Induction curve of $X + H \leftrightarrow D$ system and approximation for extreme H_0 values.** The induction curve of the simplified system (blue line) is well approximated, in its first portion ($H_0 \ll X_0$), by the relation described in Eq. D.10 (red dots). The saturated portion ($H_0 \gg X_0$) is well approximated by Eq. D.11 (green dots). The intermediate portion of the curve cannot be approximated by the simplified relation and is described by the analytical solution reported in Eq. D.9. These simulations were performed with the following parameter and initial condition values: N=40 for P_F, $X_0 = 0.5 \cdot 10^{-6}$M, $k_{+1} = 6.02 \cdot 10^3 M^{-1} s^{-1}$, $k_{-1} = 3.3 \cdot 10^{-3} s^{-1}$.

Furthermore, if the condition $Min\{X_0, H_0\} \gg \frac{1}{K_{eq,1}}$ were satisfied, the approximated solution of the system would be:

$$D_{SS} = \begin{cases} H_0 & \text{for } H_0 \ll X_0 \\ X_0 & \text{for } H_0 \gg X_0 \end{cases} \tag{D.12}$$

A linear relation occurs between D_{SS} and H_0 when $H_0 < X_0$. For $H_0 \geq X_0$, D_{SS} remains constant at X_0, as shown in Fig. D.4.

According to the logarithmic representation of the induction curve, meaningful parameters (i.e.: the induction needed to reach 50% of the maximum output) should be derived as a function of initial amounts of X_0 and H_0. Unfortunately, even for this simple system an analytical dissertation in not feasible. In even more complex systems (such as the one object of this study), numerical methods need to be applied.

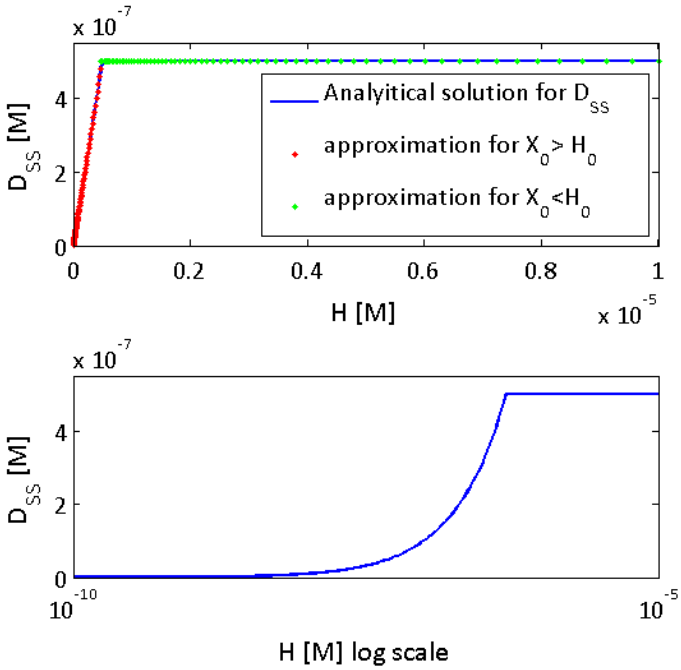


Figure D.4: **Induction curve of simplified system when the condition $Min\{X_0, H_0\} \gg \frac{1}{K_{eq,1}}$ is satisfied.** In this simulation, k_{+1} value was set to 10^9, while k_{-1} value was not varied. The top panel shows the induction curve in natural X-scale. The system behaviour computed by the analytical formula in Eq. D.9 is perfectly approximated by Eq. D.12. A sharp variation is observed around the critical value $X_0 = H_0$. The bottom panel shows the curve in logarithmic X- scale. Here, the sharp transition between the exponential growth and the linear saturation can be observed.

D.2.2 Sensitivity analysis

As already stated, the parameter values used to simulate the model were estimated by *in vitro* experiments, thus they may not reflect the actual kinetic values observable within the cell environment. The resulting induction curve does not match the experimental data and a fine tuning of the kinetic parameters of the system is required. For this reason, numerical sensitivity analysis has been performed, by varying parameters within a biologically plausible range. All k_{+x} parameters have been varied, in order to change the equilibrium constant K_{eq_x}. The same results were obtained by varying k_{-x} in accordance. Here, only the results relative to k_{+x} variation are reported, since the others are specular.

