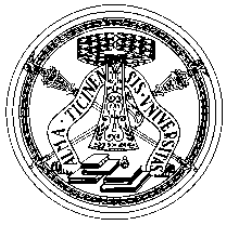
UNIVERSITA' DEGLI STUDI DI PAVIA

Dottorato di Ricerca in Bioingegneria e Bioinformatica

Dipartimento di Informatica e Sistemistica



Quantitative characterization of genetic parts and devices for the bottom-up engineering of living systems in synthetic biology

Tutor: Prof. Paolo Magni

Ph.D. program coordinator: Prof. Angelo Buizza

Ph.D. Thesis by

Lorenzo Pasotti

XXIV Ciclo (2008-2011)

Abstract (English)

Synthetic biology is an emerging area of bioengineering that aims at constructing novel biological functions to carry out specific user-defined tasks. The design rules are inspired by the engineering world: as an electronic circuit is constructed by connecting resistors, capacitors or diodes, a *genetic program* can be assembled by ligating DNA sequences, like genes, promoters, transcriptional terminators, ribosome binding sites or other regulatory elements. The ideal synthetic biology paradigm can be summarized as follows: i) choose biological parts from a library of well-characterized standard DNA components; ii) assemble them together to obtain a genetic program that encodes the desired function; iii) incorporate it in a living organism to complete the job.

As it happened in all the areas of engineering, physical standardization was introduced to facilitate the assembly of components by defining the concept of BioBrick, i.e. DNA parts with specific sequence and function that can be assembled through an easily reproducible process thanks to their common physical interface. The MIT Registry of Standard Biological Parts has been the first repository for standard components and it currently includes more than 16,000 BioBricks. However, in order to consider it as a real library of standard elements, the quantitative characterization of components is required. Although standard measurement approaches have also been recently proposed to quantify the characteristics of biological parts, the bottom-up construction of predictable biological systems is currently a major challenge. Incompatibility among components, context-dependent behaviour of parts, intrinsic noise of biological processes and nonlinearities in gene expression are some of the

limiting factors that contribute to the unpredictability of bottom-up-composed systems.

Synthetic biology could potentially yield applications of remarkable importance, like bioremediation or production of renewable fuels, new biomaterials and therapeutic molecules. However, its success depends on the definition of the working boundaries in which biological functions can be predictable, in order to enable the design of customized systems. This thesis focuses on quantitative characterization of biological parts and devices to learn their predictability boundaries, to provide useful data that support their re-use and finally to apply the synthetic biology concepts to face industrially-relevant problems in the field of bioenergy.

In Chapter 1 the current status of synthetic biology, its major achievements, promises and challenges are illustrated.

In Chapter 2 a modularity study on biological components *in vivo* is described: the activity of a representative set of promoters is quantified in disparate contexts to study how much it can change in increasingly complex *ad-hoc* constructed model systems.

In Chapter 3 the full quantitative characterization of a synthetic bacterial self-destruction device is reported. In the philosophy of providing datasheets for biological components, features like static/dynamic characteristics, compatibility and failure rate are measured.

In Chapter 4 the design of digital electronic-inspired logic functions (multiplexer and demultiplexer) is reported in a biological chassis. Mathematical models, identified by using experimental data, are used to describe the whole systems behaviour and to highlight the critical parameters for future finalization of the investigated functions.

In Chapter 5 synthetic biology principles are used to optimize the conversion of cheese whey, a food waste, into a biofuel through an engineered metabolic pathway. In particular, an ethanol-production device is incorporated in a lactose-utilizing microorganism to face two challenges at the same

time: energy production from a renewable source and disposal of a waste that is currently considered as an environmental problem.

In Chapter 6 steps towards the ambitious goal of cellulosic biomass-to-biofuel (or other compounds) conversion are described by providing quantitative characterization of cellulolytic enzymes expression and secretion in promising standard and non-standard chassis.

Finally, in Chapter 7 the overall conclusions of this thesis work are discussed.

The studies illustrated in Chapter 2, 3, 4 and 5 have been carried out in the molecular biology laboratories of the Centre for Tissue Engineering and in the Laboratory for Biomedical Informatics 'Mario Stefanelli', University of Pavia, Italy, while the study reported in Chapter 6 has been done at the Institute of Structural and Molecular Biology (School of Biological Sciences), University of Edinburgh, UK.

Abstract (Italian)

La biologia sintetica è un'area emergente della bioingegneria avente come scopo la costruzione di nuove funzioni biologiche per svolgere specifici compiti definiti dall'uomo. Le regole progettuali si ispirano al mondo dell'ingegneria: come resistenze, condensatori o diodi vengono utilizzati per costruire un circuito elettrico, sequenze di DNA come geni, promotori, terminatori di trascrizione o siti di legame ai ribosomi possono essere assemblate tra loro a formare un *programma genetico*. Il paradigma ideale della biologia sintetica può essere riassunto con le seguenti fasi progettuali: i) scegliere le parti biologiche da un catalogo di componenti caratterizzati quantitativamente; ii) assemblarle tra loro per ottenere un programma genetico codificante la funzione desiderata; iii) incorporarlo in un organismo vivente.

Come nella totalità dei campi dell'ingegneria, la standardizzazione fisica dei componenti è stata introdotta per facilitare il loro assemblaggio definendo il concetto di BioBrick, ovvero parti di DNA con sequenza e funzione definite che possono essere assemblate mediante una procedura semplice e ripetibile grazie alla loro interfaccia fisica comune. Il Registry of Standard Biological Parts del Massachusetts Institute of Technology è stato il primo archivio di componenti standard e ad oggi comprende più di 16.000 BioBrick. Tuttavia, perchè esso possa considerarsi un vero catalogo di elementi standard, è necessaria la caratterizzazione quantitativa dei suoi componenti. Sebbene siano stati recentemente proposti approcci standardizzati anche per la misurazione delle caratteristiche di parti biologiche, la costruzione *bottom-up* di sistemi biologici con comportamento predicibile è ad oggi un problema di notevole

complessità. Alcuni dei fattori che contribuiscono a tale impredicibilità sono l'incompatibilità tra componenti, il loro comportamento dipendente dal contesto specifico, la variabilità intrinseca dei processi biologici e le non-linearità nell'espressione genica.

In futuro, il grande potenziale della biologia sintetica può produrre soluzioni di notevole impatto in svariati campi applicativi, come il biorisanamento, energie rinnovabili, nuovi biomateriali o farmaci. Tuttavia, il suo successo dipende dalla definizione dei limiti in cui i sistemi biologici possono funzionare in modo predicibile, permettendo così il progetto delle funzioni desiderate.

Questa tesi è incentrata sulla caratterizzazione quantitativa di parti e dispositivi in biologia sintetica, al fine di scoprire i loro limiti di predicibilità, fornire studi quantitativi che promuovano il loro ri-utilizzo e infine applicare i concetti chiave della biologia sintetica per affrontare problemi di rilevanza industriale nel campo dell'energia.

Nel Capitolo 1 è illustrato lo stato attuale delle ricerche in biologia sintetica, i risultati più importanti ottenuti, le promesse e le sfide future.

Nel Capitolo 2 è descritto uno studio *in vivo* sulla modularità di componenti biologici. In particolare, viene quantificata l'attività di un gruppo rappresentativo di promotori in svariati contesti per studiare quanto l'attività di parti biologiche possa cambiare all'interno di sistemi modello a complessità crescente.

Nel Capitolo 3 è riportata la caratterizzazione quantitativa completa di un dispositivo di lisi batterica. Nella filosofia di generare *datasheet* per componenti biologici, vengono misurate importanti caratteristiche come il comportamento statico/dinamico, la compatibilità e il tasso di guasto.

Nel Capitolo 4 è riportato il progetto di funzioni logiche (multiplexer e demultiplexer), ispirate all'elettronica digitale, in uno *chassis* biologico. Vengono utilizzati modelli matematici, identificati mediante l'uso di dati sperimentali, per descrivere il comportamento quantitativo globale dei sistemi e per evidenziare i parametri critici per la futura finalizzazione delle funzioni progettate.

Nel Capitolo 5 i concetti della biologia sintetica sono utilizzati per ottimizzare la conversione del siero di latte, uno scarto dell'industria casearia, in un biocarburante mediante un *pathway* metabolico ingegnerizzato. In particolare, viene incorporato un pathway per la produzione di etanolo in un microorganismo in grado di metabolizzare il lattosio, al fine di produrre energia da una fonte rinnovabile e allo stesso tempo smaltire un materiale di scarto che ad oggi rappresenta un problema ambientale.

Nel Capitolo 6 è presentata la caratterizzazione quantitativa di enzimi per la degradazione della cellulosa, incorporati in diversi chassis, con l'ambizioso obiettivo di convertire la biomassa ligneo-cellulosica in biocarburanti o altre molecole di interesse mediante l'uso di parti biologiche standard.

Infine, nel Capitolo 7 sono riportate le conclusioni di questo lavoro di tesi.

Gli studi illustrati nei Capitoli 2, 3, 4 e 5 sono stati svolti presso i laboratori di biologia molecolare del Centro di Ingegneria Tissutale e nel Laboratorio di Informatica Biomedica 'Mario Stefanelli', Università degli Studi di Pavia, mentre lo studio riportato nel Capitolo 6 è stato condotto presso l'Institute of Structural and Molecular Biology (School of Biological Sciences), University of Edinburgh, UK.

Contents

List of Figures

List of Tables

Chapter 1

Synthetic Biology

The central topic of this dissertation is the quantitative study of biological systems in the emerging area of bioengineering called synthetic biology. As a consequence, the first chapter is entirely dedicated to the description of this research field, highlighting its state of the art, major achievements, promises and challenges. In particular, a general introduction of the synthetic biology field will be provided (1.1), then the essential building blocks to engineer biological systems will be described (1.2) and the physical standardization process of biological parts will be discussed (1.3). Finally, the current challenges on the quantitative characterization of biological parts will be illustrated (1.4).

The reported information has also been published in a book chapter [1].

1.1 Introduction to synthetic biology

Synthetic biology is a new research area that merges the engineering and biological sciences expertise to synthesize novel biological functions. In particular, its goal is to create novel artificial living organisms that can be constructed, like machines, to carry out specific user-defined tasks [2, 3, 4, 5]. The fundamental idea is to consider the basic cell components as separate modules and use them either to design completely new organisms or to enable existing organisms to

carry out functions that in nature they cannot commonly achieve [6]. The design rules are inspired by the engineering world, in fact, like an electronic circuit is constructed by connecting resistors, capacitors or diodes, genetic circuits can be assembled by ligating DNA sequences, like genes, promoters or other regulatory sequences [7].

Nowadays, the construction of completely novel organisms from the scratch is an extremely complex process. One of the limitations is the incomplete knowledge about many of the encoded biological functions in the genome of even simple naturally occurring organisms. Even if steps towards the full understanding of the genome organization have been done through the sequencing of a large number of genomes, this lack of knowledge prevents the bottom-up design of customized life forms [8]. The physical construction process of an entire genome is also a concrete bottleneck, since the *de-novo* synthesis of large DNA fragments is time consuming, expensive and the maximum length allowed for the fragment is limited [9]. However, in the last decades a lot of technology advances have been observed in terms of time required for DNA synthesis, cost per base pair (bp) and maximum length reached (see Fig. 1.1).

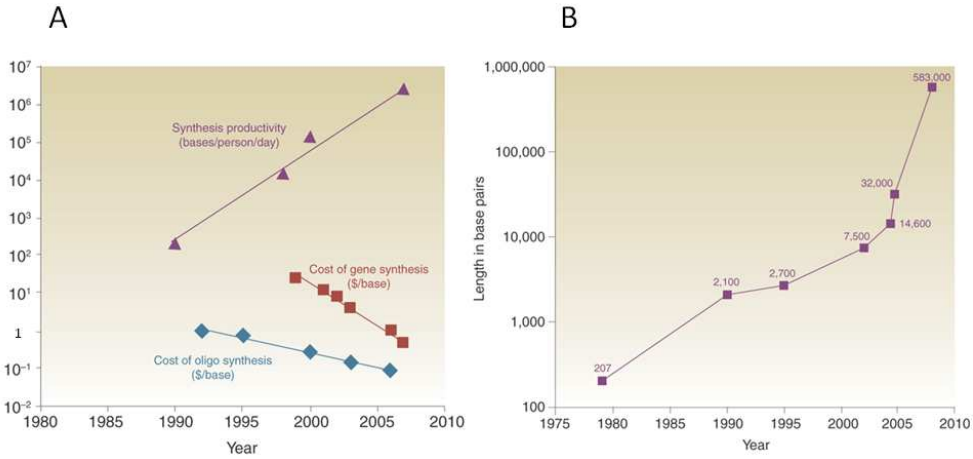


Figure 1.1: Productivity of oligonucleotide synthesis, cost of oligonucleotides and genes (A) and longest published synthetic DNA (B) over time (period up to 2008). Reference: [9]. After 2008, a synthetic DNA fragment as large as 1.08 Mbp has been reported (see text). The increasing trend of productivity and the decreasing trend of the costs of synthetic DNA are promising features for future construction of entire genomes.

The above mentioned limitations in the design and construction of completely novel synthetic organisms have brought researchers to perform feasibility studies by synthesizing the genome (sometimes with modifications) of naturally occurring microorganisms, characterized by a very small genome, and by proving that it could provide a functional life form. The first report of a *de-novo* created organism was the poliovirus, an RNA virus, whose genome (7.5 kbp) was synthesized in 2002 by the Wimmer group at the New York State University [10]. The synthetic DNA molecule was transcribed by RNA polymerase into viral RNA, which translated and replicated in a cell-free extract, resulting in the artificial recreation of a poliovirus with the same pathogenic characteristics of the naturally occurring one. This project, funded with 300,000 USD by the Pentagon in a research program against biological weapons, had the main goal of evaluating the possibility of creating an existing organism from the scratch. The most recent report of a synthetic cell is JCVI-syn1.0 (also called Synthia), a synthetic *Mycoplasma mycoides* bacterium that was developed in 2010 by the Craig Venter Institute (JCVI) [11]. The genome (1.08 Mbp) was assembled from 1-kbp fragments, purchased from a DNA synthesis company, that were assembled through novel DNA cloning methods developed at the JCVI [12, 13]. To test the DNA, the team inserted it into an empty cell of a different species of bacterium (*Mycoplasma capricolum*) as a recipient cell. The new cells were able to self-replicate, they had the expected phenotype of *M. mycoides* and they carried all the genetic features and 'watermarks' originally designed in the synthetic DNA. The provided examples mainly represent lab-reproductions of naturally occurring organisms, which are highly different from the creation of real synthetic life, even if these projects have contributed significant methodologies towards this goal. Some other research groups are facing this ambitious goal from the bottom-up, by studying the minimum constituents necessary for the emergence of life [14, 15].

The current limitations described so far in the synthesis of completely novel organisms imply that nowadays the most productive research field of synthetic biology is dealing with the bottom-up implementation of supplementary func-

tions in naturally occurring organisms [16]. In this case, the genome of the organism does not need to be artificially constructed and the supplementary functions are implemented by introducing exogenous genetic material, a *genetic program*, into the organism of interest (here called *host* or *chassis*). This program encodes the desired functions, which must not be incompatible with the life of the organism itself. Such artificial functions could yield applications of remarkable importance, like bioremediation or production of renewable fuels, new biomaterials and therapeutic molecules [17, 18]. In particular, in the future, medicine could take advantage of biological systems programmed for detection and destruction of cancer cells [19]; engineered microorganisms could be able to produce drugs or biomaterials via highly complex artificial metabolic pathways [20, 21]; renewable energy may rely on the conversion of wastes or sustainable feedstock into biofuels by novel biological catalysts [22, 23]; biological processes or environmental contaminants could be monitored by synthetic biosensors [24] and finally researchers could engineer a wide range of organisms to carry out signal processing tasks, such as logic functions [25, 26, 27], controllable waveform generation [28, 29] or information storage [30]. The first high-impact success of synthetic biology has been achieved by the Keasling lab at the University of Berkeley, where a microorganism has been engineered to produce the antimalarial drug precursor artemisinic acid, which can be subsequently transformed into the actual drug artemisinin [31]. In nature, artemisinic acid can be extracted from the rare plant *Artemisia annua*, but the plant grows only in a limited time window over the year and in specific locations and the overall yield of the drug precursor is low. For this reason, a complex metabolic pathway has been engineered and now it can yield an industrial-scale production of a low-cost antimalarial drug. The biotechnology company Amyris (California), which is now proprietor of this technology, is moving it into production with the help of the Institute of OneWorldHealth and Sanofi-Aventis pharmaceutical companies.

Synthetic biology aims at applying the key concepts of the engineering field,

such as standardization and modularity of components, to the rational design and construction of arbitrarily complex systems that accomplish the desired genetically-encoded instructions. In all the fields of engineering, the definition of standards has been one of the hallmarks that has enabled the design and construction of highly complex systems from a set of well-characterized standard components that are assembled together. In particular, physical standardization facilitates the assembly of components through a robust and reproducible procedure which can be followed for any component that conforms to the same standard. Reproducible and easy-to-carry-out methodologies to measure the characteristics of components are also necessary in order to express the same characteristics with the same units (e.g. volt and ampere to measure voltage and current). Thanks to such standard measurements, datasheets of components can be produced and collected in suitable data books in order to enable the comparison of components performances which can guide designers towards the rational choice of systems parts.

The lack of standardization in the assembly process of biological components has forced researchers to perform each assembly reaction in a different way, depending on the specific context. In fact, although recombinant DNA manipulation can be carried out through the same experimental protocols, it relies on DNA digestion by specific restriction enzymes, which vary according to the restriction sites in the specific DNA molecule. For this reason, the digestion/ligation modes can change and depend on the specific application, thus limiting this procedure to expert designers and preventing automation in the assembly process [32]. In order to enable an easy-to-carry-out assembly of basic components or devices created by different labs to compose circuits of higher complexity, a physical standardization of biological parts was necessary. In 2003, Dr. Tom Knight of the Laboratory of Computer Science, Massachusetts Institute of Technology (MIT), proposed a solution for this problem by creating the concept of BioBrick, whose name was derived from the Lego-Brick building blocks of the popular kids game Lego [33]. BioBricks are DNA sequences with

specific structure and function (like a promoter or a coding sequence) which share a common physical interface that facilitates their assembly through a standardized process. The laboratories of the MIT keep the Registry of Standard Biological Parts (Registry), an open-source archive that includes several thousands of BioBricks and that is constantly updated and maintained by the team of Prof. Randy Rettberg [34]. Tom Knight, in collaboration with the bioengineer Drew Endy, is also the inventor of the international Genetically Engineered Machine (iGEM), an annual competition that involves teams of students and researchers from universities worldwide [35]. In this contest, each team receives in spring a collection of BioBricks available in the Registry and has the possibility of assembling them to realize an original synthetic biology project, which has to be documented and presented in November at the MIT. The main goal of this competition is to create novel BioBricks that can be included in the Registry and to improve the knowledge about the existing parts. The biosensor for the detection of arsenic in drinking water, developed by the University of Edinburgh in 2006, is one of the first examples of useful artificial devices for real-world applications constructed in the field of synthetic biology starting from an iGEM project [36]. In this case, the genetic program for arsenic detection was incorporated in *Escherichia coli* bacterium.