Sensitivity analysis on model parameter $\mathbf{k_{+1}}$

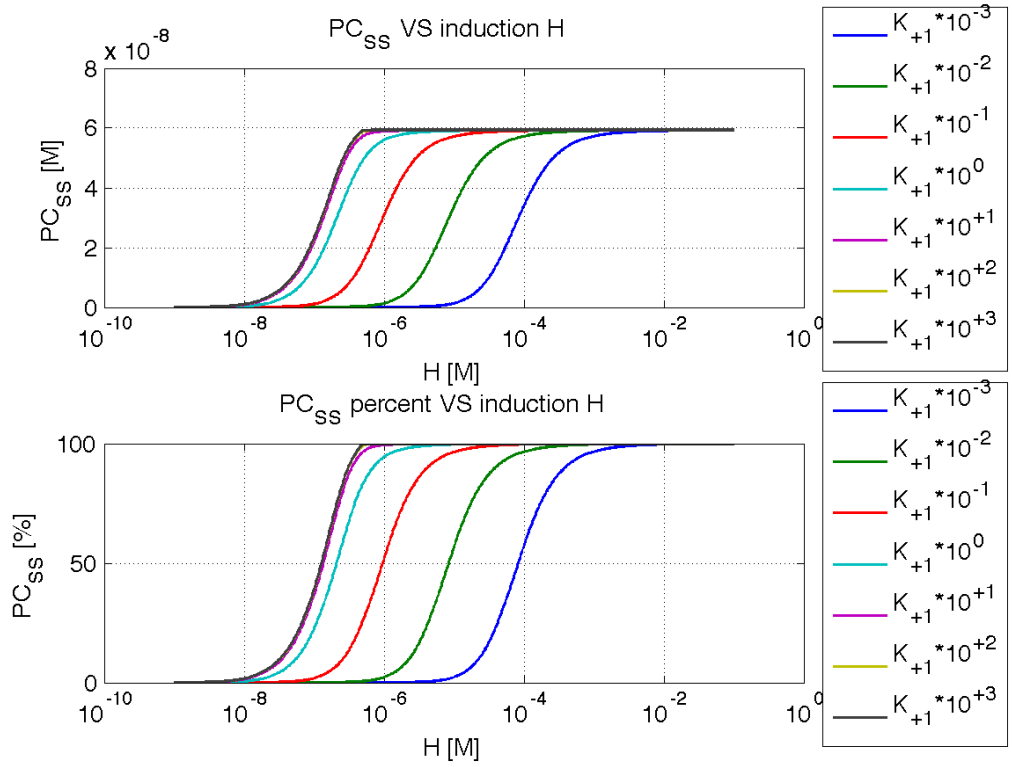


Figure D.5: **Induction curves for lux system varying k_{+1} parameter.** For increasing k_{+1} values, a change in the system output is observed: the curves result shifted to the left until a critical value is reached, above which all the curves become overlapped. No differences are observed in terms of maximum activity (i.e.: the activity at full induction). The graph above illustrates the induction curves in absolute units, while the graph below shows the normalized curves with respect to their maximum. The X-axis of both graphs is in log scale.

k_{+1} parameter has been increased and decreased of three orders of magni-

tude, while all the other parameters have been maintained at their nominal values (as reported in Tab. 5.1). The resulting induction curves are shown in Fig. D.5. No significant effect on the maximum transcription activity (P_{max}) is observed for the different curves (Fig. D.6 panel B), while the induction required to reach 50% of the maximum (H_{50}) decreases of three orders of magnitude (from $\sim 10^{-4}$ M to $\sim 10^{-7}$ M) for increasing k_{+1} values (Fig. D.6 panel A). In particular, the relation between H_{50} and k_{+1} is linear for values of k_{+1} less than $6.02 \cdot 10^3$ $M^{-1}s^{-1}$ (which is the nominal value), and shows a saturation for higher values. H* also decreases for increasing parameter values, with the shape of a decreasing sigmoidal function (Fig. D.6 panel C). These results suggest that moving the equilibrium rate constant of the first reaction does not affect the system behaviour in terms of output entity at full induction, but dramatically affects the molecule affinities in C complex formation. If the first reaction were slower than initially assumed (k_{+1} lower than nominal value), higher amounts of H would be required to occupy the whole available free promoter P_F, but at full induction the system steady state would be identical. For higher values of k_{+1}, the output change is gradually less and less evident, as no difference is observed.

Sensitivity analysis on model parameter $\mathbf{k_{+2}}$

The sensitivity analysis has been performed on k_{+2} parameter and the resulting induction curves are shown in Fig. D.7. These curves result highly affected by k_{+2} variation, both in the maximum entity and the horizontal shift. In particular, at lower k_{+2} values correspond lower P_{max} (until $\sim 4 \cdot 10^{-9}$ M), while for k_{+2} greater than or equal to the nominal k_{+2} value a saturation occurs at $P_{max} \sim 6 \cdot 10^{-8}$ M (15-fold variation of P_{max}). In Fig. D.8 panel B this trend is shown. H_{50} value shows a decreasing trend for increasing k_{+2} (Fig. D.8 panel A), which is consistent with the trend observed for k_{+1} variation, even if in this case the variation entity of H_{50} is reduced to $\sim$ one decade (from 10^{-6} M to 10^{-7} M). H* also decreases (reducing to the half) for increasing parameter values (Fig. D.8 panel C).

Considering the effect of k_{+1} and k_{+2} variation, it can be observed that the variation of a parameter regulating a reaction 'distant' from the output of the system (such as the variation of k_{+1}) has a more pronounced effect on the H_{50} parameter, while the variation of parameters which regulate reactions more 'proximal' to the output mostly affects the P_{max} value. The variation of both k_{+1} and k_{+2} parameters in concert results in a combined effect (data not shown).

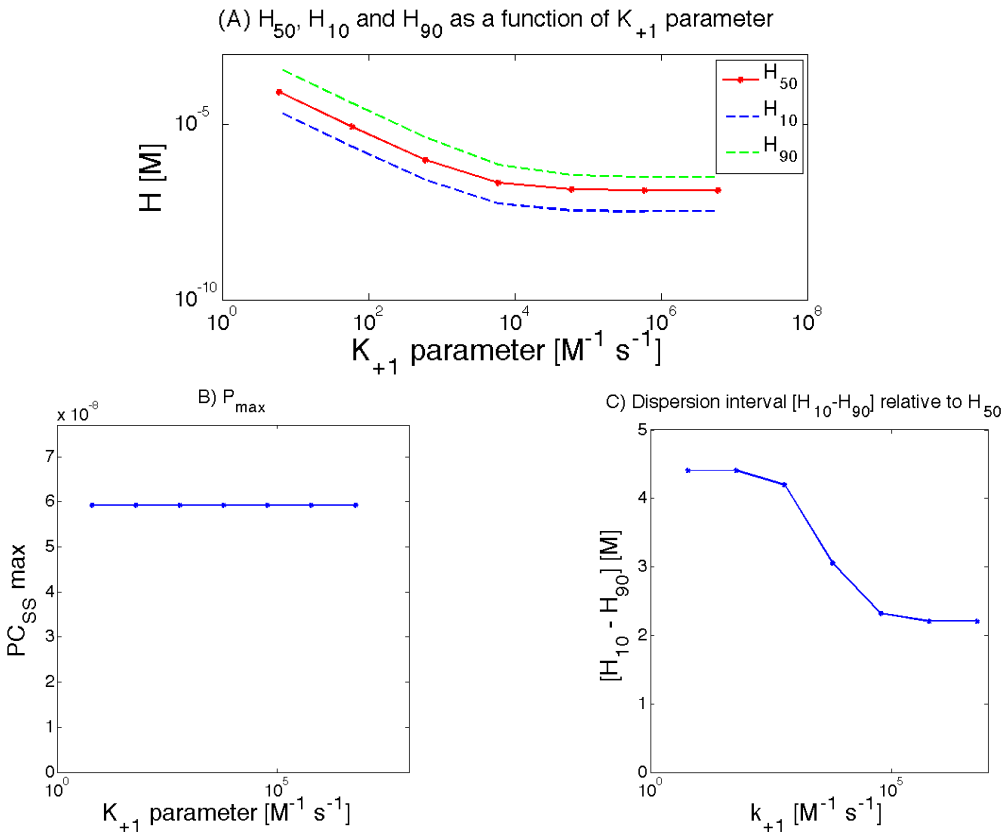


Figure D.6: **Quantitative parameters to assess the effect of k_{+1} variation on system output.** In panel A H_{50} (red line), H_{10} and H_{90} (blue and green dotted line, respectively) values are shown. Here, all the parameters show a monotonic decreasing behaviour for increasing k_{+1} values, which is almost linear below the nominal value ($6.02{\cdot}10^3 M^{-1}s^{-1}$) and does not significantly change for higher values. X-axis and Y-axis are represented in log scale. Panel B shows the P_{max} behaviour, which is not affected by k_{+1} variation. X-axis is represented in log-scale. Panel C shows H* behaviour. A sigmoidal decreasing behaviour is observed. This parameter highlights how broader respect to previous values is the range of working inductions (causing a significant change on system output from 10% to 90% of the maximum) in terms of spanned decades around the H_{50} for lower k_{+1} values, reducing to the half for higher k_{+1} values. The X-axis is represented in log-scale.

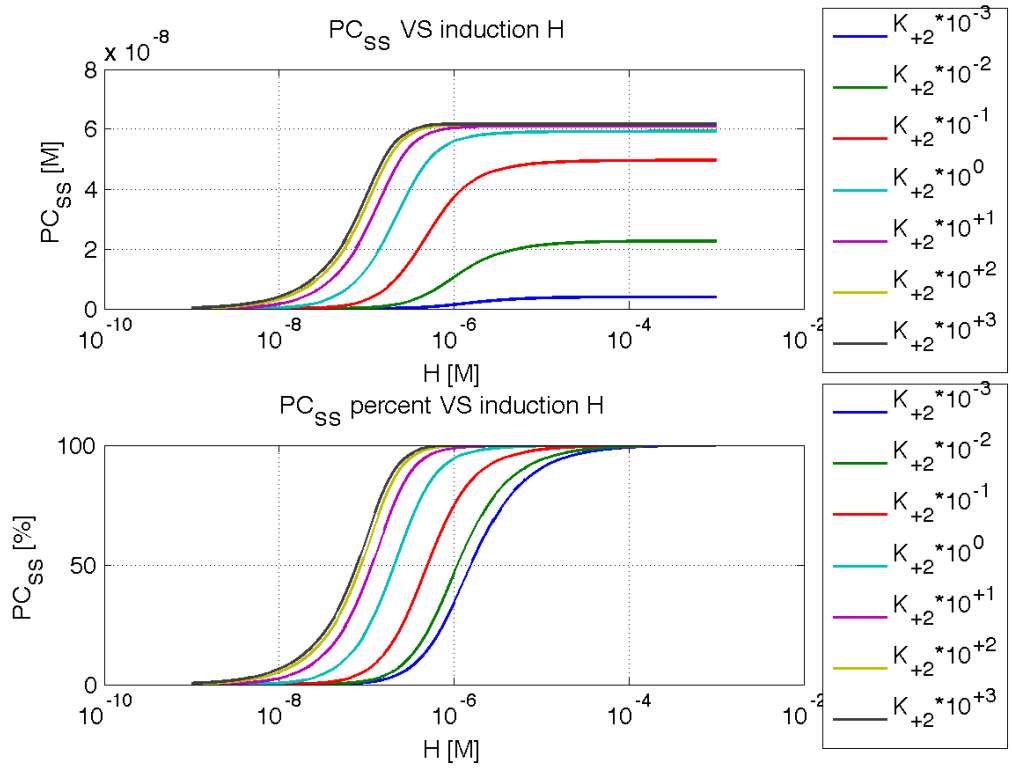


Figure D.7: **Induction curves for lux system varying k_{+2} parameter.** For increasing k_{+2} values, a change in the system output is observed: the curves result shifted to the left until a critical value is reached, above which all the curves became overlapped. A marked difference is also observed in terms of maximum activity (i.e.: the activity at full induction), which significantly increases for increasing k_{+2} values, since a saturation phase is reached. The graph above illustrates the induction curves in absolute units, while the graph below shows the normalized curves with respect to their maximum, in order to better appreciate the curve shift. The X-axis of both graphs is in log scale.