Standard measurement techniques are also needed in synthetic biology for the production of technical datasheets which can guide designers in the rational selection of single components to compose the final system. Nowadays, such choice is difficult because components are often uncharacterized or they have been characterized either in different units of measurement or in the same units but in different experimental conditions, which yield uncomparable measurements between components [8, 37]. A popular example of the latter situation is the β-galactosidase assay for the characterization of promoters [38]: first, the reporter gene encoding for the β-galactosidase enzyme (or a part of it) is assembled downstream of a promoter of interest; then the promoter activity is indirectly measured by quantifying the produced β-galactosidase through an enzymatic assay in which a chromogenic or fluorogenic substrate is added and

it is cleaved by the enzyme to yield a detectable product. However, different substrates (e.g. ONPG, S-Gal, etc.) are used to carry out this procedure and so the produced absolute measurements of promoters activity cannot be compared [37]. Recently, the same group of the MIT that proposed the BioBrick physical standard has also proposed a set of characteristics that would facilitate the re-use of biological parts, as well as methodologies for the standard measurement of the activity of parts such as promoters, ribosome binding sites and transcriptional terminators [34, 37, 39, 40, 41].

Finally, it is important to note that the enormous potentialities of synthetic biology give rise to new issues about bioethics, biosecurity, health and intellectual property, which must be considered together with its ambitious scientific challenges [11, 42]. Particular attention is paid to the so-called dual use of synthetic biology: on the one hand, for example, promising technologies for efficient manufacturing of pharmaceuticals can be obtained, while on the other hand it is theoretically possible to create novel pathogens that could become devastating biological weapons, as the feasibility studies of the Wimmer and Venter research groups demonstrated. Risks rising from synthetic biology are also considered from the ethical point of view and a recurrent issue is whether the creation of new artificial living organisms is ethically licit. Research centers, associations and committees are working worldwide to produce protocols and rules to control the research activities and the scientific information that can be freely accessed on the Internet. An example is the SYNBIOSAFE project, funded by the European Community, which has been the first project in Europe to research the safety and ethical aspects of synthetic biology, aiming to proactively stimulate a debate on these issues [43].

1.2 Functional modules for programming cells

A number of functional components (modules) are available for the construction of genetic circuits. In this section some of the most important modules are

described in order to illustrate the basic elements that can be interconnected to program user-defined biological functions.

1.2.1 Promoters

A promoter is a DNA sequence that can be bound by the RNA polymerase to start the transcription of a gene. The promoter is found upstream of the gene and the transcription, performed by the RNA polymerase, can be regulated by specific factors that recognize the promoter region. These factors can activate or repress transcription. A class of promoters, called constitutive promoters, are always active and they cannot be repressed by any factors. In naturally occurring organisms, such promoters are essential for the expression of housekeeping genes. Thanks to their features, promoters can be defined from a functional point of view as the regulatory modules for gene expression. Such modules have an essential role in the design of synthetic biological systems.

The importance of promoters can be highlighted with the help of an example. The P_{tetR} promoter allows the constant expression of the downstream genes only in absence of the protein TetR, as it binds the promoter sequence and inhibits the transcription. However, such regulatory scheme is modified in presence of the antibiotic tetracycline (or the non-toxic synthetic analogous molecule aTc), as it can bind TetR which becomes unactive and unable to bind the promoter [44]. For this reason, TetR can be defined as the repressor of P_{tetR} and this promoter can be used as a tetracycline sensor, since the transcription is dependent on the presence of the antibiotic. However, it is important to note that the constant production of TetR protein is necessary for the realization of such sensor. The simplest way to obtain the constant production of a protein is to assemble the gene encoding this protein downstream of a constitutive promoter, which ensures its constant expression. In absence of TetR, P_{tetR} acts as a constitutive promoter.

In the design of synthetic biological systems, it is important to consider that promoters functioning is highly dependent on the chosen host organism.

For example, a bacterial constitutive promoter can transcribe the downstream genes in bacteria, but not in mammalian cells, for which other promoters must be chosen.

1.2.2 Coding sequences

Coding sequences (also simply called *genes* in this work) are DNA sequences that are transcribed in mRNA by the RNA polymerase and then are translated by ribosomes to produce proteins. The encoded proteins can carry out a very wide range of functions, as activators, repressors, enzymes or reporters. Every coding sequence starts with ATG nucleotides and stops with TAA, TGA or TAG. These nucleotides are the start and stop codons for translation.

The genetic code is nearly universal, with 17 minor variations that occur in the code of a few specific organisms, for example ciliata, the *Candida* or *Mycoplasma* species and mitochondria [45]. For this reason, a coding sequence isolated from a given organism (such as *Homo sapiens*) can be introduced into any other organism (such as *E. coli*) to synthesize the same protein.

Two important constraints exist in the transfer of foreign genes: i) the final protein must not be toxic for the host organism; ii) if the protein requires post-translational modifications, the translation of the coding sequence in another organism could be unable to give a functional protein.

Reporter genes are very useful coding sequences for the characterization and debugging of biological systems. They encode for proteins whose amount is easily detectable, for example fluorescent proteins (such as GFP or RFP) or enzymes that can be quantified through specific fluorimetric/colorimetric assays (such as β-galactosidase).

1.2.3 Transcriptional terminators

Terminators are DNA sequences able to stop the transcription of a gene or an operon (i.e. a set of genes regulated by a single promoter).

As it happens for the promoters, also the terminators functioning is dependent on the chosen organism. Prokaryotic terminators can be classified into rho-independent or rho-dependent, while transcription termination in eukaryotes relies on more complex and not fully understood mechanisms that involve several signal sequences such as the polyadenylation site. Commonly used prokaryotic terminators are rho-independent, so transcription is terminated by a hairpin structure, obtained by the annealing of complementary nucleotides, which mediates the detachment of the RNA polymerase.

An important parameter that characterizes terminators is efficiency, i.e. the probability that a successful termination is achieved. For example, if a terminator with efficiency of 0.64 is assembled downstream of a gene, it gives a successful termination of gene transcription with a probability of 64%.

1.2.4 Ribosome binding sites

Ribosome binding sites (RBSs) are small sequences that can be found at the 5' end of the mRNA molecule and can bind ribosomes. After the binding, ribosomes can start the translation of mRNA when the first start codon AUG is found. So, RBSs are signal sequences present in DNA that are transcribed in mRNA but are not translated by the ribosomes. The role of the RBS upstream of a coding sequence is crucial, as its absence does not enable the tight binding of mRNA to ribosomes, thus significantly reducing the translation efficiency. There are substantial differences between prokaryotic and eukaryotic RBSs: prokaryotic ribosomes can bind the Shine-Dalgarno sequence, eukaryotic ones bind a different sequence, known as Kozak sequence. Shine-Dalgarno and Kozak are not specific sequences valid for all the organisms, but are consensus sequences, i.e. they are derived from a multiple alignment and each sequence position reports the most conserved nucleotide. For this reason, the consensus sequence can be considered as the average sequence.

As it happens for promoters and terminators, it is important to note that the RBS sequence is highly dependent on the chosen organism. For example,

a eukaryotic RBS (containing the Kozak sequence) cannot be recognized by *E. coli*, whose ribosomes bind the Shine-Dalgarno sequence.

An important parameter for the RBSs is the strength of the binding between mRNA and ribosomes. RBSs also play a crucial role in operons (i.e. a set of genes regulated by the same promoter): in this case, each gene of the operon is preceded by an RBS, so some ribosomes will bind the first RBS causing the translation of the first coding sequence, while other ribosomes will bind the second RBS causing the translation of the second coding sequence, etc. As it happens in naturally occurring living systems, only prokaryotes support the structural organization of genes in operons.

The main problem in the RBS usage is that it could affect the stability of the mRNA, so a stronger RBS will not necessarily determine a higher protein level. Recently, algorithms have been produced to predict the RBS strength given the global mRNA sequence [46].

1.2.5 Other DNA

These modules include all the DNA sequences that are not directly involved in the regulation of gene expression. Important examples of this class of modules are here reported.

- Genomic integration sites: DNA sequences that enable the integration of DNA fragments in the genome of the host cell. They include transposons and site-specific recombination sequences from bacteriophages.

- Restriction sites: DNA sequences that are recognized by specific restriction endonucleases and can be cleaved by them.

- Primer binding sites: sequences that are used to start the DNA amplification by the DNA polymerase enzyme. Such sequences are important for the *in vitro* amplification of DNA and for sequencing.

- Plasmids: small circular double-stranded DNA molecules, able to replicate in a host cell and to carry the desired synthetic genetic program.

- Replication origins: sequences that enable the replication of plasmids in a host cell. Different replication origins determine different plasmid copy numbers, so the average number of plasmids per cell can be designed thanks to such sequences.

1.3 BioBricks and Registry of Standard Biological Parts

In this section, the structure (1.3.1) and the assembly process (1.3.2) of BioBricks will be discussed, highlighting the importance of physical standardization in synthetic biology.

1.3.1 BioBrick structure

BioBricks are DNA sequences with specific structure and function; they share a common physical interface and have been designed to be assembled and incorporated in living organisms, such as *E. coli*. Each BioBrick is contained in a plasmid. The essential elements that characterize a BioBrick plasmid are: a replication origin, an antibiotic resistance marker and a specific cloning site, i.e. the region in which the BioBrick is inserted [47].

The DNA sequences adjacent to the BioBrick have specific sequences containing four different restriction sites. In particular, the BioBrick is flanked by a prefix sequence upstream and a suffix sequence downstream. The prefix includes the EcoRI (E) and XbaI (X) restriction sites, while the suffix includes the SpeI (S) and PstI (P) sites. The BioBrick structure is reported in Fig. 1.2.

The sequence and cleavage mode are reported for all the four restriction sites in Fig. 1.3. Specific DNA cleavage by all the four restriction enzymes

produces sticky ends. For example, if the EcoRI site is cleaved it produces an AATT protruding sequence in one DNA strand and TTAA in the other strand. AATT can only anneal with TTAA to reconstitute a new EcoRI site after a proper DNA ligation.

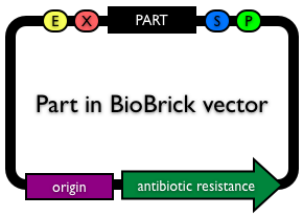


Figure 1.2: Essential elements of a BioBrick plasmid: replication origin, antibiotic resistance gene and cloning site, flanked by prefix and suffix sequences. Reference: [34].

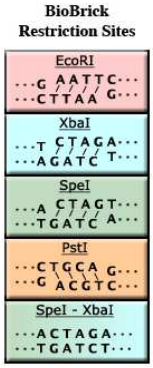


Figure 1.3: Restriction sites of a BioBrick cloning site: EcoRI (E), XbaI (X), SpeI (S) and PstI (P). The hybrid site obtained by ligating a cleaved SpeI site with a cleaved XbaI site is also reported. Reference: [34].

It is important to note that the DNA cleavage performed by SpeI and XbaI produces identical sticky ends, i.e. CTAG and GATC. For this reason, SpeI and XbaI are called *isocaudamers*, as the recognition site is different but the protruding DNA sequences are the same. This feature enables the annealing between a cut SpeI site and a cut XbaI site, thus forming, after a proper ligation, a new hybrid site, whose sequence is reported in Fig. 1.3. This site is no longer recognized by SpeI or XbaI enzymes and so it cannot be cleaved.

The described characteristics of the BioBrick restriction sites enable a standardized process of part assembly, which is reported in the next section with the help of an example.

1.3.2 Standard Assembly

Let B0034 and C0010 be two BioBricks (see Appendix A for the BioBrick nomenclature). The goal is to obtain the composite part B0034-C0010 in the plasmid backbone of C0010. In this case the following steps are necessary (see Fig. 1.4):

- digest the B0034 plasmid with E and S;
- digest the C0010 plasmid with E and X;
- ligate the excised E-S fragment containing B0034 with the opened plasmid of C0010.

In this way, the E site in the excised fragment anneals with the E site in the opened plasmid and the S site in the excised fragment anneals with the X site in the opened plasmid. It is important to note that the composite part B0034-C0010 can be used in another assembly step because all the standard restriction sites E, X, S and P are intact. In particular:

- E has been reconstituted in the ligation of the excised fragment to the opened plasmid;
- X comes from the excised fragment;
- S comes from the opened plasmid;
- P comes from the opened plasmid;
- the S-X hybrid site is not recognized by any of the E, X, S or P restriction enzymes, so the two assembled parts cannot be separated by digestion in the subsequent assembly steps.

On the other hand, if the goal is to assemble B0034 downstream of C0010 in the C0010 plasmid, the restriction sites to cut are X-P for B0034 and S-P

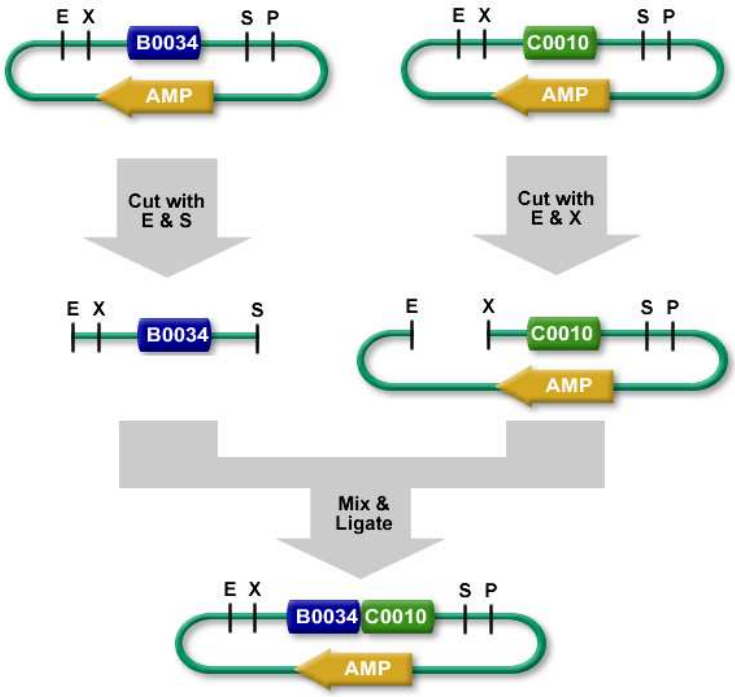


Figure 1.4: Standard Assembly process. Reference: [34].

for C0010 plasmids. In this case, the two cut P sites anneal to reconstitute a P site and the S and X sites anneal as before to form the hybrid site.

Because the integrity of all the standard restriction sites is always maintained, the described procedure is *standard*, as it can be repeated at each assembly step by using the same four restriction enzymes. An obvious constraint is that BioBrick parts do not contain E, X, S or P restriction sites, as they would be digested in unwanted points during the assembly process.

E, X, S and P sites are not the only features included in the prefix and suffix. Fig. 1.5 shows the complete DNA sequence of a BioBrick cloning site: it also includes two NotI restriction sites, between E and X and between S and P respectively for sites spacing and debugging purposes. Single nucleotides are also present between some restriction sites to avoid the accidental generation of EcoBI, EcoKI or Dam methylation sites which could 'protect' flanking DNA from digestion.

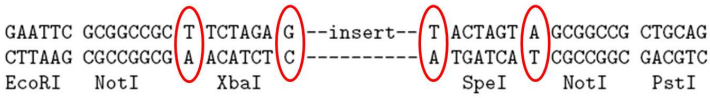


Figure 1.5: Complete DNA sequence of BioBrick prefix (upstream of the insert) and suffix (downstream of the insert), showing the restriction sites used for the Standard Assembly process (EcoRI, XbaI, SpeI and PstI), the two NotI sites and the additional nucleotides (red circles) flanking the mentioned sites to avoid unwanted DNA methylation.

The BioBrick structure and Standard Assembly process described so far refer to the original BioBrick format, now known as BioBrick Standard 10, invented by Tom Knight. After the dissemination of this standard, other research groups proposed different physical standards. An example is the BioBrick Standard 21 (or BglBrick) in which prefix and suffix sequences are composed by EcoRI-BglII and BamHI-XhoI respectively, where the isocaudamer sites are BglII and BamHI [18]. Such format is convenient when sequences encoding for protein domains have to be ligated *in frame*, as the DNA scar left after an assembly process does not include stop codons and it is always 6-nucleotide long, so the reading frame is maintained. Moreover, the two encoded amino acids in the scar are glycine and serine, often used in the linkers of many fusion proteins. On the other hand, the DNA scar left after a Standard 10 assembly contains a stop codon (TAG) in frame (see Fig. 1.5), which impairs the functional assembly of DNA parts encoding protein domains to obtain a fusion protein.

Nowadays the term BioBrick is a trademark of the BioBricks Foundation [48], a non-profit organization founded by engineers and scientists of MIT, Harvard and UCSF that promotes the development and responsible usage of BioBrick-based technologies and develops methods to improve the usage of existing components and the construction of novel standard biological parts.

In the development and improvement of BioBrick standards, the BioBricks Foundation has implemented a Request For Comment (RFC) process in its web site, inspired to the Internet Engineering Task Force RFC process, proposed in the past in order to define, evaluate, introduce and review new standards

in the field of interest. RFCs are short documents that can be examined by the community and from them new standards can rise. In particular, an RFC has to carry out at least one of these tasks: i) propose a standard; ii) describe methods and protocols (e.g. how to assemble two parts); iii) provide general information (e.g. how to design a transcriptional terminator) or iv) comment, extend or replace an older RFC [49].

1.4 Measuring the activity of biological parts

Apart from the physical assembly standardization to combine parts, it is important to have suitable, standardized methods to measure their characteristics. Thanks to the *standard measurements*, in the engineering world it is possible to characterize the modules of interest in comparable units, so that their performance can be compared, and their parameters and measured characteristics can be annotated in *datasheets*.

In this section, datasheets for biological components are introduced (1.4.1) and the standard measurement kits for biological parts proposed so far are described (1.4.2).

1.4.1 Datasheets in synthetic biology

In order to consider the Registry a real handbook of biological parts, which can be consulted by designers to build up new functions, a quantitative characterization of the components, both basic and composite, is needed. This process requires the definition of important parameters and the setting up of new methodologies to measure them in a standardized way, in order to make comparable the results obtained in different laboratories, with different measurement instruments.

Here, the parameters that have been considered important in the definition of a datasheet of biological parts are shown. Most of the information contained

in this section comes from [37, 39, 40, 41].

As shown in the previous section, biological components can have different functions. For every category of modules, it is therefore necessary to define the parameters of interest for the compilation of detailed datasheets. Even if each category of components has typologies of specific parameters, some parameters are common to all modules having well defined inputs and outputs. Inspired by datasheets of electronic components, the following elements have been proposed:

- *static characteristic*, which is the input-output characteristic at the steady state, important to allow the prediction of interconnected modules behaviour;
- *dynamic behaviour*, which is the response time of the system, i.e. the required time to reach the steady state;
- *compatibility* with other biological elements, which can be exogenous inputs (e.g. inducer molecules), experimental conditions (e.g. temperature, pH, medium, etc.) or a defined chassis; to this aim, it is important to report any incompatibility or crosstalk among components;
- *reliability* of the component, also called failure rate, which is the time (or number of cellular divisions) that passes before a significant part of the recombinant population stops showing the programmed function;
- *metabolic demand*, which represents the amount of resources (nucleotides, tRNA, polymerases, ribosomes, etc.) required for the proper functioning of the module.