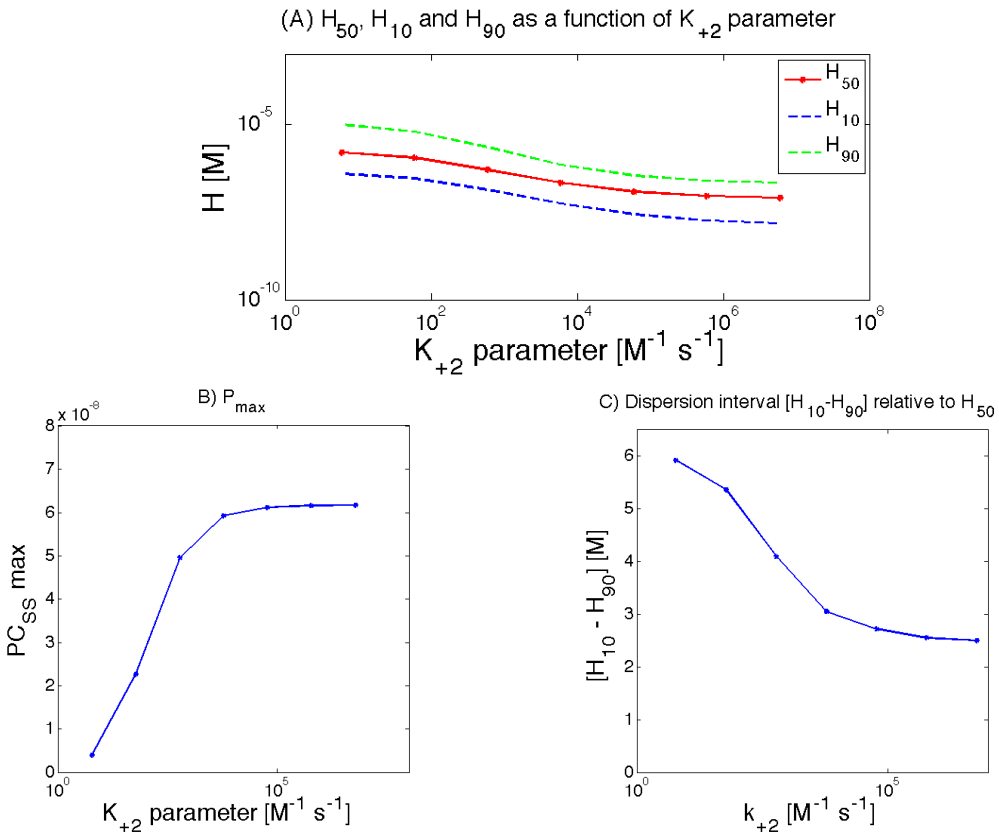


Figure D.8: **Quantitative parameters to assess the effect of k_{+2} variation on system output.** In panel A H_{50} (red line), H_{10} and H_{90} (blue and green dotted line, respectively) values are shown. Here, all the parameters show a monotonic decreasing behaviour for increasing k_{+2} values. The relation has a reverse sigmoidal shape around the nominal value ($6.02{\cdot}10^3 M^{-1}s^{-1}$). X-axis and Y-axis are represented in log scale. Panel B shows the P_{max} behaviour, which is highly affected by k_{+2} variation. P_{max} increases until a saturation occurs, for k_{+2} values higher than the nominal one. X-axis is represented in log-scale. Panel C shows H* behaviour. A decreasing trend is observed. This parameter highlights how broader the range of working inductions (causing a significant change on system output from 10% to 90% of the maximum) is in terms of spanned decades around the H_{50} for lower k_{+2} values. The X-axis is represented in log-scale.

Sensitivity analysis on model parameter $\mathbf{k}_{+3}$

The variation of k_{+3} parameter results in a change in the binding affinity between free promoter P_F and the activated complex C. The resulting induction curves (Fig. D.9) show significant differences both in terms of maximum entity and H_{50}. For increasing k_{+3} values, P_{max} shows a 60-fold change, from $\sim 1.1 \cdot 10^{-9}$ to $6.4 \cdot 10^{-8}$ M (Fig. D.8 panel B). The saturation value of $6.4 \cdot 10^{-8}$ M corresponds to the total available promoter (P_{F_0}).
H_{50} value varies from 10^{-6} M to 10^{-7} M with a reverse sigmoidal behaviour (Fig. D.8 panel A). The H_{10} and H_{90} values are not symmetrically distributed around the H_{50} value: induction curves corresponding to the higher k_{+3} values, indeed, show a smoother behaviour in their first tract, while a sharper trend is observable in the last portion of the curves.
H^* is >2-fold reduced (Fig. D.8 panel C), resulting in sharper curves for higher k_{+3} values. For these curves, the working induction range that allows the regulation of promoter activity from 10% to 90% of the maximum covers almost one induction decade, while the curves obtained with lower k_{+3} values have a regulation range of at least two induction decades.
It is important to underline that every H_0 value used to simulate the model corresponds to a physiologically tolerable induction for *E. coli* cell.

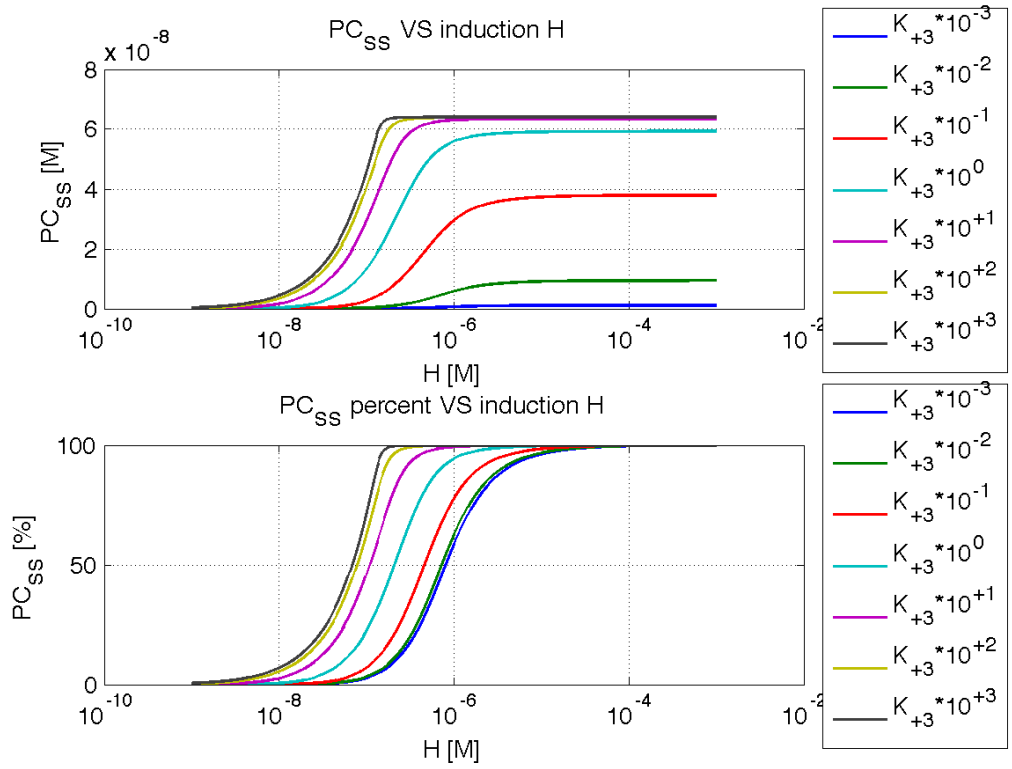


Figure D.9: **Induction curves for lux system varying k_{+3} parameter.** k_{+3} variation provokes important changes in the system output. Induction curves are reduced in their maximum entity and shifted to the right for lower parameter values. While for k_{+3} values deviating of one order of magnitude from the nominal value the resulting induction curves are remarkably different, for major variations they result overlapping with the two most extreme conditions. The graph above illustrates the induction curves in absolute units, while the graph below shows the normalized curves compared to their maximum. The X-axis of both graphs is in log scale.

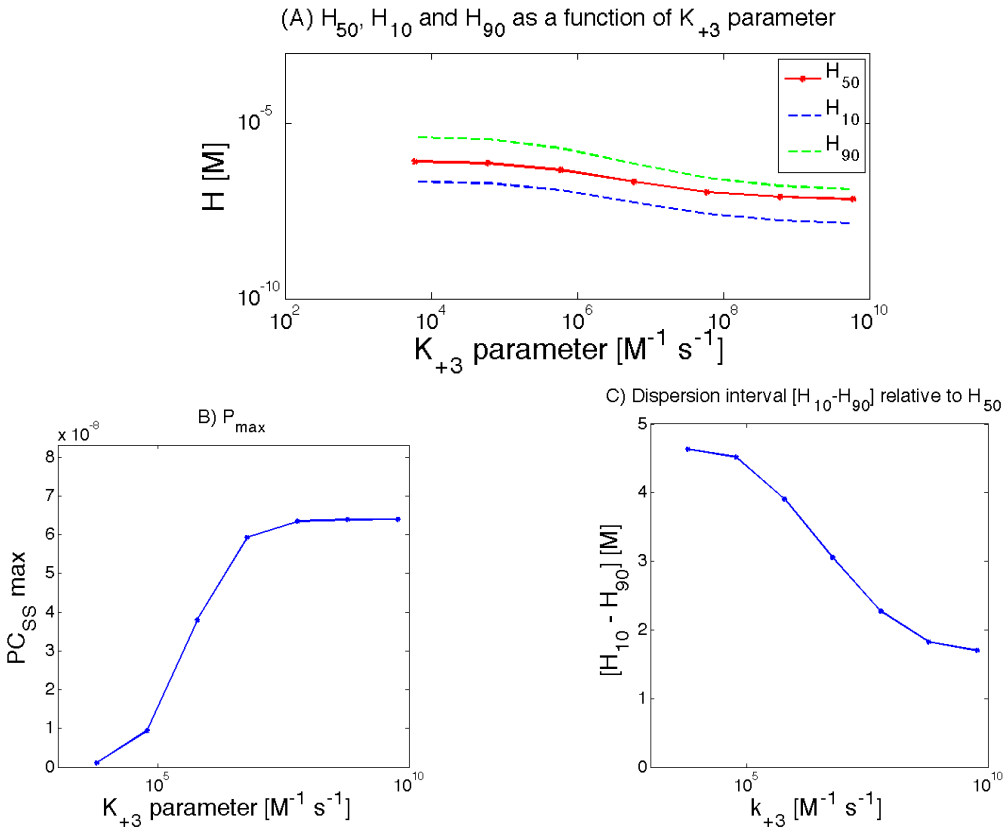


Figure D.10: **Quantitative parameters to assess the effect of k_{+3} variation on system output.** In panel A H_{50} (red line), H_{10} and H_{90} (blue and green dotted line, respectively) values are shown. All the parameters show a monotonically decreasing behaviour for increasing k_{+2} values. The relation has a reverse sigmoidal shape around the nominal value ($6{\cdot}10^6 M^{-1}s^{-1}$). The H_{10} and H_{90} values are not symmetrically distributed around the H_{50} value for the highest k_{+3} values. X-axis and Y-axis are represented in log scale. Panel B shows the maximum activity as a function of k_{+3} values. A huge variation ($\sim$ 60-fold) in P_{max} value is highlighted and a saturation trend is present for the higher parameter values. X-axis is represented in log-scale. Panel C shows H* behaviour. This value >2-fold decreases for higher parameter values and a reverse sigmoidal shape is observed. The X-axis is represented in log-scale.

Bibliography

[1] 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.

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

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

[4] R. Kitney and P. Freemont. Synthetic biology–the state of play. *FEBS letters*, 2012.

[5] C. Bernasconi, S. Garagna, F. Milano, M. Zuccotti, and C. A. Redi. Le sfide della biologia sintetica e la fine del naturale. Cellule e genomi - X corso. Ibis, 2012.

[6] C. M. Agapakis and P. A. Silver. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*, 5(7):704–713, 2009.

[7] L. Gold. Post-transcriptional regulatory mechanisms in *Escherichia coli*. *Annual Review of Biochemistry*, 57(1):199–233, 1988.

[8] B. Canton. Engineering the interface between cellular chassis and synthetic biological systems. PhD thesis, Department of Biological Engineering, MIT, 2008.

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

[10] J. R. Kelly, A. J. Rubin, J. Davis, C. Ajo-Franklin, J. Cumbers, M. Czar, K. de Mora, A. Glieberman, D. Monie, and D. Endy. Measuring the

activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*, 3(1):4, 2009.

[11] T. F. Knight. Idempotent vector design for standard assembly of BioBricks. Technical report, DTIC Document, 2003.

[12] A. S. Khalil and J. J. Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379, 2010.

[13] D. Rebatchouk, N. Daraselia, and J.O. Narita. NOMAD: a versatile strategy for in vitro DNA manipulation applied to promoter analysis and vector design. *PNAS*, 93(20):10891–10896, 1996.

[14] A.P. Arkin and D. Endy. A standard parts list for biological circuitry. Technical report, MIT, 1999.

[15] D. Ferber. Microbes made to order. *Science*, 303(5655):158–161, 2004.

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

[17] M.A. Speer and T.L. Richard. Amplified insert assembly: an optimized approach to standard assembly of BioBrick™ genetic circuits. *Journal of Biological Engineering*, 5(1):1–11, 2011.

[18] iGEM main page. [http://www.igem.org/Main_Page].

[19] The BioBricks Foundation, [http://biobricks.org/].

[20] Registry Of Standard Biological Parts, [http://partsregistry.org].