These parameters would support the re-usability of characterized biological components. However, standardized methodologies for their assessment do not exist yet. Research has focused on the standardization of input and output signals based on transcription, thus defining a standard carrier. The rate at

which RNA polymerase moves past a particular location in the DNA, measured in *polymerase per second (PoPS)*, has been identified as standard measurement unit able to represent the input or the output (or both) of most biological devices, just as electric current is used to characterize electronic devices. A constitutive promoter, for example, is a module having PoPS as output and no inputs; the sensor of tetracycline/aTc described in 1.2.1 is a module having PoPS as output and the concentration of tetracycline/aTc as input; a module made up of RBS-GFP-Terminator has fluorescence intensity (or the amount of GFP produced) as output and PoPS as input. The latter module can be defined as a fluorescence *actuator*, that is, it receives a signal (PoPS) from another device and responds with a specific action (GFP production). From this example it is clear that it is very important for input/output biological devices to share the same carrier. In fact, thanks to a standard signal, 1) the features of similar devices, like promoters, are easily comparable and 2) the behaviour of a system made up by the interconnection of different, well-characterized modules, can be theoretically predicted. The parameter PoPS, however, is not directly accessible, since by now it is not possible to measure the number of RNA polymerases which flow through a specific point of DNA over time *in vivo*. Therefore, reporter genes have been used to measure this parameter indirectly. In particular, let us assume that we want to assess the parameter PoPS(t) of a promoter of interest and also that the RBS-GFP-Terminator part is assembled under the control of this promoter. The following system of differential equations can be used to describe the dynamics of GFP transcription (Eq.1.1), translation (Eq.1.2) and maturation (Eq.1.3) for this genetic circuit:

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

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

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

where [M] is the concentration of mRNA per cell, [I] the concentration of non mature GFP (non fluorescent), [G] the concentration of mature GFP (fluorescent), n is the number of copies per cell of the circuit, PoPS is the synthesis rate of mRNA per DNA copy, γ_M is the extinction rate of mRNA per cell, ρ is the protein synthesis rate per mRNA, α is the maturation rate of GFP, γ_I is the extinction rate of non mature GFP per cell and finally γ_G is the extinction rate of mature GFP per cell. All extinction rates are to be considered as the sum of two contributes: $\gamma = d + \mu$, where d is the molecule degradation rate (mRNA or protein), while μ is the molecule dilution rate due to cellular division; therefore the last parameter coincides with the population growth rate.

Assuming the steady state of the concentrations of the three quantities of interest, and assuming also that $d \gg \mu$ for mRNA and that $d \ll \mu$ for the non mature and mature protein, it is possible to obtain an expression for PoPS at the steady state:

$$PoPS^{SS} = \frac{\gamma_M(\alpha + \gamma_I)\gamma_G[G]^{SS}}{\rho\alpha n} \tag{1.4}$$

where the apex 'SS' means 'steady state'.

Moreover, if we consider Eq.1.3, it is possible to define:

$$S_{cell} = \alpha[I] = \frac{d[G]_{tot}}{dt} \cdot \frac{1}{CFU} \tag{1.5}$$

as the synthesis rate of mature GFP per cell, where G_{tot} is the number of GFP molecules in a culture and CFU (*colony forming units*) is the number of cells in the same culture. It is therefore also possible to write the $PoPS^{SS}$ parameter as:

$$PoPS^{SS} = \frac{\gamma_M(\alpha + \gamma_I)S^{SS}_{cell}}{\rho\alpha n} \tag{1.6}$$

Thanks to Eq.1.4 and 1.6 it is possible to estimate $PoPS^{SS}$ given the other parameters of the differential model, the GFP produced per cell (or in the

whole culture) and the number of cells in the culture. The calculation of produced GFP molecules and of the number of cells in the culture requires a proper calibration of the measurement instruments, because these instruments are able to detect the optical density (OD), proportional to the number of cells, and the fluorescence intensity (F), proportional to the number of GFP molecules.

Having the PoPS of a promoter, this can be used as an instrument to measure the output of another module which requires PoPS as input, just as a tester is used to inject a known current to measure the parameters of a component or an electronic circuit. The concepts described above have led to the production of a datasheet for BioBrick BBa_F2620, the first prototype of datasheet of a biological component [40].

Even if PoPS has interesting features as a signal carrier, unfortunately the procedure to obtain it is not simple. In fact, the assessment of the model parameters requires specific experiments for every experimental condition used (like the strain, the medium, etc.). For example, the degradation rate of mRNA requires the use of real time PCR. The not easy measurement of these parameters increases the complexity of PoPS assessment. Moreover, the measurements of PoPS carried out in independent laboratories or using different measurement instruments, have been proved to give highly variable estimations. Under these conditions, the measurement unit PoPS cannot be considered standard, since it is difficult to estimate and it is not robust. In order to facilitate the evaluation of promoters strength, a standard methodology has been developed, which will be discussed in the following section.

1.4.2 Standard measurement kits for biological parts

In order to try to compensate the variability in the assessment of the transcriptional activity of promoters due to different measurement instruments or operators, this activity has been expressed in measurement units relative to

an *in vivo* standard reference, rather than absolute units (like PoPS) [37]. In several scientific fields where absolute measurements are not reliable or easy to carry out, in fact, it is convenient to refer the measurements to a comparative standard. A practical example in the field of bioengineering is the comparison of microarray results among studies made in independent laboratories. This comparison is impossible because of numerous differences, such as platforms employed, image processing algorithms and other experimental conditions such as probe length, area of the spots containing probes, molecules per spot, etc. To solve the problem of comparison, which can also be very important in the field of medicine, many works have proposed to use standard samples of RNA to normalize the signals and therefore to standardize the results obtained from different experiments. As far as promoters are concerned, a similar approach has been proposed for *E. coli*. As in the previous section, let us assume that we want to quantify the strength of a promoter of interest, φ, and to assemble RBS-GFP-Terminator under the control of said promoter. The relative activity of φ, expressed in RPUs (Relative Promoter Units), is defined as the ratio between the absolute activity of φ and the absolute activity of the BioBrick promoter BBa_J23101 (TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGC, assembled to RBS-GFP-Terminator, too), both expressed in PoPS, under identical experimental conditions and with the same measurement instrument:

$$Relative\ activity\ of\ promoter\ \varphi\ (RPUs) = \frac{PoPS_{\varphi}^{SS}}{PoPS_{J23101}^{SS}} \tag{1.7}$$

With this definition, a promoter with RPU=1 has an activity equivalent to BBa_J23101. Substituting Eq.1.6 in Eq.1.7 and maintaining the symbology introduced in the previous section, the following expression for RPUs is obtained:

$$Relative\ activity\ of\ promoter\ \varphi\ (RPUs) = \frac{\frac{\gamma_{M,\varphi}(\alpha_\varphi+\gamma_{I,\varphi})S^{SS}_{cell,\varphi}}{\rho_\varphi \alpha_\varphi n_\varphi}}{\frac{\gamma_{M,J23101}(\alpha_{J23101}+\gamma_{I,J23101})S^{SS}_{cell,J23101}}{\rho_{J23101}\alpha_{J23101}n_{J23101}}} \tag{1.8}$$

In order to simplify Eq.1.8, besides the hypotheses on parameters made in the previous section, we assume that:

1. since the two promoters are measured under the same experimental conditions, $\alpha_\varphi = \alpha_{J23101}$;
2. since the same reporter device RBS-GFP-Terminator is assembled under both promoters, the mRNA produced will be identical and it is assumed $\gamma_{M,\varphi} = \gamma_{M,J23101}$ (if promoters initiate transcription at the same site);
3. since the plasmids must be the same for φ and for J23101 (to standardize the experimental conditions), the copy number is equal by hypothesis, therefore $n_\varphi = n_{J23101}$;
4. if $\mid \mu_\varphi - \mu_{J23101} \mid \ll \alpha$ then $\frac{\alpha+\mu_\varphi}{\alpha+\mu_{J23101}} \approx 1$ obtained from experimental tests.

These assumptions have allowed to write:

$$Relative\ activity\ of\ promoter\ \varphi\ (RPUs) = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,J23101}} \tag{1.9}$$

The parameter S_{cell} of both promoters is easily detectable, under the hypotheses of steady state, as:

$$S_{cell} = \frac{dF_{tot}}{dt} \cdot \frac{1}{OD} \tag{1.10}$$

or

$$S_{cell} = \mu \cdot F \tag{1.11}$$

To calculate the RPUs of a promoter, being Eq.1.9 a ratio it is not necessary to calibrate the instruments for the calculation of GFP molecules and of CFU respectively. Moreover the previous simplifications allow the measurement of RPUs without measuring the maturation rate α, γ_M, ρ and n for the specific experimental setup, thus reducing the calculation of the activity of the promoter of interest only to the measurements of fluorescence (F) and absorbance (OD). S_{cell} can be calculated from either population measurements (Eq.1.10), where F_{tot} is the total fluorescence of the culture, or single-cell measurements (Eq.1.11), where F is the per-cell fluorescence. Given a specific experimental condition, it is therefore possible to define a ranking of promoters through their RPU value, thus allowing the rationalization of the choices of a biological systems engineer. By measuring the absolute activity (PoPS) of the reference promoter BBa_J23101, it is also possible to evaluate the absolute activity of all the promoters for which RPUs have been measured under the hypotheses described so far and for each specific experimental condition.

This method has been successfully used for the quantitative characterization of different promoters [50, 51] and it has also been applied to carry out the characterization of BioBricks for the iGEM competition [35]. However, the steady state for gene and protein expression is a very simplistic and not easily validable hypothesis. In order to standardize the growth conditions of the culture in which the constructs of interest are incorporated, it has been suggested to make the measurements during the exponential phase of bacterial growth, where the nutritional resources are sufficient to support the gene expression of all the population: therefore the hypothesis is that the steady state is valid in this phase. It is important to underline, for completeness, that the methodologies used to characterize the promoters against a reference standard have been employed for the definition of standardized measurement methods to characterize also RBSs and terminators through the definition of proper kits in *E. coli* [41] and to characterize promoters in mammalian cells [52].

Chapter 2

Composition of biological systems through standard bricks: a modularity study on basic parts and devices

Modularity is a crucial aspect in the engineering world, as it enables to achieve predictable outcomes when different functional modules are interconnected. Synthetic biology aims to apply key concepts of engineering to design and construct new biological systems that exhibit a predictable behaviour. Even if physical and measurement standards have been proposed to facilitate the assembly and characterization of biological components, real modularity is still a major research issue. The success of synthetic biology depends on the definition of the working boundaries in which biological functions can be predictable, in order to enable the bottom-up composition of customized systems.

Here, the modularity of transcription-based biological components has been investigated in *ad-hoc* constructed model systems. After a background section in which the current status of this research is illustrated (2.1), the tested systems and the obtained results will be presented (2.2) and conclusions will

be provided (2.3).

2.1 Background

Standardization of components, abstraction, modularity and predictability are the main engineering principles for which the emerging field of synthetic biology lays the foundations [7, 32]. Following these principles, the ultimate goal is to design and construct new biological systems that exhibit a predictable and user-defined behaviour, starting from a set of quantitatively characterized standard components, exactly as it is accomplished in all the fields of engineering [3, 6]. Standard biological parts, such as BioBricks, are the building blocks that may enable the composition of these systems, according to their abstraction hierarchy and standardized assembly process [18, 23]. The Registry of Standard Biological Parts [34] encloses a collection of thousands of online-browsable BioBricks whose structure and function are listed and, when properly characterized, they could be exploited for the bottom-up construction of the desired composite systems [40]. Such paradigm may contribute to the construction of biological solutions for a large variety of applications, from medicine to renewable energy production [19, 22].

However, the tremendous potential of synthetic biology is limited by the intrinsic complexity of living systems [8]. In fact, although abstraction and physical standardization concepts have been successfully proposed by the definition of BioBricks, exhaustive characterization of components is currently a major challenge [53]. Standard measurement techniques have been introduced for promoters, terminators and ribosome binding sites (RBSs) [37, 41]. In the case of promoters, the Relative Promoter Unit (RPU) method was proposed to improve the reproducibility of measurements among different instruments and different labs. This method relies on the measurement of the activity of promoters relative to a standard reference promoter, assayed *in vivo* in the same experimental conditions. RPUs have been recently used to characterize many

promoters in literature [51, 54] and this is also a popular approach among the international Genetically Engineered Machine (iGEM) competition, as it has been exploited by many teams to share their quantitative measurements in the Registry [35].

The central challenge about quantitative characterization is the modularity of components [55]. Modularity is a crucial aspect in the engineering world, as it enables to achieve predictable outcomes when different functional modules are interconnected [56]. Even if the successful interconnection of biological modules has been reported in many studies [53, 57], most of them relied on trial-and-error approaches or time-consuming debugging of the constructed systems [31, 41, 58] and nowadays the behaviour of composite parts is hard to predict from individually characterized components. Recent efforts in such research have focused on model-based design and qualitative prediction of genetic circuits behaviour [25, 53].

Crosstalk or incompatibility among components [7], context-dependent behaviour of parts [50], intrinsic noise that characterizes biological processes [59] and nonlinear effects on gene expression caused by cell machinery overburdening [60] are some of the limiting factors that contribute to the unpredictability of bottom-up-composed systems. For example, promoters activity may be affected by their flanking sequences and this effect contributes to the unpredictability of these components when moved to a different physical context [50].

As synthetic biology key concepts are based on engineering, the success of this emerging field depends on the definition of the working boundaries in which biological functions can be predictable, in order to enable the bottom-up composition of customized systems. Recently, some experimental studies have been reported to assess modularity of biological components. Davis et al. [50] proposed design rules to engineer insulation of promoters in order to improve their predictability when different DNA sequences surround them. Hajimorad et al. [60] reported the working limits in which the superposition of the effects

could be valid in a model system composed by one to three independent gene expression cassettes as a function of the system copy number. Both works are promising starting points for studying context-dependent activity of biological parts and devices.

Here, the modularity of transcription-based biological parts and devices has been investigated in *ad-hoc* constructed model systems.

2.2 Systems design and results

To study the context-dependent variability of biological parts, the quantitative behaviour of promoters was studied in three increasingly complex conditions in *E. coli*, a popular chassis in synthetic biology. First, the activity of a set of promoters was quantified *in vivo* via different biological measurement systems (i.e. different plasmids, reporter genes and ribosome binding sites) relative to a reference promoter in identical conditions, using the Relative Promoter Unit (RPU) approach (2.2.1). Second, promoter activity variation was measured when two independent gene expression cassettes were assembled in the same system (2.2.2). Third, the modularity of input devices was tested in a functionally interconnected framework, composed by a variable input device connected to a fixed output device (a logic inverter) expressing GFP (2.2.3).

2.2.1 Characterization of a set of promoters via different biological measurement systems

In order to estimate the activity variation when a promoter is characterized via different biological measurement systems, a representative set of five widely used promoters was assembled to three different reporter expression devices in two different plasmid vectors and characterized in terms of RPU in the TOP10 strain (see Fig. 2.1A and B). Relying on relative measurements, RPUs enable the comparison among different fluorescent proteins, used as reporters.

Moreover, a promoter quantified via different measurement systems should give the same results among the conditions (see Eq.1.8 and 1.9). The considered promoters and tested conditions are listed in Fig. 2.1C and D.

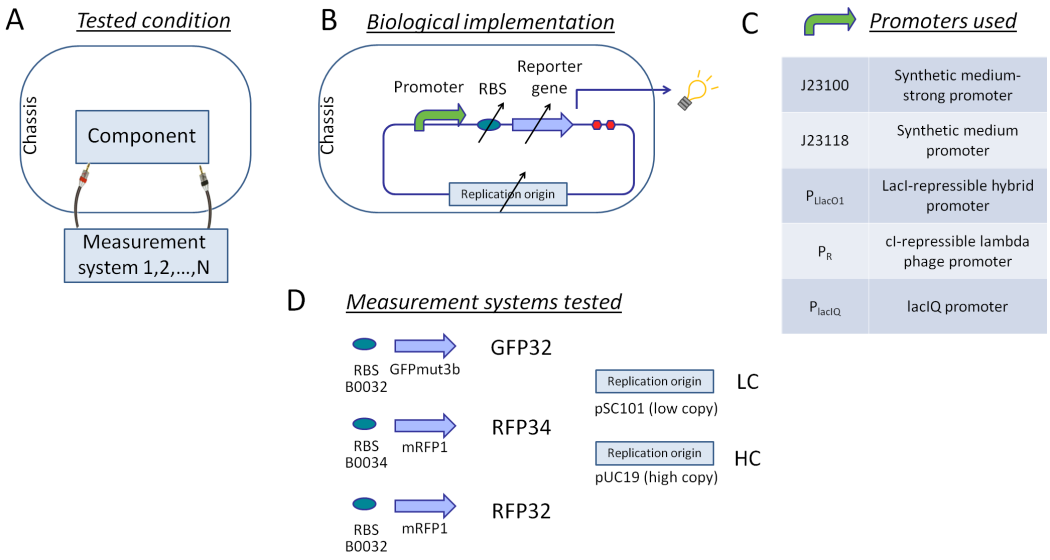


Figure 2.1: Study of individual promoters activity quantified via different measurement systems. A) Schematic functional representation of the tested framework. B) Promoters are tested via different biological measurement systems (i.e. RBSs, reporter genes and/or plasmid copy number) and their activity is computed relative to a standard reference promoter with the RPU approach. C) Promoters used in this study. D) Three reporter devices (GFP32, RFP34 and RFP32) and two copy number conditions (LC or HC) are used as different biological measurement systems.

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 [34, 41, 61]. Only P_R activity is statistically different among the three tested measurement systems (P<0.05, ANOVA), yielding a CV of 22% among the three measured mean activities.

The observed variability may be caused by downstream sequence-dependent

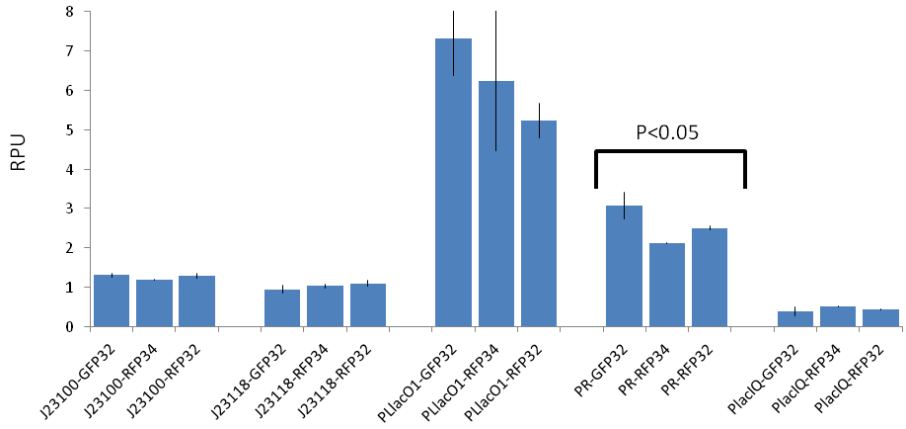


Figure 2.2: Measured RPU values for the five investigated promoters, individually characterized via three reporter devices: GFP32, RFP34 and RFP32. Error bars represent the standard deviation of the mean activity computed on three clones. Promoters showing a statistical difference in the mean activities among the three conditions are highlighted.

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 [50, 51].

Promoters with GFP32 and RFP34 reporter devices were also tested in a high copy number plasmid. Results showed that the RPU activity of the strongest promoters (P_{LlacO1} and P_R) is much weaker in high copy than in low copy context (up to 4.4- and 2.3-fold respectively, see B.2.3 in Appendix B). Such results are in accordance with other studies in which devices in high copy number showed saturation effects in strong promoters activity [54] or a nonlinear response in the DNA-mRNA device transfer curves [60].

2.2.2 Promoter activity variation when independent expression modules are physically combined

A subset of the reporter expression cassettes tested above were assembled in order to study the context-dependent variability in promoters activity when the system is slightly more complex. In this framework, constructs are combined in the same plasmid, but do not functionally interact with each other (see

Fig. 2.3A and B). In order to quantify the activity of two promoters in the same system, two reporter genes (GFP and RFP) had to be used. GFP32 reporter device driven by the medium-strength J23118 promoter was kept constant in all the constructed composite systems, while RFP34 reporter device was driven by one of the other four promoters (J23100, P_{LlacO1}, P_R or P_{lacIQ}). The relative position of the cassettes was also cross-exchanged (see Fig. 2.3B).

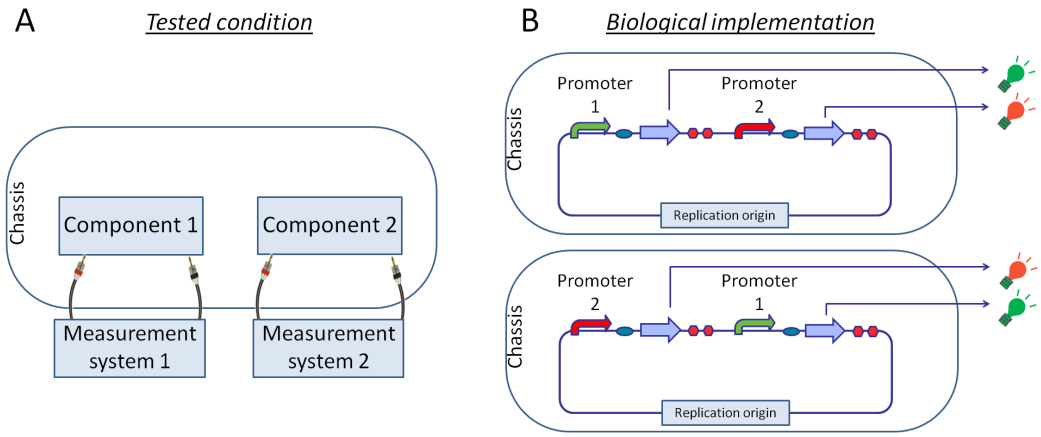


Figure 2.3: Study of promoters activity when two independent expression modules are physically combined. A) Schematic functional representation of the tested framework. B) Promoters with GFP32 or RFP34 are assembled in the same plasmid and quantified. The resulting activity will be compared to the one measured in the individual characterization study to investigate context-dependent activity changes. The two gene expression cassettes are also cross-exchanged to investigate the relative position-effect.

Fig. 2.4 shows the resulting activity of combined promoters in two different *E. coli* strains. The activity of the individually characterized promoter, via the same measurement system, is also reported for each group. Among the groups where at least one of the mean activities showed statistical difference ($P<0.05$, ANOVA), CVs of 33%, 7%, 33% (for J23118, P_R, P_{lacIQ} in TOP10) and 27% (P_{LlacO1} in KRX) were observed.

Part of the variability may be due to upstream sequence-dependent promoter activity change. In particular, the promoter driving the 'upstream' expression cassette has the same flanking sequences as the one in the individually characterized condition (i.e. the same vector sequence upstream and the same reporter expression device downstream), but the promoter driving the

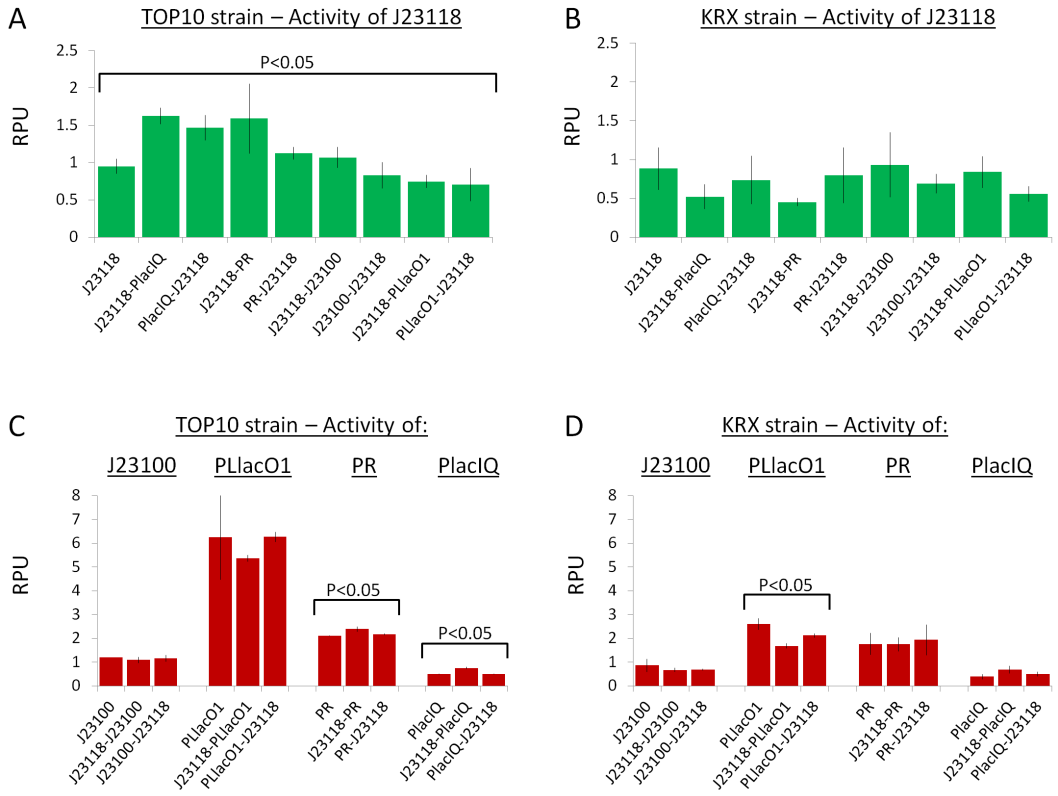


Figure 2.4: Measured RPU values for promoters assembled in the same plasmid in TOP10 and KRX strains. Green and red bars indicate that the promoter has been characterized via GFP32 and RFP34 reporter device respectively. Error bars represent the standard deviation of the mean activity computed on three clones. For each group, the RPU value of the individually characterized promoter is also reported. Promoters showing a statistical difference in the mean activities among the tested conditions are highlighted.

'downstream' cassette has a different upstream sequence (i.e. the transcriptional terminator of the other cassette). If the data of the promoters driving the 'downstream' cassette are excluded from the variability analysis, only the J23118 promoter in TOP10 shows a significant difference among the tested contexts, with a CV of 35%. This demonstrates that flanking sequences significantly contribute to context-dependent differences, but they are not sufficient to explain all the observed variability.

2.2.3 Modularity test for input devices connected to a NOT gate

The modularity of biological devices was studied when dealing with functionally interconnected circuits. The basic idea driving this part of the study is illustrated in Fig. 2.5A. Considering interconnected systems composed by different input blocks (X_1, X_2, ..., X_N) and a fixed output block (Z) downstream, if the signals provided by the input blocks are the same ($in_1 = in_2 = ... = in_N$), the output signals must be identical ($out_1 = out_2 = ... = out_N$) even if the input blocks are structurally different.

To test this condition, the model systems shown in Fig. 2.5B were constructed (with the inputs shown in Fig. 2.5C, D and E) and tested in the KRX *E. coli* strain, which over-expresses the lacI repressor. The circuits are composed by a variable input device interconnected to a fixed output device. The considered input modules were: i) a set of constitutive promoters of different strengths (four J231xx-family members, here called INPUT1), ii) a lacI-regulated promoter (P_{LlacO1}, here called INPUT2) and iii) a luxR-regulated promoter (P_{lux}, here called INPUT3). They provide a transcriptional signal that drives the output device. The latter is a logic inverter (or NOT gate), i.e. a promoterless tetR repressor-expressing cassette connected with a tetR-repressible promoter (P_{tetR}) downstream that can be inhibited by TetR. P_{tetR} expresses GFP as the system output through the GFP32 reporter device. The lacI-, luxR- and tetR- regulated systems, as well as the J231xx constitutive

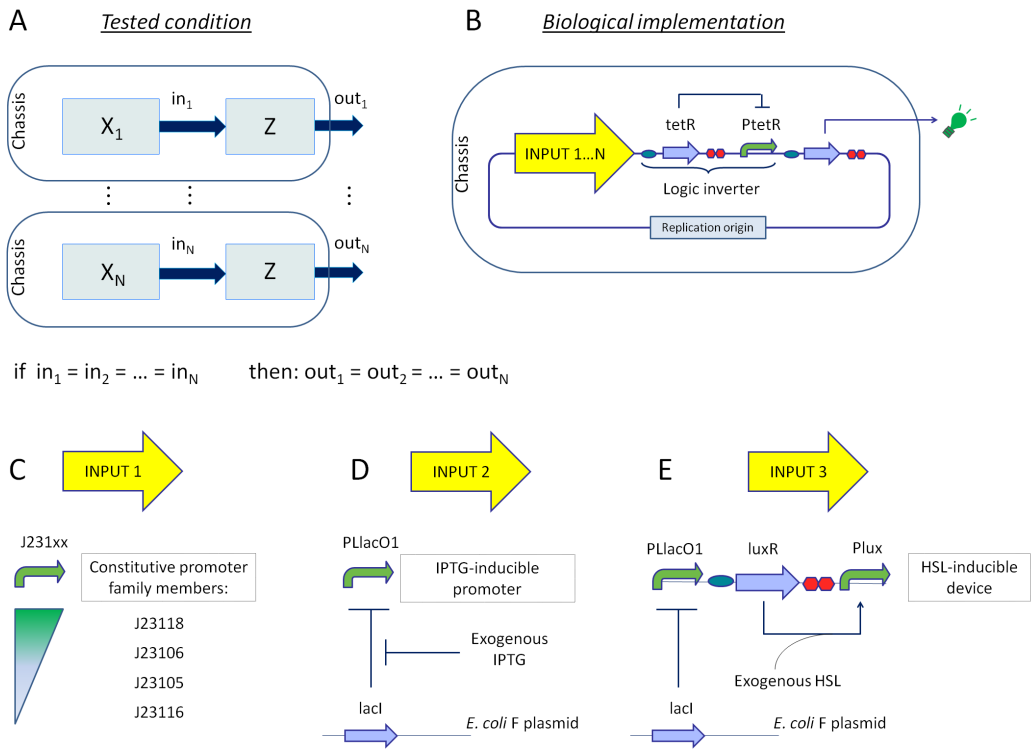


Figure 2.5: Modularity study for input modules in interconnected systems. A) Schematic functional representation of the tested framework. B) Different input devices are assembled upstream of a tetR-based logic inverter. They provide transcriptional signals that drive the inverter. C) INPUT1: a set of four constitutive promoters of different strengths. D) INPUT2: P_{LlacO1} promoter, which is repressed by the endogenously-overexpressed lacI and can be induced by IPTG. E) INPUT3: luxR-based HSL-inducible device. P_{lux} can be induced by LuxR-HSL complex. The luxR gene is produced by the weak basic activity of repressed P_{LlacO1} in absence of IPTG. IPTG = isopropyl β-D-1-thiogalactopyranoside; HSL = N-3-oxohexanoyl-L-homoserine lactone.

promoters, are all widely used in the design of genetic circuits [28, 51, 62]. If modularity persists, identical transcriptional signals trigger identical GFP outputs for any connected input device.

To investigate the validity of this framework, at first the input devices were individually characterized in terms of RPU via RFP measurements. Fig. 2.6A, B and C report the characterization of the individual input devices. The four considered constitutive promoters yielded a 10-fold activity range (Fig. 2.6A). The induction curves of the two inducible inputs showed that they were tightly controlled (i.e. the basic activity of promoters in the uninduced state was very low) and they could be regulated over a wide range of activities, yielding a maximum activity of ∼2.5 RPUs (lac) and ∼4.2 RPUs (lux). Their induction curves were fitted with a Hill function (Fig. 2.6B and C).

Subsequently, the interconnected systems composed by the different input devices and the NOT gate downstream were considered. The transcriptional activity of the different input devices was tuned either by changing the constitutive promoter (INPUT1) or by exogenously adding different inducer amounts to the bacterial cultures (IPTG for INPUT2 and HSL for INPUT3). The response of the logic inverter output device was measured in terms of RPU via GFP, thus yielding an input-output transfer function for each of the three input devices used. The three transfer functions should be identical, as input systems in a modular framework are functionally interchangeable. Fig. 2.6D shows the three resulting input-output curves for the logic inverter, as well as the identified parameters of the curve fitting. The V_{max} and n values showed a modest variation among the three conditions (CV of 6% and 17% respectively), while the K_m values showed a much higher variation (CV of 44%). Such high difference of the switch point-related parameter may be caused by the input promoters activity variation from the individually characterized context to the interconnected circuit.

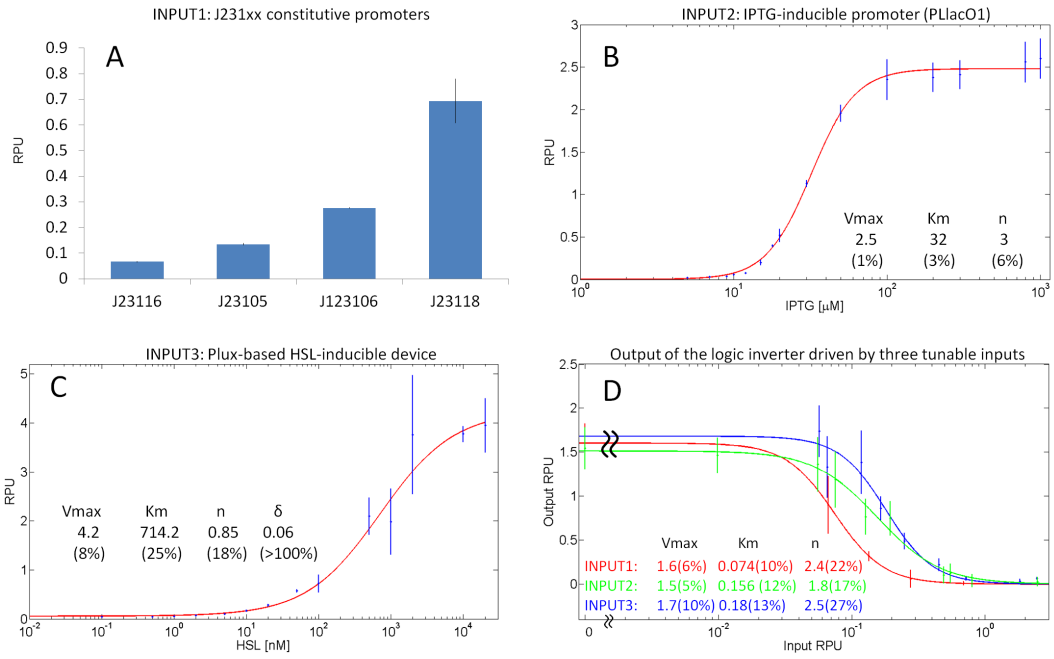


Figure 2.6: Individual characterization of the three investigated input modules (A, B, C) and logic inverter transfer function when driven by the different inputs (D) in KRX strain. Input modules were individually characterized in terms of RPU via RFP34, while the logic inverter was measured via GFP32. Data points are the mean activities computed on three clones and error bars represent the standard deviations. Induction and input-output curves were fitted (solid line) and identified parameters are reported with their estimated CV in brackets. The δ parameter of INPUT2 and output curves was fixed to zero.

2.3 Conclusions

The apparent unpredictability of genetic circuitry is one of the five *hard truths*, recently captured as the major challenges for synthetic biology, that nowadays prevent the fully rational design of biological systems following engineering principles [8]. When parts are put together, they may not behave as expected for several reasons, like the incompatibility of two or more parts or the excessive overloading of transcriptional/translational machinery of the host cell. The term 'retroactivity' has also been introduced to define the unwanted characteristics change of a component upon interconnections and this phenomenon is currently under investigation by a number of research groups [56, 63]. Software tools have been developed that aid the bottom-up design of genetic circuits and improve their predictability [46]. Experimental studies of context-dependent variability and modularity have also been reported for biological parts to elucidate the predictability boundaries of components behaviour [50, 60]. Here, the aim of this work was to expand such experimental investigations by providing useful data on context-dependent variability of transcription-based components and testing the modularity of devices with simple *ad-hoc* constructed model systems in *E. coli*. The RPU approach was applied in order to generate reproducible results among the experiments and to enable the sharing of the data presented in this work, as they are expressed in standard measurement units. To compute RPU, absolute promoters activity was divided by the absolute activity of a reference promoter.

The first preliminary goal was to estimate the activity variation for a set of five popular promoters when measured in TOP10 strain via different measurement systems: three different reporter devices were assembled to the promoters in a low copy vector and were used to quantify their activity *in vivo*. The reference promoter was quantified via the same measurement system, so if promoters activity does not change with the reporter device assembled downstream, the studied promoters should theoretically yield the same RPU activity values when measured via different devices. Only one of the five promoters showed

a significant activity difference among the three tested reporter devices, yielding a CV of 22%. Plasmid copy number dependence was also studied for the tested promoters and results showed that the RPU activity of the strongest promoters is significantly weaker (up to 4.4-fold) in high copy than in low copy context. This result is in accordance with other studies [54, 60]. It confirms the unsuitability of high copy vectors for the characterization of biological parts, because the activity of components with a high energy demand could be underestimated when present in a huge number of copies per cell.

The second goal was to estimate how much the RPU activity changes when a promoter with a given reporter device is moved from the individual context to a complex system composed by two non-interacting gene expression cassettes in the same plasmid. Results showed a maximum activity variation of 33% and 27% for TOP10 and KRX strains respectively.

In these two studies, some of the activity variations could be explained by the physical un-insulation of promoters, in fact it is known that upstream and downstream sequences can significantly change the activity of a promoter. This can explain the variation in the RPU activity of individually characterized promoters when the downstream reporter device is varied; it can also explain the activity variation of a promoter (relative to the individual characterization case) when the upstream sequence is changed by assembling another expression cassette in this position. However, not all the observed variability could be explained by un-insulation, in fact, considering the systems with two combined cassettes in the same plasmid, the activity of the promoter driving the 'upstream' cassette could differ from the one in the individually characterized context, even if the promoter had the same upstream and downstream sequences. Such effect has also been found by Hajimorad et al. [60] while testing the superposition of the effects of different independent gene expression cassettes in the same plasmid.

The third goal of this work was to evaluate the interchangeability of different input modules driving the same output device. The transcriptional sig-

nals generated by the investigated input components were measured in terms of RPU via RFP and then the components were used to regulate a tetR-based logic inverter with GFP downstream. Three input modules were tested: four constitutive promoters of different strengths (considered as one module, transcriptionally tunable by changing the promoter itself), an IPTG-inducible promoter and an HSL-inducible device. In a modular framework, the three steady-state input-output transfer functions of the logic inverter should be the same, as devices driven by identical signals should yield identical outputs. However, although the obtained input-output curves had similar shapes, their switch point showed a CV of 44% among the conditions. One or more of the above mentioned factors could be responsible for this variability, e.g. the different RBS-gene downstream of the promoter in characterization and test stages could cause a downstream sequence-dependent activity change of the input.

Previously published works [25, 53, 61, 64] also reported successful modularity of input devices in interconnected systems, such as logic gates, actuators or feed-forward circuits. Although the mentioned works proved the correct functioning of complex and valuable systems, the modularity assessment often relied on qualitative behaviour comparisons, without providing variability indexes for quantitative comparisons.

The overall variability results of this work suggest that increasingly complex conditions yield an increasing variability in components activity. Designers of synthetic biological systems must take into account such variability entity when dealing with the bottom-up composition of systems. In particular, interconnected modules have a high importance in this design process, as the tuning of gene expression by different constitutive and inducible promoters is a major task and the knowledge about the variability of this operation is necessary.