[21] R. P. Shetty, M. Lizarazo, R. Rettberg, and T. F. Knight. Assembly of BioBrick standard biological parts using three antibiotic assembly. *Methods in Enzymology*, 498:311–326, 2011.

[22] S.N. Cohen, A.C.Y. Chang, H.W. Boyer, and R.B. Helling. Construction of biologically functional bacterial plasmids in vitro. *PNAS*, 70(11):3240–3244, 1973.

[23] J. C. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, and J. D. Keasling. BglBricks: A flexible standard for biological part assembly. *Journal of Biological Engineering*, 4(1):1–12, 2010.

[24] A. P. Arkin et al. Setting the standard in synthetic biology. *Nature Biotechnology*, 26(7):771–773, 2008.

[25] C.E. French, K. de Mora, N. Joshi, A Elfick, J. Haseloff, and J. Ajioka. Synthetic biology and the art of biosensor design. In *The Science and Applications of Synthetic and Systems Biology: Workshop Summary*, page 178. National Academy Press, 2011.

[26] S. Fields and O.U. Song. A novel genetic system to detect protein–protein interactions. *Nature*, 340:245–246, July 1989.

[27] J. C. Anderson, C. A. Voigt, and A. P. Arkin. Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, 3(1), 2007.

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

[29] L. Martin, A. Che, and D. Endy. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS ONE*, 4(11):e7569, 2009.

[30] 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, 2011.

[31] 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.

[32] 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.

[33] G. Rodrigo, J. Carrera, T.E. Landrain, and A. Jaramillo. Perspectives on the automatic design of regulatory systems for synthetic biology. *FEBS letters*, 2012.

[34] J. M. Pedraza and A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307(5717):1965–1969, March 2005.

[35] K. Biliouris, P. Daoutidis, and Y. N. Kaznessis. Stochastic simulations of the tetracycline operon. *BMC Systems Biology*, 5(9), 2011.

[36] Y. N. Kaznessis. Models for synthetic biology. *BMC Systems Biology*, 1(1), 2007.

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

[38] A. E. Mayo, Y Setty, S Shavit, A Zaslaver, and U Alon. Plasticity of the cis-Regulatory Input Function of a Gene. *PLoS Biology*, 4(4), 2006.

[39] 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, 2008.

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

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

[42] S. Basu, Y. Gerchman, C.H. Collins, F.H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130–1134, 2005.

[43] M. Stamatakis and N. V. Mantzaris. Comparison of Deterministic and Stochastic Models of the lac Operon Genetic Network. *Biophysical Jurnal*, 96(3):887–906, February 2009.

[44] L. Pasotti, S. Zucca, M. Lupotto, M. G. Cusella De Angelis, and P. Magni. Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *Journal of Biological Engineering*, 5:8–8, 2011.

[45] R. Young, I.N. Wang, and W.D. Roof. Phages will out: strategies of host cell lysis. *Trends in microbiology*, 8(3):120–128, 2000.

[46] 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* Under revision, 2012.

[47] G. Lipps. Plasmids: current research and future trends. Caister Academic Pr, 2008.

[48] L. Serrano. Synthetic biology: promises and challenges. *Molecular systems biology*, 3(1), 2007.

[49] X.C. Rao, S. Li, J.C. Hu, X.L. Jin, X.M. Hu, J.J. Huang, Z.J. Chen, J.M. Zhu, and F.Q. Hu. A novel carrier molecule for high-level expression of peptide antibiotics in *Escherichia coli*. *Protein Expression and purification*, 36(1):11–18, 2004.

[50] S. Goh and L. Good. Plasmid selection in *Escherichia coli* using an endogenous essential gene marker. *BMC Biotechnology*, 8(1):61, 2008.

[51] J. Kroll, S. Klinter, C. Schneider, I. Voß, and A. Steinbüchel. Plasmid addiction systems: perspectives and applications in biotechnology. *Microbial Biotechnology*, 3(6):634–657, 2010.

[52] D. Boyd, D. S. Weiss, J. C. Chen, and J. Beckwith. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *Journal of Bacteriology*, 182(3):842–847, Feb 2000.

[53] F. Martinez-Morales, AC Borges, A. Martinez, KT Shanmugam, and LO Ingram. Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction. *Journal of Bacteriology*, 181(22):7143–7148, 1999.

[54] G. Posfai, M. D. Koob, H. A. Kirkpatrick, and F. R. Blattner. Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the *Escherichia coli* O157: H7 genome. *Journal of Bacteriology*, 179(13):4426–4428, 1997.

[55] S. Datta, N. Costantino, et al. A set of recombineering plasmids for gram-negative bacteria. *Gene*, 379:109–115, 2006.

[56] K. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS*, 97(12):6640–6645, 2000.

[57] H.M. Ellis, D. Yu, T. DiTizio, et al. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *PNAS*, 98(12):6742–6746, 2001.

[58] A. M. Campbell. Chromosomal insertion sites for phages and plasmids. *Journal of Bacteriology*, 174(23):7495, 1992.

[59] A. Haldimann and B.L. Wanner. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *Journal of Bacteriology*, 183(21):6384, 2001.

[60] P.P. Cherepanov and W. Wackernagel. Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, 158(1):9–14, 1995.

[61] D.H.S. Block, R. Hussein, L.W. Liang, and H.N. Lim. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Research*, 2012.

[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] D.J. Lee, L.E.H. Bingle, K. Heurlier, M.J. Pallen, C.W. Penn, S.J.W. Busby, and J.L. Hobman. Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiology*, 9(1):252, 2009.

[64] L. P. Yomano, S. W. York, S. Zhou, K. T. Shanmugam, and L. O. Ingram. Re-engineering *Escherichia coli* for ethanol production. *Biotechnology letters*, 30(12):2097–2103, 2008.

[65] P.C. Turner, L.P. Yomano, L.R. Jarboe, S.W. York, C.L. Baggett, B.E. Moritz, E.B. Zentz, KT Shanmugam, and L.O. Ingram. Optical mapping and sequencing of the *Escherichia coli* KO11 genome reveal extensive chromosomal rearrangements, and multiple tandem copies of the *Zymomonas mobilis pdc* and *adhB* genes. *Journal of industrial microbiology & biotechnology*, pages 1–11, 2012.