Finally, it is worth noting that in the model systems considered here, promoters have very similar predicted transcription start sites, so, given a specific downstream device, the produced mRNA should be the same for all the pro-

moters [37, 44, 61]. All the tested cultures had a growth rate similar to the strain bearing the standard reference promoter (data not shown). Moreover, the mRFP1, GFPmut3b and tetR genes, used to characterize promoters and to interface devices, have comparable lengths. Taken together, these features contribute to the simplification of the designed model systems, while deviations from these conditions still have to be tested and may yield higher context-dependent variability. In general, nonlinearities in components activity can be function of these and other factors, like the codon usage of the expressed genes, promoters/RBSs strength and bacterial strain used, so similar studies should be conducted to investigate such factors.

Chapter 3

Characterization of a synthetic bacterial self-destruction device

Datasheets are highly important tools in the engineering world and their production can be useful to support the re-use of components in synthetic biological systems. In the philosophy of providing datasheets for biological parts, this chapter reports the quantitative characterization of a biological actuator which is able to trigger cell lysis in *E. coli*.

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 parts for cell lysis have been developed in literature and their quantitative characterization is an essential step to enable the engineering of synthetic lytic systems with predictable behaviour.

Here, a 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 a well characterized N-3-oxohexanoyl-L-homoserine lactone (HSL) -inducible promoter and the transfer function, lysis dynamics, protein release capability and genotypic and pheno-

typic stability of the device have been evaluated. Finally, its modularity has been tested by assembling the device to a different inducible promoter, which can be triggered by heat induction. A detailed background section on the state of the art of devices and tools to engineer cell lysis will be reported (3.1) and the results about the characterization of the considered device will be illustrated (3.2). According to the obtained results, applications for such device will be discussed and the conclusion notes for this study will be provided (3.3). The contents of this chapter have been published in [64].

3.1 Background

Naturally occurring lytic and temperate bacteriophages have the ability to provoke the host cell lysis through the expression of specific proteins during the lytic cycle. In many phages, like the T4 phage and the lambda phage, these proteins have been identified and widely studied [65]. In particular, holins form stable and non-specific lesions in the cytoplasmic membrane that allow the lysozymes to gain access to the peptidoglycan layer. Lysozymes are generally soluble proteins with one or more muralytic activities against the three different types of covalent bonds (glycosidic, amide, and peptide) of the peptidoglycan polymer of the cell wall [66, 67]. The combined work of holin and lysozyme results in the degradation of the two cell membranes of gram-negative bacteria, thus causing cell lysis. Antiholin is a third protein involved in this process as it inhibits holin and is responsible for the regulation of its activity [68]. The described lytic mechanism can be exploited for the release of useful recombinant proteins which cannot be secreted by the engineered host strain [69].

E. coli is a widely used organism for recombinant protein production, but its secretion capabilities are limited and recombinant protein targeting to the growth medium has shown to work only with a small set of proteins [70]. For this reason, cell disruption techniques are required to gain the intracellularly

expressed protein of interest. Mechanical techniques, such as cell ultrasonication, usually result in protein denaturation caused by the heat produced during the process and some of them are also unfeasible on industrial scale, whereas non-mechanical techniques, such as chemical membrane degradation with detergents or enzymes, involve the purchase of expensive reagents [69]. The engineering of a lysis system that is triggered by a user-defined signal can avoid the use of common cell disruption techniques for the recovery of intracellularly expressed proteins.

Another important application of an inducible lytic system is the programmed cell death of a bacterial population, which might be useful in those processes where the microorganism must be eliminated at a specific time, after having completed its work.

In literature, inducible lysis systems have been proposed. T4 phage holin and T7 phage lysozyme genes have been used to construct lytic *E. coli* strains to achieve the gentle disruption of cells upon IPTG induction using the P_{lac} promoter or the DE3 inducible system [69, 71]. The T7 lysozyme was used both to cut bonds in the cell wall and to tightly regulate holin gene by inhibiting the T7 polymerase basal expression in uninduced DE3 inducible system. The same genes have been used under the control of a glucose starvation-inducible promoter to allow cell autolysis upon glucose exhaustion in the medium [72]. Heat- and UV-inducible promoters have been used to regulate the lambda phage lysis cassette *SRRz* for high throughput enzyme release [73].

The Registry of Standard Biological Parts hosts several lysis protein coding sequences and devices, but most of them, despite their important potential applications, remain uncharacterized. The BioBrick device BBa_K112808 consists of a promoterless operon composed by the T4 phage genes *t* and *e*, encoding a holin and a lysozyme respectively [74]. Downstream of the transcriptional terminator of the operon, the T4 rI gene, encoding a *t*-specific antiholin, is present under the control of a weak constitutive promoter (see Fig. 3.1A for a detailed overview of this lysis device). The T4 antiholin is

able to heterodimerize with the t gene product, thus preventing the holin from forming pores in the inner membrane [75]. When the lysis device is assembled under the control of an inducible promoter, a weak constitutive expression of rI prevents cell lysis, which may be caused by the basal expression of the promoter in the OFF state, and in this way the tight regulation of the inducible lysis can be obtained. In this work, BioBrick parts have been used to quantitatively characterize the BBa_K112808 lysis device by assembling it to BBa_F2620, a well characterized N-3-oxohexanoyl-L-homoserine lactone (HSL)-inducible device based on the P_{lux} promoter (see Fig. 3.1B for a working diagram of this device) [40], thus yielding the new BioBrick BBa_K173015. BBa_F2620, whose transfer function has already been measured, has been used to drive the expression of BBa_K112808 over a range of transcriptional input values and to characterize the lysis behaviour as a function of the transcriptional strength in *E. coli*. Finally, in order to study the modularity of this device, a different inducible input device, which can be triggered by heat induction (see Fig. 3.1C for a working diagram of this device), has been assembled upstream, thus yielding the new BioBrick BBa_J107014, and the validation of this composite part has been carried out. Fig. 3.1D shows the maps of the three main plasmids used in this work, containing the promoterless lysis device (pLC-T4Lys$^-$), the HSL-inducible lysis device (pLC-T4LysHSL) and the heat-inducible lysis device (pLC-T4LysHeat) respectively, all of them in a low copy number vector.

3.2 Results

3.2.1 Lysis assays

Lysis was assayed in a 96-well microplate by measuring the optical density at 600 nm (OD_{600}) dynamics of TOP10 bearing HSL-inducible lysis device pLC-T4LysHSL, induced with HSL and uninduced. Induced and uninduced TOP10 bearing the promoterless lysis device pLC-T4Lys$^-$ were used as a negative control in all the experiments. Lysis entity, i.e. the maximum percent

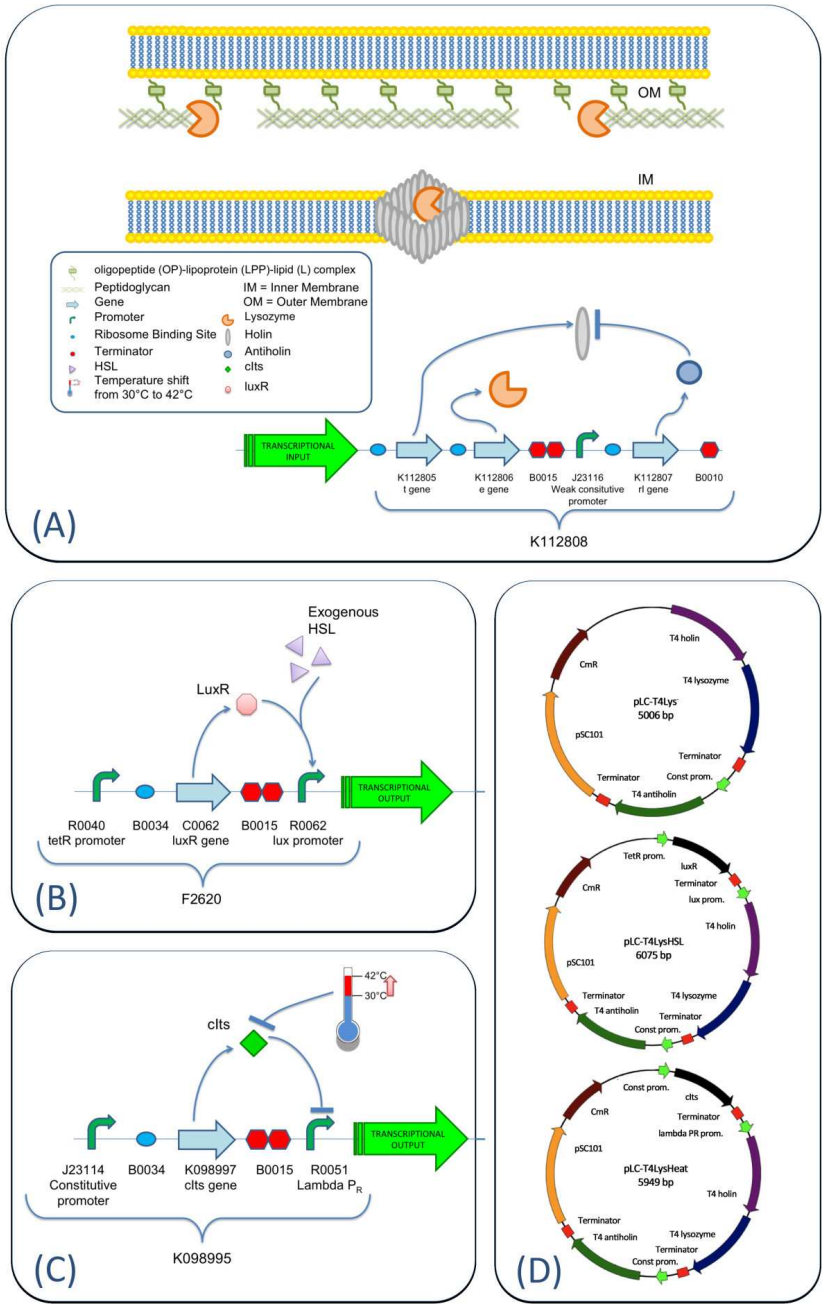


Figure 3.1: Overview of the devices used in this work. The lysis device is composed by a holin (t)-lysozyme (e)-encoding operon and a weakly expressed antiholin rI gene (A). BBa_F2620 HSL-inducible input device is composed by a luxR expression cassette, driven by the P_{tetR} promoter, and the P_{lux} promoter, which is normally off. Its transcription can be induced by the LuxR transcription factor in presence of HSL (B). BBa_K098995 heat-inducible input device is composed by a cIts expression cassette, driven by the BBa_J23114 constitutive promoter, and P_R repressible promoter from lambda phage. cIts is a heat-sensitive repressor of P_R: when temperature is 30°C the repressor keeps the promoter in the OFF state, while a temperature shift to 42°C can induce the transcription of P_R (C). Plasmid maps of the promoterless lysis device pLC-T4Lys$^-$, the HSL-inducible lysis device pLC-T4LysHSL and the heat-inducible lysis device pLC-T4LysHeat, all of them in a low copy vector (D).

reduction of OD_{600} caused by lysis induction, was computed in all the assays as an indirect measurement of the amount of lysed cells.

A typical lysis profile is reported in Fig. 3.2, where TOP10 bearing pLC-T4LysHSL were induced with HSL 100 nM at t = 0 h ($OD_{600} = 0.2$, exponential phase), t = 4 h ($OD_{600} \sim 1.3$, early stationary phase) or t = 20 h ($OD_{600} \sim 2$, 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 = 0 h, t = 4 h and t = 20 h, respectively (Tab. 3.1).

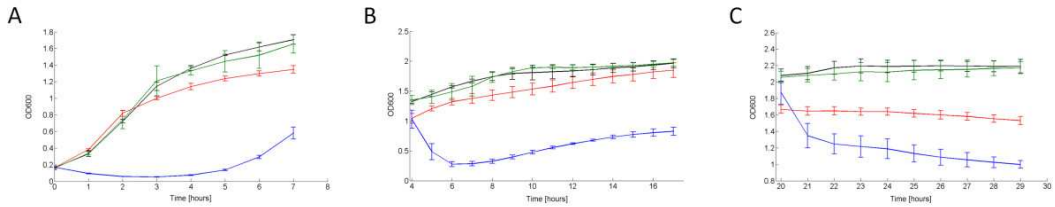


Figure 3.2: Lysis profile of TOP10 bearing the HSL-inducible lysis device in low copy number when induced in different growth phases in microplate reader. OD_{600} of TOP10 with pLC-T4LysHSL induced with HSL 100 nM (blue line) and uninduced (red line), pLC-T4Lys$^-$ induced with HSL 100 nM (green line) and uninduced (black line). Induction was performed in exponential phase (A), early stationary phase (B) and late stationary phase (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.

TOP10 with pLC-T4LysHSL

Growth phase:	**Exponential**	**Early stationary**	**Late stationary**
Lysis entity [%]	76.28±0.3	75.43±1.1	50.1±2.5
Lysis delay after induction [min]	15	15	15
Doubling time [min]	49.8±1		
Doubling time of negative control [min]	43±1.3		

Table 3.1: Quantitative characterization of TOP10 bearing the HSL-inducible lysis device in low copy plasmid grown at 37°C in microplate. Induction was carried out with HSL 100 nM. Mean lysis entity and lysis delay after the induction are reported for the different growth phases together with their standard error, measured on 3 independent experiments. The doubling time of the uninduced lysis device and its negative control are reported too.

The mean doubling time of uninduced TOP10 bearing pLC-T4LysHSL, evaluated on all the experiments, was 49.8 ± 1.1 min, while the doubling time of the uninduced negative control pLC-T4Lys$^-$ was 43 ± 1.3 min, thus demon-

strating 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 (Tab. 3.1). In all cases, after about 2-3 h from the induction, cells start growing again, suggesting the onset of mutants that have lost the inducible lysis phenotype.

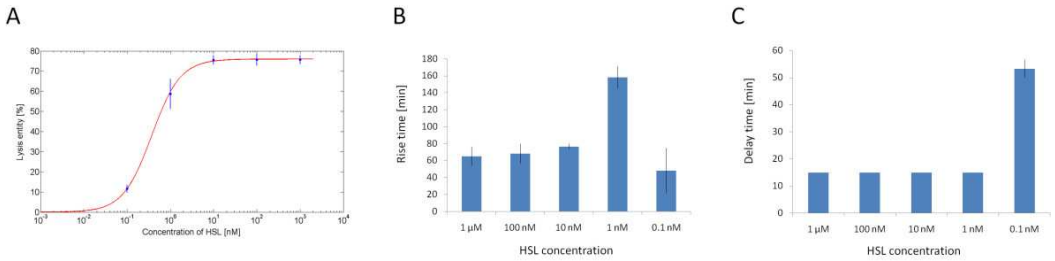


Figure 3.3: Transfer function, rise time and delay time of the HSL-inducible lysis device in low copy plasmid in early stationary phase (OD_{600} = 0.9) in microplate reader. Lysis entity of TOP10 cells with pLC-T4LysHSL induced with different concentrations of HSL (A). The experimental data (circles) were fitted with a Hill function (line, V_{max} = 76, K_{50} = 0.37, n = 1.3). For each concentration, the rise time, i.e. the time to rise from the 10% to 90% of the lysis entity (B) and the delay time before the OD_{600} drop after induction (C) are also shown. Error bars represent the 95% confidence interval of the estimated mean.

Fig. 3.3A shows the measured transfer function of pLC-T4LysHSL induced at t = 4 h (OD_{600} = 0.9, early stationary phase) with different HSL concentrations. It demonstrates that lysis entity can be effectively tuned. The rise time, i.e. the time required to rise from 10% to 90% of the lysis entity, and the delay time before the OD_{600} drop are also reported for each HSL concentration (Fig. 3.3B and C). As these parameters show, the lysis dynamics is highly nonlinear, in fact the delay and rise times change as a function of the induction entity. In particular, the delay time is equal for all the HSL concentrations except for the smallest one (0.1 nM), in which it is about 40 min longer than the one measured for the other concentrations. The rise time increases as a function of the HSL concentration, reaching its highest value at 1 nM HSL (more than 2.5 hours). However, at 0.1 nM the rise time is less than 1 hour.

Analogous studies were conducted for the HSL-inducible lysis device in high copy plasmid. However, for its evident instability and for the too high doubling time of uninduced cells (see C.2.2 in Appendix C), the lysis device

in high copy plasmid has not been considered for further studies.

Lysis was also assayed for TOP10 bearing pLC-T4LysHeat low copy vector containing a heat-inducible lysis device. Induction was triggered in a 96-well microplate by shifting the incubation temperature from 30°C to 42°C. The results are shown in Fig. 3.4 and summarized in Tab. 3.2. The lysis entity was comparable to the one in the HSL-inducible device in low copy. The doubling time of the heat-inducible lysis device was similar to the negative control, but both were much higher than those reported in Tab. 3.1 because cultures were grown at 30°C instead of 37°C, causing a slower growth rate. Also the time delay after induction in the thermoinducible lysis device is much higher than in the HSL-inducible device. This should be due to the different response time of the two input devices.

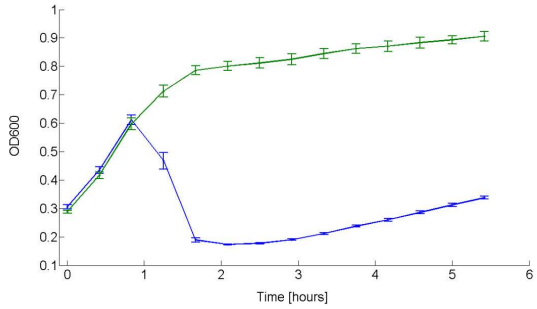


Figure 3.4: Lysis dynamics of TOP10 bearing the thermoinducible lysis device in low copy plasmid grown in microplate reader. OD_{600} of TOP10 with pLC-T4LysHeat induced with a temperature shift from 30°C to 42°C in the microplate reader (blue line). Heat-induced pLC-T4Lys$^-$ (green line) is shown as the negative control. Induction was performed in exponential phase at $OD_{600} = 0.3$. Error bars represent the 95% confidence interval of the estimated mean. For clarity of presentation, data points shown here are resampled with a 30-minute sampling time.

TOP10 with pLC-T4LysHeat	
Lysis entity in exponential phase [%]	73±0.3
Lysis delay after induction [min]	55
Doubling time [min]	66.1±1.7
Doubling time of negative control [min]	65.9±0.6

Table 3.2: Quantitative characterization of TOP10 bearing the thermoinducible lysis device in low copy plasmid grown at 30°C in microplate. Induction was carried out by shifting the temperature from 30°C to 42°C. Mean values estimated on three wells in the same experiment are reported with standard errors.