[66] K.E.J. Tyo, P.K. Ajikumar, and G. Stephanopoulos. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology*, 27(8):760–765, 2009.

[67] J. H. Leveau and S. E. Lindow. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *Journal of Bacteriology*, 183(23):6752–6762, Dec 2001.

[68] M. Hajimorad, P. R. Gray, and J. D. Keasling. A framework and model system to investigate linear system behavior in *Escherichia coli*. *Journal of Biological Engineering*, 5:3–3, 2011.

[69] J. R. Kelly. Tools and reference standards supporting the engineering and evolution of synthetic biological systems. PhD thesis, Massachusetts Institute of Technology, 2008.

[70] J. E. Cabrera and D. J. Jin. Active transcription of rRNA operons is a driving force for the distribution of RNA polymerase in bacteria: effect of extrachromosomal copies of rrnB on the in vivo localization of RNA polymerase. *Journal of Bacteriology*, 188(11):4007–4014, Jun 2006.

[71] A. M. Stevens and E. P. Greenberg. Quorum sensing in *Vibrio fischeri*: essential elements for activation of the luminescence genes. *Journal of Bacteriology*, 179(2):557–562, Jan 1997.

[72] 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.

[73] J. T. Kittleson, S. Cheung, and J. C. Anderson. Rapid optimization of gene dosage in *E. coli* using DIAL strains. *Journal of Biological Engineering*, 5(10), July 2011.

[74] 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):1–20, 2010.

[75] M. Fussenegger. Synchronized bacterial clocks. *Nature*, 463(7279):301–302, 2010.

[76] A. Tamsir, J.J. Tabor, and C.A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical wires. *Nature*, 469(7329):212–215, 2010.

[77] P.S. Daugherty, M.J. Olsen, B.L. Iverson, and G. Georgiou. Development of an optimized expression system for the screening of antibody libraries displayed on the *Escherichia coli* surface. *Protein Engineering*, 12(7):613–621, 1999.

[78] D. Dubnau and R. Losick. Bistability in bacteria. *Molecular Microbiology*, 61(3):564–572, 2006.

[79] L. Endler, N. Rodriguez, N. Juty, V. Chelliah, C. Laibe, C. Li, and N. Le Novère. Designing and encoding models for synthetic biology. *Journal of The Royal Society Interface*, 6(Suppl 4):S405–S417, 2009.

[80] J. Rohwer, A.J. Hanekom, and J. H. Hofmeyr. A universal rate equation for systems biology. *Proc. 2nd Int. Symp. on Experimental Standard Conditions of Enzyme Characterizations*, pages 175–187, 2007 (ESEC 2006), Beilstein Institute, Frankfurt am Main, Germany.

[81] A. B. Goryachev, D. J. Toh, and T. Lee. Systems analysis of a quorum sensing network: design constraints imposed by the functional requirements, network topology and kinetic constants. *Biosystems*, 83(2):178–187, 2006.

[82] T.S. Gardner and J.J. Faith. Reverse-engineering transcription control networks. *Physics of life reviews*, 2(1):65–88, 2005.

[83] I.H. Segel. Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems. Wiley New York, 1993.

[84] M. L. Urbanowski, C. P. Lostroh, and E. P. Greenberg. Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein. *Journal of Bacteriology*, 186(3):631–637, 2004.

[85] H.H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269(5224):650, 1995.

[86] J. Peccoud and M. Isalan. The PLOS ONE Synthetic Biology Collection: Six Years and Counting. *PLoS ONE*, 7(8):e43231, 2012.

[87] J. Estrada and R. Guantes. Dynamic and structural constraints in signal propagation by regulatory networks. *Molecular BioSystems*, 2013.

[88] Y. Mileyko, R. Joh, and J. Weitz. Small-scale copy number variation and large-scale changes in gene expression. *PNAS*, 105(43):16659–16664, 2008.

[89] C. Fuqua, M.R. Parsek, and E.P. Greenberg. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. *Annual Review of Genetics*, 35(1):439–468, 2001.

[90] S. Basu, R. Mehreja, S. Thiberge, M.T. Chen, and R. Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *PNAS*, 101(17):6355–6360, 2004.

[91] L. Pasotti, M. Quattrocelli, D. Galli, M. G. Cusella De Angelis, and P. Magni. Multiplexing and demultiplexing logic functions for computing signal processing tasks in synthetic biology. *Biotechnology Journal*, 6(7):784–795, July 2011.

[92] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, et al. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462, 1997.

[93] T. Durfee, R. Nelson, S. Baldwin, G. Plunkett, V. Burland, B. Mau, J. F. Petrosino, X. Qin, D. M. Muzny, M. Ayele, et al. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *Journal of Bacteriology*, 190(7):2597–2606, 2008.

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

List of publications

Articles in peer reviewed journals

- 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* Accepted after revision, 2012.

- 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.

- 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.

- L. Pasotti, S. Zucca, M. Lupotto, M. G. Cusella De Angelis and P. Magni. Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *Journal of Biological Engineering*, 5:8, 2011.

Book chapters

- P. Magni, S. Zucca, L. Pasotti, Biologia sintetica e standardizzazione: verso la programmazione razionale di funzioni biologiche artificiali, *In* L. Marini, A. Carlino (eds) Il post-umano e l'etica del nuovo. Dal corpo bionico al corpo sintetico, *Carocci Ed.* 2011 ISBN: 9788843065660

Contributions to conference proceedings

- S. Zucca, L. Pasotti, M. G. Cusella De Angelis and P. Magni, Characterization of promoter strength at different DNA copy numbers using RPUs.

Congresso Nazionale di Bioingegneria - Atti. Roma, Italy, June 26-29, 2012, ISBN/ISSN: 978 88 555 3182-5

- 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, Italy, June 26-29, 2012, ISBN/ISSN: 978 88 555 3182-5

- N. Politi, L. Pasotti, S. Zucca, M. G. Cusella De Angelis and P. Magni, Modellizzazione e implementazione di un circuito di regolazione con feedback negativo in *Escherichia coli. Congresso Nazionale di Bioingegneria - Atti.* Roma, Italy, June 26-29, 2012, ISBN/ISSN: 978 88 555 3182-5

- 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, Holland, October 1-2, 2011, Cambridge:The iGEM Foundation

- S. Zucca, L. Pasotti, M.G. Cusella De Angelis and P. Magni, Characterization of promoter strength at different DNA copy numbers using RPU approach In: F. Geraci, R. Marangoni, M. Pellegrini and M.E. Renda. *Proceeding of BITS 2011.* Pisa, Italy, June 20-22, 2011, p. 134-135, PISA:Edizioni Plus, ISBN: 9788846730695

- G. Zambianchi, L. Pasotti, S. Zucca, M.G. Cusella De Angelis and P. Magni, Realization of a double knob in *Escherichia coli* to optimize fermentation of lactose of dairy whey into ethanol. In: F. Geraci, R. Marangoni, M. Pellegrini and M.E. Renda. *Proceeding of BITS 2011.* Pisa, Italy, June 20-22, 2011, p. 132-134, PISA:Edizioni Plus, ISBN: 9788846730695

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

- L. Pasotti, S. Zucca, M. G. Cusella De Angelis and P. Magni, Sharable quantitative measurements of biological parts using the Relative Promoter Units approach. *International Conference on Synthetic Biology (Genopole)*, Evry, France, December 15-16, 2010.

- 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, MA, November 6-8, 2010, Boston: MIT

- S. Zucca, L. Pasotti, D. Galli, M. G. Cusella De Angelis and P. Magni, Engineering an alcoholic fermentation pathway in *E. coli* using the Synthetic Biology approach. *Congresso Nazionale di Bioingegneria - Atti*. Torino, Italy, July 8-10, 2010, Bologna: Patron, p.165-166, ISBN/ISSN: 9788855530828

- L. Pasotti, S. Zucca, M. G. Cusella De Angelis and P. Magni, Standard measurements in Synthetic Biology: towards the rational design of novel biological functions. *Congresso Nazionale di Bioingegneria - Atti*. Torino, Italy, July 8-10, 2010, Bologna: Patron, p.127-128, ISBN/ISSN: 9788855530828

- L. Pasotti, S. Zucca, M. G. Cusella De Angelis and P. Magni, Validation of a standard approach for quantitative characterization of promoter strength, *Proceedings of Advances in Synthetic Biology 2010*, March 4-5, 2010, London, UK

- S. Zucca, L. Pasotti, M. Lupotto, M. G. Cusella De Angelis and P. Magni, Study of an inducible lysis device in *E. coli*, *Proceedings of Advances in Synthetic Biology 2010*, March 4-5, 2010, London, UK

- L. Pasotti, G. Chiesa, E. Del Fabbro, R. De Molfetta, L. Diamante, M. Lupotto, M. Meroso, S. Schiavi, A. Turcato, V. Vitelli, G. Zambianchi, S. Zucca, D. Galli, M. G. Cusella De Angelis and P. Magni, Ethanol? Whey not!, Proceedings of iGEM 2009. Boston, MA, October 31 - November 2, 2010, Boston: MIT

Ringraziamenti

Vorrei ringraziare il Prof. Paolo Magni per avermi dato la possibilità di incentrare le attività del mio dottorato di ricerca su un'area scientifica tanto promettente e stimolante quale è la Biologia Sintetica. Lo ringrazio in particolare per la la guida e il sostegno durante questi tre anni e per l'allegria con cui ha sempre accompagnato ogni attività lavorativa.
Vorrei anche ringraziare la Prof. Maria Gabriella Cusella De Angelis per aver messo a disposizione le strutture necessarie per le attività di laboratorio e per il suo sostegno indiscusso alle nostre ricerche.
Ringrazio Lorenzo Pasotti per essere stato un collega e un mentore in questi tre anni. Lo ringrazio per la pazienza con cui ha assistito e preso parte ad ogni aspetto della mia attività di ricerca. Ringrazio anche i miei "colleghi di laboratorio" Nicoló Politi e Michela Casanova per il supporto e l'aiuto che mi hanno dato.
Un sentito ringraziamento alla Dott. ssa Giuseppina Micoli della Fondazione Salvatore Maugeri per avere in questi anni messo a disposizione la sua strumentazione, il suo tempo e la sua preparazione per la definizione delle tecniche di misura utilizzate per portare avanti importanti linee di ricerca, anche se non sono state trattate in questo lavoro di tesi e alla Dott. Laura Benedetti, fonte inesauribile di preziose informazioni per sopravvivere alla vita di laboratorio.
Ringrazio tutti i ragazzi che hanno preso parte ai progetti iGEM nel 2009, 2010 e 2011 per gli insegnamenti che mi hanno lasciato e per la simpatia e l'amicizia con cui hanno accompagnato le interminabili estati in laboratorio, in particolare Giacomo, Matteo, Federica e Tommaso e un caro saluto a Giuseppe.
Un grosso grazie va anche ai miei "compagni di avventura" Nadia, Francesca e Matteo, con cui ho condiviso, in questi tre splendidi anni, le gioie e i dolori legati al nostro comune destino di dottorandi e a tutto il personale del laboratorio di Informatica Biomedica Mario Stefanelli, nonché ai professori che, prima da studente e poi da dottoranda, hanno avuto un ruolo fondamentale nella mia formazione, in particolare il Prof. Riccardo Bellazzi. Un caro saluto va al Prof. Mario Stefanelli, che ricordo con affetto e stima, per i preziosi consigli e insegnamenti di vita.

Il ringraziamento più sentito va sicuramente alla mia famiglia per avermi sempre sostenuta in qualsiasi scelta io abbia effettuato. Grazie a Bruno per esserci stato e avermi sempre capita, motivata e aiutata a non perdere di vista le cose importanti. Grazie a Nicole per la sua amicizia e la sua presenza costante. Grazie alle mie amiche di sempre Chiara e Monica, che hanno una sconfinata fiducia nelle mie capacità, che mi ha sempre motivata a migliorare. Grazie a tutte le persone che in questi anni hanno fatto parte della mia vita e che mi hanno permesso di affrontare serenamente e pienamente questi tre anni di attività di ricerca.