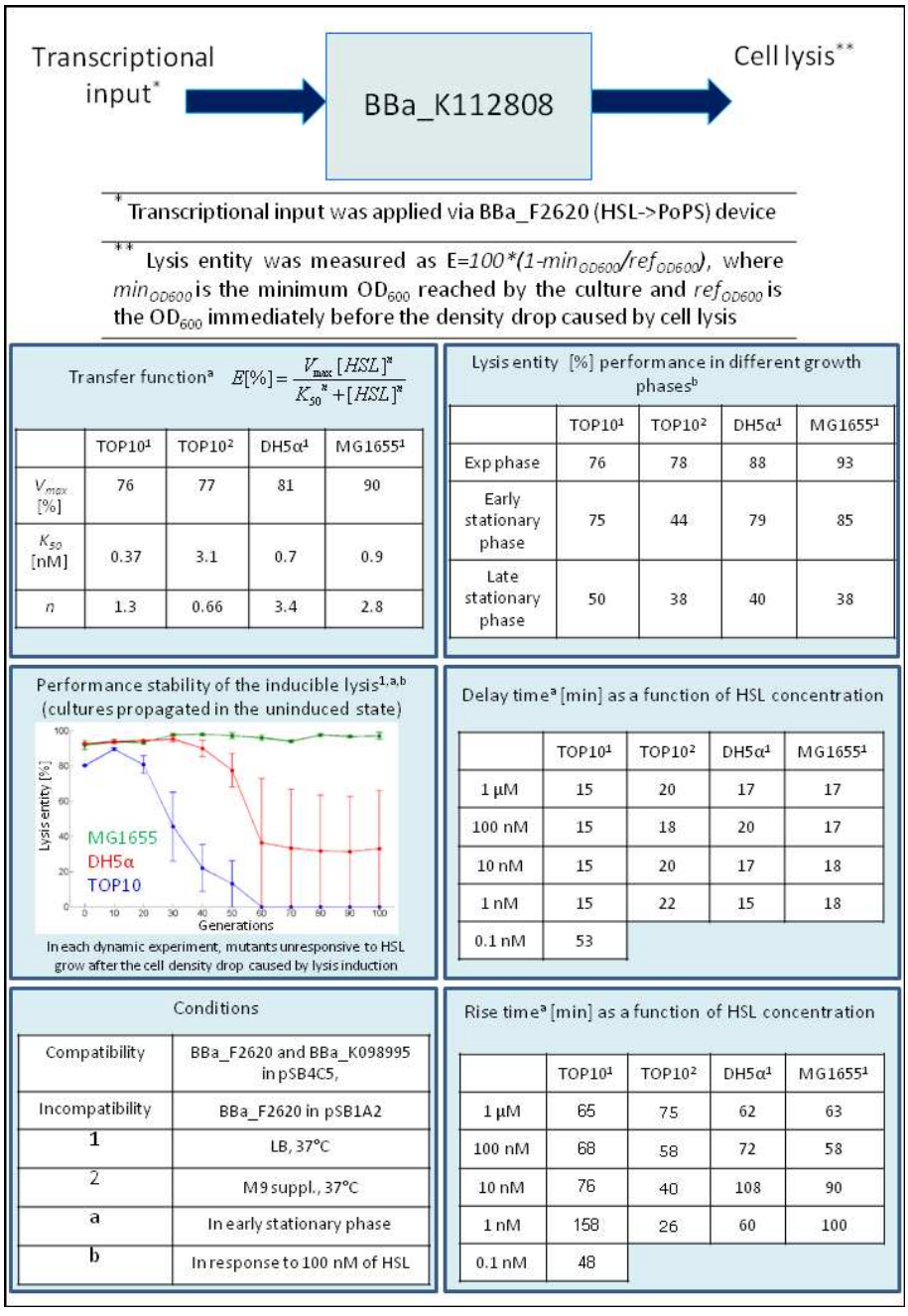
Experiments on pLC-T4LysHSL and pLC-T4LysHeat were also performed in a different growth medium (M9 with glycerol as carbon source, supplemented with thiamine and casamino acids) and in two other *E. coli* strains (DH5α and MG1655) grown in LB, giving consistent results when compared to the TOP10 strain in LB medium results described in this section (see [64] for a detailed description of the results). Most of the obtained quantitative results are summarized in a specific datasheet for the BBa_K112808 device, which is shown in Fig. 3.5.

3.2.2 Protein release assays

A co-transformed TOP10 strain (here called TOP10-rfp-lys) bearing both a high copy plasmid with a Red Fluorescent Protein (RFP) constitutive expression cassette pHC-RFP and the HSL-inducible lysis device pLC-T4LysHSL was induced with HSL to study the RFP release in the medium. Uninduced TOP10-rfp-lys, TOP10 bearing pHC-RFP (induced and uninduced) and TOP10 bearing pLC-T4Lys$^-$ were chosen as controls in this assay. Fig. 3.6A and B show respectively the OD_{600} during the experiment and the fluorescence of the supernatant, which is proportional to the RFP molecules released in the growth medium. It is evident that only TOP10-rfp-lys + HSL 100 nM could lyse and that cell lysis was accompanied by RFP release in the culture supernatant. For this culture, after 125 minutes from induction time, 96 $\pm$ 0.04% of the RFP molecules had been released in the growth medium (see Tab. 3.3). Surprisingly, the three negative control cultures released about 25% of the molecules.

3.2.3 Analysis of mutants

After performing lysis assays, the lysed cultures grew again (see Section 3.2.1). When the re-grown cultures of TOP10 bearing pLC-T4LysHSL were diluted 1:1000 in fresh selective LB medium and let grow to an OD_{600} = 0.35 (expo-



	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 3.5: Datasheet produced for BBa_K112808.

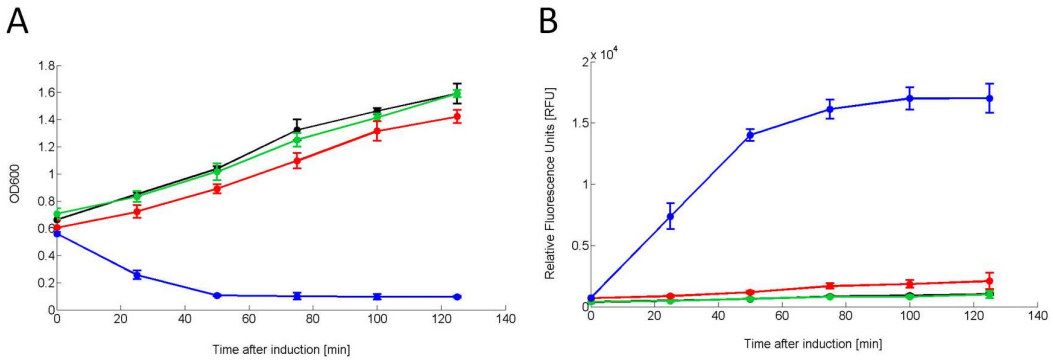


Figure 3.6: OD_{600} time course of TOP10-rfp-lys grown in 15-ml tubes upon induction with HSL 100 nM (A) and RFP fluorescence time course in the supernatant (B). Culture absorbance (A) and supernatant fluorescence (B) of TOP10-rfp-lys induced with HSL 100 nM (blue line). Uninduced TOP10-rfp-lys (red line), TOP10 bearing pHC-RFP induced with HSL 100 nM (green line) or uninduced (black line) are the negative controls. Induction was carried out in the exponential phase at $OD_{600} \sim 0.55$. Error bars represent the 95% confidence interval of the estimated mean.

RFP release in a lytic strain

Strain	**Secreted RFP** [%]
TOP10-rfp-lys	24.84±0.6
TOP10-rfp-lys + HSL 100 nM	95.95±0.04
TOP10 with pHC-RFP	25.38±1.8
TOP10 with pHC-RFP + HSL 100 nM	24.68±0.8

Table 3.3: RFP release efficiency of TOP10-rfp-lys cultures grown at 37°C in 15-ml tubes and induced with HSL 100 nM. Mean values estimated on three tubes in the same experiment are reported with standard errors.

nential phase), they did not lyse upon induction with HSL 10 nM, suggesting that the cells have completely lost the inducible lysis phenotype (data not shown). The restriction analysis of the mutant plasmids after DNA digestion with EcoRI-PstI is reported in Fig. 3.7.

Plasmids were purified from two single colonies isolated from four mutant cultures. One of these cultures was obtained starting from 5 μl of glycerol stock (here called mut_{gly} culture), while the other three cultures were obtained starting from single colonies streaked from the glycerol stock (here called mut_{sc1}, mut_{sc2} and mut_{sc3} cultures respectively). In all the mutant clones the pSB4C5 vector band (∼3.2 kbp) was present, but the inserts were highly different from

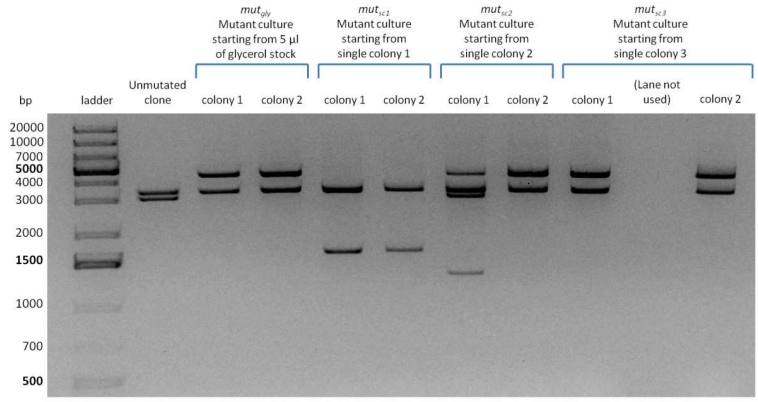


Figure 3.7: Restriction analysis of pLC-T4LysHSL mutants. Plasmid DNA was digested with EcoRI and PstI. In all the screened clones two bands (vector backbone and insert) are present except in colony 1 of mut$_{sc2}$, in which four bands can be observed. As sequencing showed, colony 1 of mut$_{sc2}$ had plasmids carrying two different mutated inserts in the same clone, one of which containing an EcoRI restriction site that, when digested, produces 2 bands.

the unmutated culture insert (∼2.8 kbp), suggesting that mutations occurred just in the HSL-inducible lysis device. The sequencing of the clones disclosed all the mutations, as shown in Fig. 3.8. Two clones (colony 2 of mut$_{gly}$ and colony 2 of mut$_{sc3}$) showed the insertion sequence IS10R after the nucleotide 576 of luxR (BBa_C0062) gene (Fig. 3.8B); two clones (colony 2 of mut$_{sc2}$ and colony 1 of mut$_{sc3}$) showed the insertion of IS10R in the same place as before, but in the opposite direction (Fig. 3.8C); two clones (colony 1 and 2 of mut$_{sc1}$) showed the deletion of the DNA comprising the P_{lux} promoter and the holin-lysozyme operon (Fig. 3.8D); one clone (colony 1 of mut$_{gly}$) showed the insertion of IS10R after the nucleotide 318 of the holin (BBa_K112805) gene (Fig. 3.8E); one clone (colony 1 of mut$_{sc2}$) had two different mutated plasmids in the same colony: one of them showed the insertion of IS10R before the nucleotide 1 of the luxR RBS (BBa_B0034), while the other showed the insertion of IS5 before the start codon of luxR (Fig. 3.8F). All the described mutations cause the impairment of the lysis phenotype. The molecular weights of the restriction fragments shown in Fig. 3.7 are all consistent with the found mutations.

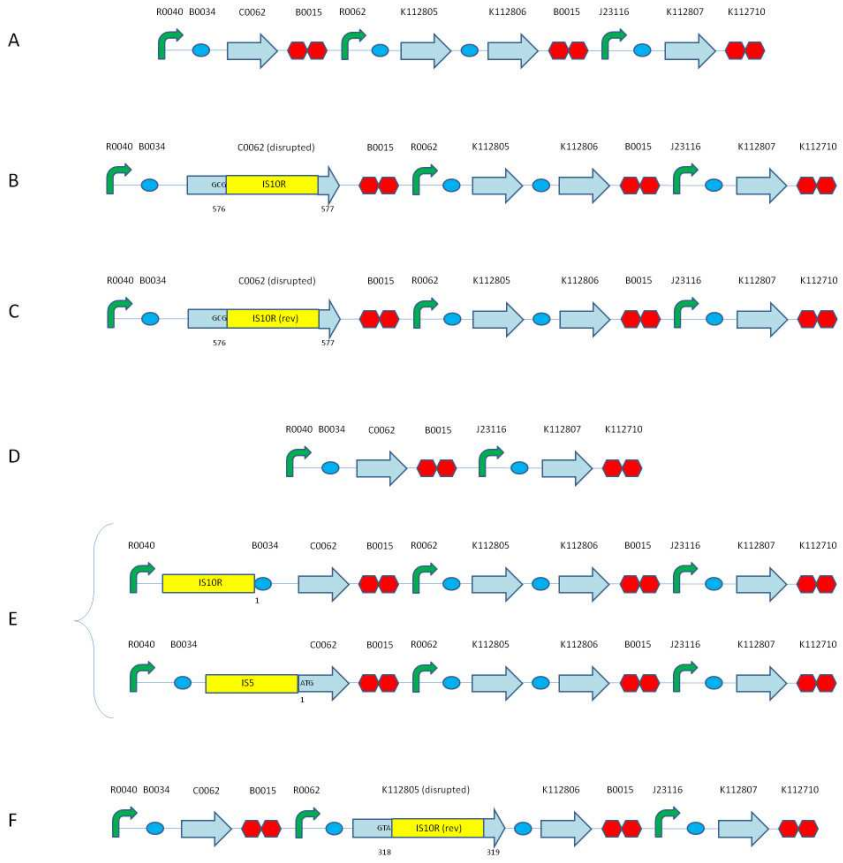


Figure 3.8: Mutations occurred in the insert of pLC-T4LysHSL in mutant clones. Unmutated BBa_K173015 (A); luxR disruption mediated by IS10R (B); luxR disruption mediated by IS10R (insertion in reverse direction relative to (B)) (C); deletion of the DNA fragment flanked by BBa_B0015, containing the P_{lux} promoter and the holin and lysozyme genes (D); two different mutated plasmids in the same clone: insertion of IS10R and IS5 (containing an EcoRI restriction site) upstream of the first nucleotide of BBa_B0034 RBS and luxR, respectively (E); t gene disruption mediated by IS10R (insertion in reverse direction relative to (B)) (F). The number under the parts denotes the nucleotide of the basic part flanking the insertion sequence that has disrupted the part. The disrupted genes are luxR (BBa_C0062), encoding the transcriptional activator of P_{lux} promoter (BBa_R0062), and t (BBa_K112805), encoding the holin.

3.2.4 Evolutionary stability of bacteria bearing pLC-T4LysHSL

In order to study the reliability of the lysis device when cell disruption is not triggered, bacteria bearing pLC-T4LysHSL were propagated for 100 generations without HSL and their inducible lysis phenotype was tested over time in terms of lysis entity. The datasheet in Fig. 3.5 shows the results of this study in TOP10, DH5α and MG1655 grown in selective LB. Only MG1655 could maintain the phenotype of interest for 100 generations, thus showing an excellent stability of the uninduced lysis genes. Among these strains, TOP10 was the first to lose the lysis capability, which started to decrease after 30 generations and was completely lost after 60. DH5α showed a high lysis entity variability because two of the three replicates completely lost the lysis capability after 60 generations, while the other one never lost it even in 100 generations, thus giving very wide standard errors of the measured mean.

3.3 Conclusions

In this work, the quantitative characterization of a BioBrick lysis device of the Registry of Standard Biological Parts is reported. Its activity has been measured in *E. coli* using a well characterized HSL-inducible promoter and the transfer function, lysis dynamics, protein release capability, modularity and genotypic and phenotypic stability of the device have been evaluated.

Low copy number has 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. Lysis entity in late stationary phase was lower than in the other growth phases. These results are consistent with the published findings for which *E. coli* membrane disruption mediated by T4 phage lysis gene t or e was more efficient when the genes were expressed in the exponential growth phase than in

the stationary phase, where less or almost no disruption occurred [69, 71, 73]. 96% of the total amount of intracellular proteins was successfully released into the growth medium upon induction of the lysis device in low copy plasmid. These features demonstrate that in this condition the expression of the lysis genes is tightly controlled and makes the device suitable for recombinant protein release upon gentle disruption of cell membranes. However, in all the experiments mutant cells that were unresponsive to induction arose after the bacterial lysis. This intrinsic instability makes the device unsuitable for the programmed cell death of a bacterial population.

Mutant analysis showed that two main classes of DNA modifications occurred to eliminate the HSL-inducible lysis phenotype. The most frequent one consisted in the insertion of IS10R or IS5 within the device sequence to impair the expression of the lysis genes by disrupting their regulatory parts upstream or the lysis genes themselves. The other class exhibited the deletion of genes included between two identical DNA sequences. In particular, the DNA fragment including the P_{lux} promoter and the holin and lysozyme genes was most probably deleted from the plasmid by a replication slippage mechanism between two identical transcriptional terminators (BBa_B0015) of 129 bp flanking the fragment. These mutations were consistent with published DNA mutations commonly occurring in *E. coli* [76], in fact gene disruptions mediated by insertion sequences, which occurred in the majority of the mutant cultures analyzed in this work, have been found to be responsible of 95% of the mutation events in TOP10 strain [77]. On the other hand, replication slippage events have already been found to occur between two BBa_B0015 terminators in a previously studied BioBrick device [39].

The HSL-inducible lysis device stability was also studied during continual bacterial growth for 100 generations without induction. The inducible lysis phenotype started decreasing after only 30 generations and was completely lost after 60, but DH5α and MG1655 strains gave a better performance. In particular, DH5α showed a significantly lower lysis capability after 60 generations, while MG1655 stably maintained it during all the 100 generations.

The described lysis device has been shown to be compatible with other modular input devices. When assembled with a heat-inducible BioBrick device upstream and triggered with a temperature shift to 42°C, the lysis device worked as expected, thus demonstrating the possibility of triggering cell lysis with any transcription-based input device. This feature enables the intriguing possibility to control cell disruption in response to a user-defined exogenous signal.

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 re-usability of this lysis device. Such information is summarized in the biological datasheet reported in Fig. 3.5.

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.

Chapter 4

Mathematical models of genetic circuits for the design of logic functions in a biological chassis

Mathematical modelling is widely used in all the fields of engineering to guide the design of systems composed by a set of well-characterized parts. In this way, parameters can be tuned *in silico* and suitable components can be selected to be included in the system. For the same reasons, modelling can also play an important role in the design of biological functions. Even if modularity is not always valid for biological components and unwanted behaviour of parts could be exhibited when put in a different system, modelling can be used to guide the choice of components that give the desired output, thus avoiding a trial-and-error approach.

This chapter shows the mathematical modelling of two proof-of-concept biological functions, designed to mimic the behaviour of logic functions: a multiplexer (mux) and a demultiplexer (demux). Mathematical models were defined for the two desired systems and, after a preliminary design and construction of genetic devices that could implement the desired functions, a set of parameters were identified from experimental data. Finally, the models were

exploited to propose future modifications in the designed circuits in order to yield the desired functions.

In particular, the mux and demux logic functions will be introduced and a design of the basic logic gates and the final logic functions will be provided in a biological context (4.1). Experimental results about the physically constructed circuits and model simulations will be shown, highlighting the predicted working conditions of the designed devices and the crucial parameters (4.2). Finally, the overall results of the present study will be discussed and guidelines to yield functional mux and demux devices will be provided (4.3). The contents of this chapter have been published in [62] and [78].

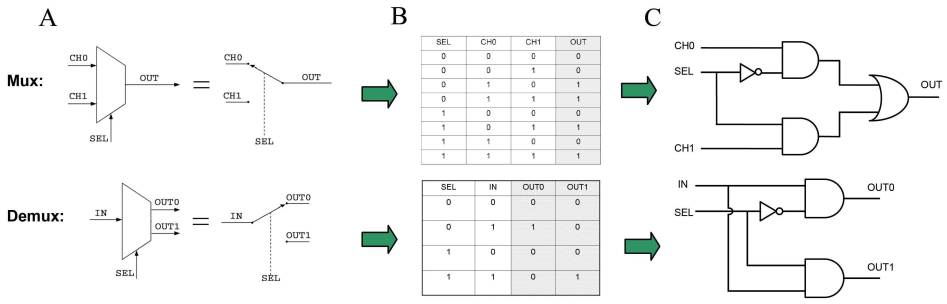
4.1 Systems design

4.1.1 Multiplexer and demultiplexer

Multiplexing is a function in which one of multiple input signals is conveyed into a single output channel, whereas in demultiplexing a single input signal is conveyed into one of multiple output channels. The choice of the input channel in multiplexing and of the output channel in demultiplexing is controlled by a selector. The devices implementing these two functions are called multiplexer (mux) and demultiplexer (demux), respectively, and can be considered as controlled switches [79]. They have a remarkable importance in electronic, telecommunication, and signal processing systems, such as central processing units or analog to digital converters.

Herein, only two-input mux (mux 2:1) and two output demux (demux 1:2) devices were considered. Furthermore, only digital ON/OFF signals, i.e. signals that could only assume Boolean values (0 or 1), were considered. In this framework, mux and demux can be considered as logic functions and using truth tables and Karnaugh maps methods it is possible to design a modular logic network for each of the devices [80]. In particular, the overall behaviour

of mux and demux is summarized in Fig. 4.1. According to their logic functions, the basic logic gates composing the networks are AND, OR and NOT for mux, AND and NOT for demux.



SEL	CH0	CH1	OUT
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

SEL	IN	OUT0	OUT1
0	0	0	0
0	1	1	0
1	0	0	0
1	1	0	1

Figure 4.1: Mux and demux schematic structures (A), truth tables (B), and logic networks obtained by Karnaugh map method (C). CH0 and CH1 are the two input channels of mux; OUT is the output channel of mux; IN is the input channel of demux; OUT0 and OUT1 are the two output channels of demux; SEL is the selector. In the truth tables, 1 and 0 represent ON and OFF signals, respectively.

The goal of this work is to design genetic networks that could mimic mux and demux in a biological chassis, such as *E. coli*, using standard BioBrick parts available in the Registry of Standard Biological Parts and to discuss the working conditions of these devices, also thanks to the definition and the implementation of suitable mathematical models. Protein/DNA and autoinducer/protein interactions were used to build up the required biological logic gates.

4.1.2 Design of biological gates

Fig. 4.2 summarizes, from both a digital electronics and a biological point of view, the proposed AND (Fig. 4.2A), OR (Fig. 4.2B), and NOT (Fig. 4.2C) gates.

In particular, the AND gate was implemented using quorum-sensing complexes, such as the lux operon from *Vibrio fischeri* [81] and las operon from *Pseudomonas aeruginosa* [82]. In the lux system, the P_{lux} promoter (BBa_R0062) can be activated only by the simultaneous expression of luxR (BBa_C0062)

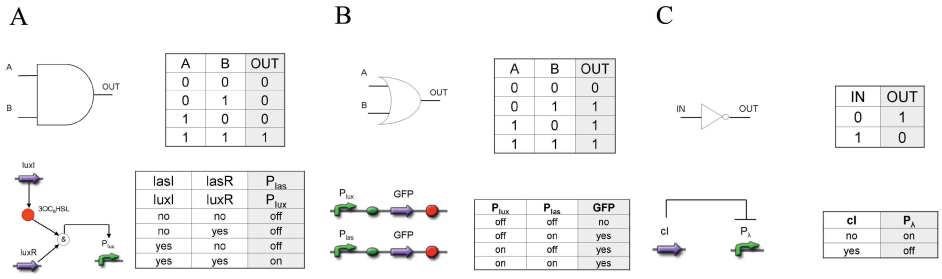


Figure 4.2: Logic gate design. The AND gate was implemented with the lux (or las) system. Only the simultaneous presence of LuxI (or LasI) and LuxR (or LasR) can activate P_{lux} (or P_{las}) (A). The OR gate was implemented by replicating the gene encoding the function of interest downstream of two independent promoters (B). The NOT gate was implemented through the cI-P_λ repressible system (C).

and luxI (BBa_C0061) genes, which encode the LuxR transcription activator and the synthase of the autoinducer N-3-oxohexanoyl-L-homoserine lactone ($3OC_6HSL$), required to activate LuxR [3, 83]. The las system shows similar behaviour: it is composed of the P_{las} promoter (BBa_R0079) and lasR (BBa_C0079) and lasI (BBa_C0078) genes [84, 85, 86], which work through the N-3-oxododecanoyl-L-homoserine lactone ($3OC_{12}HSL$) autoinducer and can implement an AND gate as well [26].

A biological OR gate was implemented simply by replicating a gene of interest under the control of two independently inducible promoters, P_{lux} and P_{las}.

A biological NOT gate was implemented using a λ-cI system from lambda phage [87]. This system includes the P_R (also called P_λ) repressible promoter (BBa_R0051), which is normally on, but can be turned off upon the expression of cI gene (BBa_C0051), which encodes a specific repressor for this promoter [28].

4.1.3 Design of biological mux and demux

The interconnection of the described biological logic gates, according to the logic networks shown in Fig. 4.1, gives life to the final circuits that could theoretically yield mux and demux behaviour. To supply two universal devices

that can detect any kind of input signal and that can express any protein as an output, mux and demux circuits were designed without inputs and outputs. In this way, the final users can tailor the circuits by assembling the specific input and output devices for the desired applications, following the guidelines herein. Fig. 4.3 shows the resulting circuits and all the possible interactions between the networks elements.

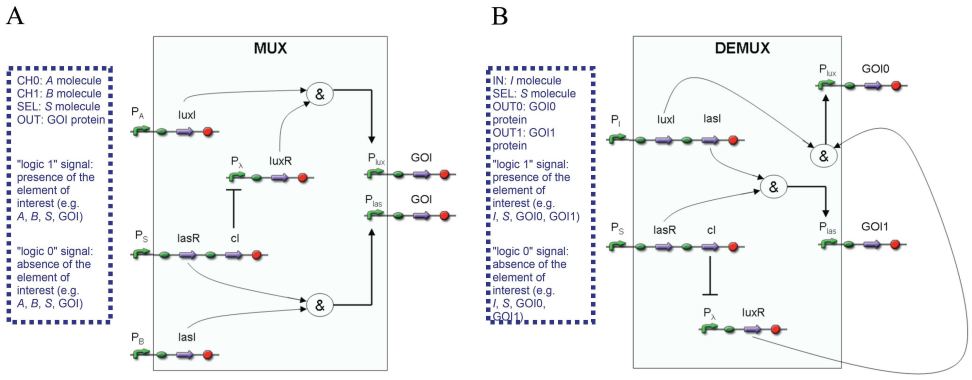


Figure 4.3: Biological mux (A) and demux (B) circuits. P_A, P_B, P_S and P_I indicate promoters activated by the molecules A, B, S and I, respectively. GOI, GOI0 and GOI1 indicate the genes of interest.

In Fig. 4.3, generic input (P_A, P_B, P_S for mux; P_I and P_S for demux) and output (GOI for mux; GOI0 and GOI1 for demux) elements are present. In particular, P_A, P_B, P_S and P_I are promoters that can be activated by the molecules A, B, S and I, respectively, whereas GOI, GOI0 and GOI1 are the genes of interest that encode for the generic output proteins GOI, GOI0 and GOI1, respectively. According to these genetic circuits, if A, B, S and GOI correspond to CH0, CH1, SEL and OUT, respectively, for mux and if I, S, GOI0 and GOI1 correspond to IN, SEL, OUT0 and OUT1, respectively, for demux, the biological truth tables of mux and demux are analogous to the Boolean truth tables given in Fig. 4.1.

Three final biological devices for mux and two for demux (see Fig. 4.4A) were planned to build up the designed circuits. Each planned device has an RBS at the 5' end and a promoter or a transcriptional terminator at the 3' end. For this reason, mux and demux circuits both conform to the polymerase

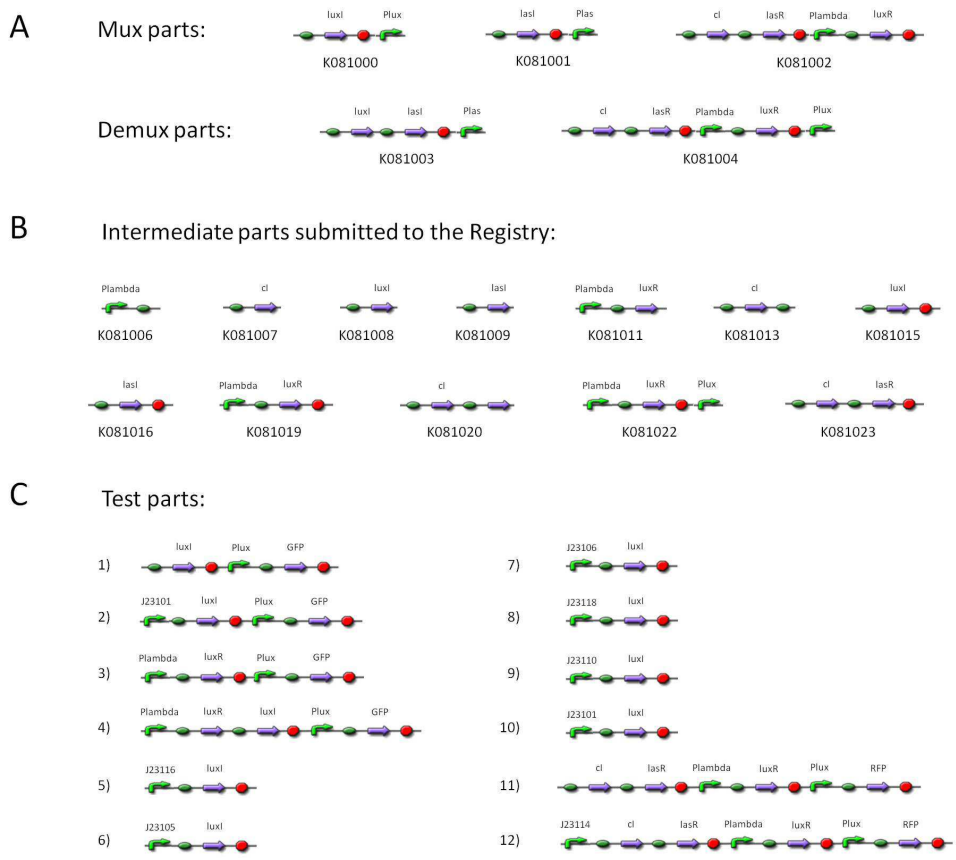


Figure 4.4: (A) Final devices for mux and demux. (B) Intermediate parts submitted to the Registry. (C) Test parts used for subcircuit validation. The reporter genes are GFPmut3b (BBa_E0040) or mRFP1 (BBa_E1010). All the RBSs are BBa_B0030 and all the transcriptional terminators are BBa_B1006, except for test parts 5-10, in which the terminator is BBa_B0015; J231xx are constitutive promoters. All the test parts are in a high copy plasmid.

per second (PoPS) device boundary standard in a multi-input multi-output framework [40].

Herein, mathematical modelling was used to study the theoretical behaviour of the designed networks and experiments performed on a set of constructed genetic subcircuits using GFP or RFP as reporters. These subcircuits are essential parts of both mux and demux and, when interconnected and placed together with the other modules, they may contribute to yield the theoretical behaviour of mux and demux.

4.2 Results

4.2.1 Physical assemblies

All the five planned final devices, which can lead to mux and demux functions, are shown in Fig. 4.4A and were built up and submitted to the Registry of Standard Biological Parts, as well as an exhaustive documentation, which includes user assembly handbooks and gene network schemes and description. BBa_K081000 and BBa_K081001 devices are contained in pSB1A2 high copy number plasmid (pUC19-derived pMB1 replication origin) carrying Ampicillin resistance, whereas BBa_K081002, BBa_K081003 and BBa_K081004 devices are contained in pSB1AK3 high copy number plasmid (same replication origin as pSB1A2) carrying Ampicillin and Kanamycin resistance. Intermediate parts (shown in Fig. 4.4B) were shared and documented in the Registry as well. To validate the behaviour of the implemented circuits, additional composite parts were built up (see Fig. 4.4C) and tested in *E. coli* TOP10 strain.

4.2.2 Parts characterization

The AND gate composed of the lux system was quantitatively characterized. TOP10 with test parts 1-4 allowed the validation of the four rows of the AND gate reported in Fig. 4.2A. This logic gate is expected to give a high output

only when both luxI and luxR genes are expressed. Results show that, under these conditions, the output is about 100-fold higher than the low output values (Fig. 4.5).

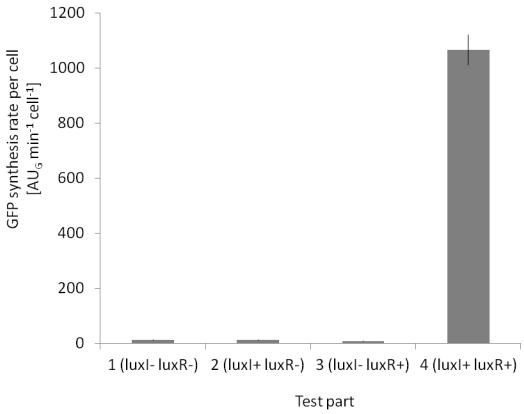


Figure 4.5: Results of the AND gate designed with the lux system. Error bars represent the standard deviation of the measurements of three assayed clones. AU_G indicates arbitrary units of GFP.

Test part 3 was also used to evaluate the behaviour of P_{lux} in the presence of different concentrations of exogenous $3OC_6HSL$ when luxR was constitutively expressed. The static characteristic is reported in Fig. 4.6 as a function of $3OC_6HSL$ and this result is consistent with previous reports [40].

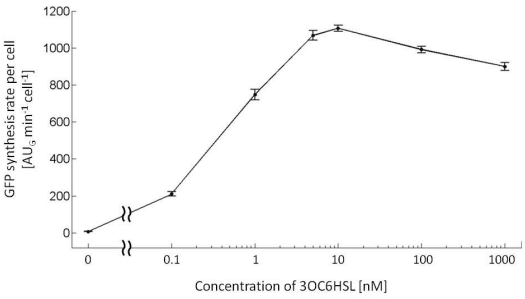


Figure 4.6: Test part 3 behaviour in the presence of exogenous $3OC_6HSL$. Error bars represent the standard deviation of the measurements of three assayed clones. AU_G indicates arbitrary units of GFP.

Moreover, it is important to note that the maximum GFP output for test part 3 is the same as for test part 4, thus demonstrating the full induction of this system by exogenously adding $3OC_6HSL$ or by endogenously producing $3OC_6HSL$ via LuxI (compare Fig. 4.5 and Fig. 4.6).

Test parts 5-10 were used to characterize the capability of RBS-luxI to produce $3OC_6HSL$ when induced at different expression strengths. This information is essential when dealing with mux CH0 and demux input because RBS-luxI is directly interfaced with an input promoter in both devices. These test parts are constitutive generators of LuxI, driven by promoters of different strength. The constitutive promoters BBa_J23101, BBa_J23105, BBa_J23106, BBa_J23110, BBa_J23116 and BBa_J23118 were ranked by RFP measurement (Fig. 4.7A). These promoters, taken from the Registry, were contained in BBa_J61002 high copy vector, which carries RBS-RFP-Terminator downstream of the SpeI restriction site, adjacent to the promoter. For this reason, they were ready to be measured without performing any assembly.

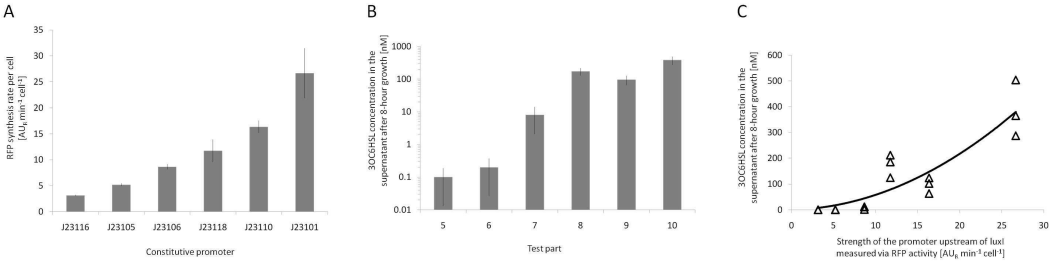


Figure 4.7: Characterization of the LuxI protein generators (test parts 5-10). (A) Constitutive promoters ranked by RFP measurements. Error bars represent the standard deviation of the measurements of three assayed clones. (B) $3OC_6HSL$ production after 8 h growth. Error bars represent the standard deviation of the measurements of three independent experiments. (C) $3OC_6HSL$ concentration as a function of promoter strength (diamonds) with a second-degree polynomial fit (solid line, R^2 >0.8). AU_R indicates arbitrary units of RFP.

$3OC_6HSL$ was measured in the supernatant of cultures with test parts 5-10 (Fig. 4.7B) through the BBa_T9002 biosensor, whose estimated detection limit was 0.1 nM (data not shown). The concentration of $3OC_6HSL$ in the growth medium of these cultures was correlated to the upstream promoter strength, giving high concentrations of autoinducer when stronger promoters produce LuxI and low concentrations in case of weaker promoters (Fig. 4.7C), following a nonlinear trend.

The NOT gate was also characterized when coupled to the previously studied luxR-P_{lux} system, which was a fundamental part of the AND gate, through

test parts 11-12 by RFP measurement. The genetic circuit of test part 11 was expected to express luxR through the P_λ promoter, which was not repressed because cI was not produced. Therefore, in the presence of $3OC_6HSL$, the P_{lux} promoter can be turned on. When a promoter is placed upstream of this circuit, thus yielding test part 12, cI is produced and it can repress P_λ. In this case, luxR cannot be produced and P_{lux} cannot be turned on upon $3OC_6HSL$ exogenous addition. Induction results for test part 12 show that, when the expression of cI is driven by BBa_J23114 promoter, RFP activity is lower than in test part 11 for inducer concentrations up to 100 nM of $3OC_6HSL$ (Fig. 4.8). This difference is probably due to the partial repression of P_λ promoter in test part 12, so luxR expression is lower than in test part 11 and the activation of the P_{lux} promoter is weaker, as expected. However, this difference cannot be appreciated for 1-10 μM inductions. As illustrated below, mathematical model predictions are in accordance with this experimental result and the right working conditions of the NOT gate are discussed later. It is important to note that BBa_J23114 activity is about 33% of BBa_J23116, the weakest promoter of Fig. 4.7A (data not shown). This suggests that the NOT gate is highly sensitive to the transcriptional input and even a low transcription rate can repress the P_λ promoter.

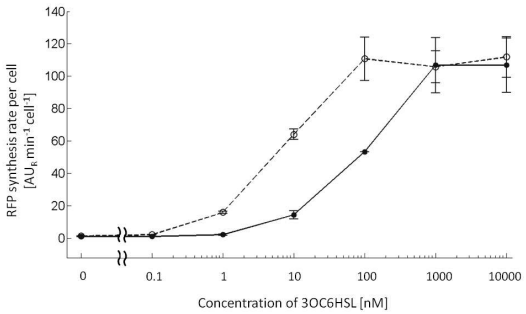


Figure 4.8: Results of the NOT gate. P_{lux} activity as a function of $3OC_6HSL$ for test part 11 (dashed line and open circles) and 12 (continuous line and filled circles). Error bars represent the standard deviation of the measurements of three assayed clones. AU_R indicates arbitrary units of RFP.

4.2.3 Mathematical model simulations

Input (P_A, P_B for mux and P_I for demux) and selector (P_S) promoter values were varied to span all the truth table combinations for both devices. Fluorescence output values, in terms of RFP (for mux) or RFP and GFP (for demux) synthesis rate per cell, were predicted after 16 h, starting from an inoculum of $3.6{\cdot}10^7$ cells. In all the simulations, initial conditions were set to zero for all the species, except LuxR, which was already supposed to be present in cells at a steady-state level ($\frac{\alpha_\lambda}{\gamma_X}$, see Eq.D.7 in Appendix D). Fig. 4.9A shows the simulated results for mux and demux as a function of the input/selector promoters activity in different ON or OFF conditions.

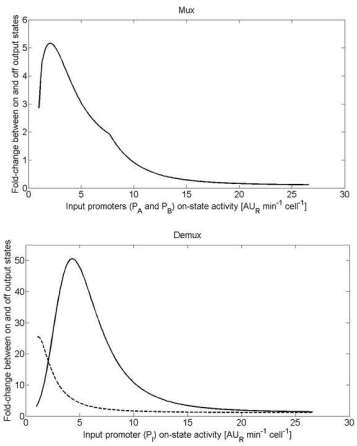
When the ON state was 3 AU_R min^{-1} $cell^{-1}$ (the same strength as that of BBa_J23116) and the OFF state was 0.03 (i.e. 1% of the ON state), mux and demux output values showed the expected behaviour with a 5- and 11-fold change[1] between the ON and OFF states, respectively (Fig. 4.9A, case 1). However, among the ON state output values of mux, RFP could vary up to 5.5 fold. This behaviour was due to the OR gate architecture, in which RFP was produced either by the strong promoter P_{lux} or by the weaker promoter P_{las}. To simulate systems in a less ideal condition, the OFF state activity was increased to 0.3 (i.e. 10% of the ON state) for all the promoters. The difference between ON and OFF states was lower than that of the previous case, in fact, it was 1.7- and 1.9-fold for mux and demux, respectively (Fig. 4.9A, case 2). Nevertheless, both systems were still functional. If input promoters activity in the ON state was increased from 3 to 26.6 AU_R min^{-1} $cell^{-1}$ (the same strength as that of BBa_J23101), with the selector activity as in case 1, both systems lost the expected behaviour (Fig. 4.9A, case 3). In fact, a threshold between ON and OFF states cannot be defined in mux or demux. On the other hand, when the P_S ON activity was increased to 26.6 (with an OFF activity of 1%, keeping input promoters for ON and OFF states at 3 and 0.03, respectively),

[1]Fold change was computed as the ratio between the minimum value of the ON state and the maximum value of the OFF state.

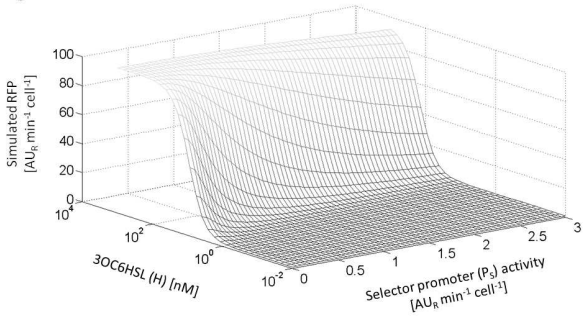
A

Simulated cases:				case 1 (on/off) state = (3/0.03) AU_R min^{-1} $cell^{-1}$ for input/selector promoters		case 2 (on/off) state = (3/0.3) AU_R min^{-1} $cell^{-1}$ for input/selector promoters		case 3 Input promoter(s) (on/off) state = (26.6/0.266) AU_R min^{-1} $cell^{-1}$ P_S promoter (on/off) state = (3/0.03)	
MUX									
input A	input B	selector S	output RFP	Simulated RFP [AU_R min^{-1} $cell^{-1}$]		Simulated RFP [AU_R min^{-1} $cell^{-1}$]		Simulated RFP [AU_R min^{-1} $cell^{-1}$]	
0	0	0	0	1		1.1		1.1	
0	0	1	0	1.1		1.6		1.5	
0	1	0	0	1.6		4.4		7.5	
0	1	1	1	7.6		7.6		8.4	
1	0	0	1	41.3		8.5		99.4	
1	0	1	0	1.2		1.7		71.5	
1	1	0	1	41.8		11.7		105.8	
1	1	1	1	7.7		7.7		78.4	
DEMUX									
input I	selector S	output RFP	output GFP	Simulated RFP [AU_R min^{-1} $cell^{-1}$]	Simulated GFP [AU_G min^{-1} $cell^{-1}$]	Simulated RFP [AU_R min^{-1} $cell^{-1}$]	Simulated GFP [AU_G min^{-1} $cell^{-1}$]	Simulated RFP [AU_R min^{-1} $cell^{-1}$]	Simulated GFP [AU_G min^{-1} $cell^{-1}$]
0	0	0	0	1	0.7	1	1.3	1	0.8
0	1	0	0	1	0.9	1	6.5	1	1.2
1	0	1	0	41.2	6.4	8.3	34.2	99.3	64.5
1	1	0	1	1.1	66.6	1.1	66.6	98.8	73.4

B



C



D

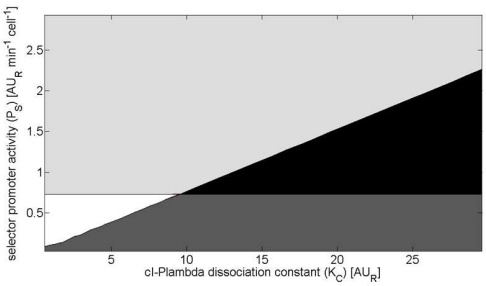


Figure 4.9: (A) Simulated output values of mux and demux under three conditions (the reported values for the two demux output channels correspond to RFP and GFP activities and are expressed in different arbitrary units). (B) Fold-change difference between ON and OFF output states in mux and demux as a function of input promoter (P_A and P_B for mux; P_I for demux) ON state activity. P_S ON state activity was set at 3 AU_R min^{-1} $cell^{-1}$; the OFF state activity was set at 1% of the ON activity for all the promoters. In the demux graph, the continuous line refers to RFP, whereas the dashed line refers to GFP. (C) Model prediction of RFP for the NOT gate coupled to the luxR-P_{lux} system (see test part 11) as a function of $3OC_6HSL$ (H) and the activity of the promoter upstream of cI (P_S). (D) Simulated output of demux as a function of P_S and K_C, referring to case 2, panel A. ON/OFF threshold values for RFP and GFP were set to 4.5 AU_R min^{-1} $cell^{-1}$ and 50 AU_G min^{-1} $cell^{-1}$, respectively. Dark and light gray zones indicate the region in which only RFP or GFP, respectively, is on. Black and white zones indicate the region in which both or none, respectively, of them are on. AU_R and AU_G indicate arbitrary units of RFP and GFP, respectively.

both devices worked correctly (simulation not shown).

The fold change between the ON and OFF output states was also studied as a function of the input activities to identify the working conditions of both devices. The results are reported in Fig. 4.9B.

Moreover, the mathematical model was used to study the behaviour of the subsystem composed by the NOT gate coupled to the luxR-P_{lux} system, described above, to better understand the causes of the unexpected experimental results obtained with test parts 11 and 12 (Fig. 4.8). Fig. 4.9C reports the predicted RFP as a function of $3OC_6HSL$ and P_S (i.e. the promoter upstream of cI). Model predictions are consistent with the experimental data described in the previous section (Fig. 4.8). In fact, RFP is an increasing function of $3OC_6HSL$ and it is repressed by P_S. However, when $3OC_6HSL$ increases, the difference between RFP values in repressed (high P_S value) and unrepressed (low P_S value) states decreases, thus impairing the function of the NOT gate. High $3OC_6HSL$ concentration values are related to the strong activity of P_A and P_I promoters. This can explain why high input promoter values are detrimental to the correct behaviour of both mux and demux, as illustrated in the simulation results above. If the LuxR maximum synthesis rate is decreased, simulations show that RFP can be more strongly repressed at high $3OC_6HSL$ concentrations and the NOT gate exhibits a better performance in the presence of strong P_A and P_I promoters (see D.3.1).

Sensitivity analysis was carried out on the cI-P_λ dissociation constant (K_C), which was critical during simulation trials. Fig. 4.9D reports the predicted GFP and RFP activities in demux as a function of K_C and P_S when input promoter P_I is ON (3 AU_R min^{-1} $cell^{-1}$). Fig. 4.9D shows the K_C and P_S ranges leading to output states in which both channels are ON or OFF (black and white regions). The same unwanted phenomena also affect the selection capability in mux (simulation not shown), in which the two input channels are not conveyed to the output in a mutually exclusive fashion.

Finally, the effect of crosstalk between the lux and las systems was in-

vestigated. The results indicate that orthogonality between the two quorum sensing systems is a critical requirement for mux and demux, since even low-entity crosstalk parameter values can be detrimental (see D.3.2).

4.3 Conclusions

Mux and demux are two fundamental devices in electronics. They are used in several applications, for example, in communication devices, arithmetic logic units, and, in general, in applications that involve channel sharing. Analogously, they could play a crucial role in building complex genetic circuits, in fact, both of them can be used as controlled genetic switches.

Two gene networks that could lead to mux and demux logic functions were designed, built up, and are available in the open source Registry of Standard Biological Parts. Their behaviour was characterized by performing simulations through an *ad-hoc* mathematical model, also defined thanks to *in vivo* experiments expressly performed as a part of this study.

Because the designed mux and demux elements conform to PoPS device boundary standard, they can be considered as general black box (PoPS in - PoPS out) devices, which can be tailored by users to detect any kind of input signal and to express any protein as an output by assembling the desired input and output devices, which must conform to the PoPS functional standard. For these reasons, their application field is very wide.

Mathematical model simulation results showed that both circuits behave as expected, but only under specific conditions. In fact, thanks to the availability of a mathematical model, it was possible to highlight some crucial points that should be considered as guidelines for future work. First, to have mux and demux that work as expected, their input and selector promoters must produce the downstream protein with a synthesis rate within a limited range. For the circuits presented herein, this range barely meets the requirements for real applications, since promoters characterized by too strong or too leaky ac-

tivity make the systems lose the correct behaviour. To overcome this problem, input RBSs can be tuned in future versions of the devices, as described in [41] and [46], to reach the synthesis rate values identified herein as a necessary condition for correct operation. To the same aim, the RBS upstream of luxR can also be tuned to improve the performance of the NOT gate in the presence of strong input promoters. Second, crosstalk between the lux and las systems may impair the whole behaviour of the systems. Even though its effect cannot be decreased, the actual amount of crosstalk should be measured with *ad-hoc* experiments on lux and las biological elements to enable more accurate analysis through the mathematical model and a better definition of the working conditions. Third, even tuning input/selection promoter synthesis rates to achieve the expected behaviour, variation in the values of other model parameters can produce unwanted phenomena. An example is the cI-P_λ dissociation constant, which causes the non-mutually exclusive input or output selection in mux and demux, respectively, for some of the simulated values. Quantitative tests were performed on a set of constructed circuit subsystems containing biological logic gate elements. An AND gate composed of quorum-sensing elements and a NOT gate, coupled with a part of the AND gate, were characterized. The tests performed validated all the input combinations of the AND gate, which gave the expected output GFP levels. Tests also allowed the steady-state characteristic of a part containing a constitutively expressed luxR gene and a reporter gene under the control of P_{lux} promoter to be measured (test part 3). Considering Fig. 4.2A, this configuration is the equivalent of keeping the second input of the AND (lux) gate constant, while the first input changes by providing different concentrations of $3OC_6HSL$, which simulates the presence of the LuxI enzyme. The measured steady-state characteristics were useful to estimate the cutoff point between high and low states of the AND gate. The RBS-luxI module, which is involved in the mux and demux inputs, was characterized by placing promoters of different strengths upstream and measuring the concentration of the LuxI enzyme product, $3OC_6HSL$, under all conditions. This can be useful to determine the critical input promoter

strength required to produce a sufficient amount of $3OC_6HSL$ to activate P_{lux} through LuxR.

Knowledge of the autoinducer production capability of RBS-luxI (test parts 5-10), together with the induction behaviour of luxR-P_{lux} (test part 3), can support the re-usability of the implemented biological parts in different contexts. In fact, RBS-luxI can be used as a sender of chemical signals in the form of $3OC_6HSL$ to engineer quorum-sensing networks for cell-to-cell communication systems. On the other hand, test part 3 can be used as a receiver of such chemical signals that can produce GFP as a function of the received $3OC_6HSL$ molecules. When properly coupled with a heat-sensitive cI repressor, such as cI857, test part 3 can also behave as an AND gate where inputs are $3OC_6HSL$ and heat.

NOT gate results (test parts 11-12) showed that, within a specific range of input values, the constructed logic gate was functional, because P_λ could be repressed by cI and its behaviour was also predicted by mathematical model simulations. In this case, P_λ was turned off when cI expression was driven by a very weak promoter. This confirms the unsuitability, also predicted by the mathematical model, in real applications of the constructed NOT gate because even slight basal expression of the mux or demux selector promoter may partially repress luxR production. As discussed above, optimization is required.

In conclusion, part of the implemented logic functions have been shown to work in several tests and they can be re-used in more complex artificial systems designed to process Boolean signals. Although the parts that could lead to mux and demux devices are available in the Registry, future improvement is required to tune these parts for real applications, taking into account the mathematical model analysis of the systems.

Chapter 5

Production of biofuels from dairy industry wastes with synthetic biology

In this chapter, synthetic biology concepts have been proposed to solve an industrial problem. In particular, the main goal is the production of a biofuel (ethanol) from a waste of dairy industries (cheese whey), thus trying to face two challenges at the same time: energy production from a renewable source and disposal of a waste that is nowadays considered as an environmental problem. After an introduction about cheese whey, bioethanol and ethanol-producing microorganisms (5.1), the synthetic biological system design will be illustrated (5.2) and the results will be presented (5.3) and discussed (5.4).

5.1 Background

5.1.1 Cheese whey

Whey is a component of milk that separates after curdling when rennet or an edible acidic substance is added. It represents one of the main by-products

of the cheese industry. It is a mixture of different substances, some of which potentially valuable if extracted and individually processed. The main solid components of whey are soluble proteins, lipids, lactose, mineral salts and vitamins (see Tab. 5.1) [88, 89, 90, 91]. Dairy industries generally produce wheys of different content depending on the specific productive process and kind of cheese [92].

Components	**Sweet whey [%]**	**Acid whey [%]**	**Scotta [%]**
Total solids	6.4	6.2	5.67
Proteins	0.8	0.75	0.39
Fat	0.5	0.04	0.07
Lactose	4.6	4.2	4.7
Ash	0.5	0.8	0.53
Lactic acid	0.05	0.4	-

Table 5.1: Typical composition of different cheese whey classes. Sweet and acid whey are obtained from curdling by using rennet and acids respectively. Scotta is the denomination of the whey resulting from ricotta production process.

In Italy (as in other countries with an important dairy tradition) the amount of whey produced every year is huge: on average, at least 8-10 Mtons of whey are produced to obtain $\sim$1 Mton of cheese, while the world whey production is over 160 Mtons per year [92, 93, 94].

Whey is commonly regarded as the 'environmental problem' of the cheese factories because if the liquid is introduced directly into the river systems, it contributes to the organic pollution of the environment and gives rise to water asphyxia. Its high nutritional load, in fact, causes the proliferation of microorganisms which deplete oxygen levels in the surface water systems [88]. For this reason, the Italian law classifies whey as a 'special waste'. Two indicators are commonly used to measure the nutrient load: the Biochemical Oxygen Demand within 5 days ($B.O.D.5$) and the Chemical Oxygen Demand ($C.O.D.$). $B.O.D.5$ indicates the amount of dissolved oxygen needed by biological organisms to break down organic material in a given liquid sample in 5 days, while $C.O.D.$ indicates the oxygen amount needed for the complete oxidation of organic and inorganic compounds in a given liquid sample.

Disposal of whey is an important problem even if for a long time this product has been discharged into water streams, used for land spreading or to feed farm animals, since the $B.O.D.5/C.O.D.$ reduction treatments require a great economic effort [92, 95]. Thanks to the current technologies, it is possible to obtain high-value substances, such as lactose and whey-proteins, to employ in pharmaceutical and food industry, by using membranes and filtration [88]. However, lactose extraction from whey is not an economically convenient process because the cost of the extraction is higher than the economic value of lactose itself. As a consequence, even if large amounts of proteins are extracted, the residual liquid of cheese whey (called *permeate*) remains a special waste because of the high lactose concentration ($\sim$45 g/l) [88]. According to the industry needs, both whey and permeate can be concentrated (up to $\sim$200 g/l of lactose) to facilitate storage and transportation.

For these reasons, new solutions to valorize whey (or its derivatives) must be found. In this work, lactose fermentation is used to produce ethanol fuel, with the main goal of decreasing $B.O.D.5$ and $C.O.D.$ while generating a useful product.

5.1.2 Bioethanol

Biofuels can be defined as fuels derived from recent biological materials, and are thus distinguished from fossil fuels, which are derived from ancient biological materials. Because of the current concerns about oil reserves limited availability and green house gas (GHG) emissions connected to the commonly used fuels, the importance of biofuels is increasing since they can supply alternative, renewable and sustainable energy sources. Examples of biofuels are ethanol, diesel, methane, methanol, propanol, butanol, hydrogen and ethers.

This chapter mainly focuses on ethanol biofuel (or *bioethanol*) production.

Ethanol is an alcohol that is a colorless liquid at ambient temperature, extremely flammable and volatile. Even if the attention will be focused on its

importance as a biofuel, it is also used in beverages (e.g. spirits, wine and beer), cosmetics and disinfectants. A number of microorganisms are able to produce ethanol from different carbon sources by fermentation (see below). The produced ethanol can be extracted from a liquid mixture by distillation and finally it can be concentrated up to ∼99.7% by dehydration to yield fuel-grade bioethanol.

The currently used substrates for bioethanol production are sugar (e.g. sugarcane and sugar beet), starchy (e.g. cereal grains and potatoes) and cellulosic (e.g. paper and non-edible plant biomass) feedstock. Sugar feedstock could be directly fermented into ethanol by microorganisms such as yeasts, while starchy and cellulosic feedstock must always be pre-treated to release fermentable sugars from complex materials.

Brazil is the country where bioethanol is mostly used as a fuel. It is produced from sugarcane feedstock, which is used as a substrate by the ethanol-producing yeast *S. cerevisiae* in industrial plants. Even if sugarcane can be easily cultivated in countries like Brazil, this feedstock does not represent an economically-convenient solution in other countries. The development of cost-effective technologies for fuel ethanol production is a priority that implies the selection of the most appropriate feedstock and the selection and definition of a suitable process configuration. Whey can be considered as a *free* feedstock for bioethanol production from lactose.

Ethanol fermentation (or alcoholic fermentation) is a biochemical process in which sugar (or a fermentable carbon source) is transformed into carbon dioxide and ethanol. This process generally occurs anaerobically and it is exploited by a number of microorganisms (such as yeasts) to gain cellular energy in absence of oxygen.

Before fermentation takes place, one mole of glucose is transformed into two moles of pyruvate in the metabolic pathway called *glycolysis*, which is governed by the following reaction:

$$glucose + 2NAD^+ + 2ADP + 2P_i \rightarrow 2pyruvate + 2NADH + 2H^+ + 2ATP + 2H_2O$$

NADH can be oxidized back to NAD^+ in the subsequent alcoholic fermentation process, thus allowing the glycolysis to continue. In alcoholic fermentation, one mole of pyruvate is decarboxylated to carbon dioxide and acetaldehyde (which acts as electron acceptor) and then acetaldehyde is reduced to ethanol by electrons coming from NADH, in the process regenerating NAD^+. The essential enzymes involved in the described steps are the pyruvate decarboxylase and the alcohol dehydrogenase, which catalyze the $pyruvate \rightarrow acetaldehyde + CO_2$ and the $acetaldehyde \rightarrow ethanol$ reactions respectively.

It is important to note that other kinds of fermentations can also support the regeneration of NAD^+ by generating a variety of by-products, such as lactate, acetate, butyrate, butanol, isopropanol, etc. The fermentation class is a characteristic of individual species of organisms, for example lactic fermentation is exploited by yogurt-making bacteria to yield lactic acid as the main fermentation product.

If the starting fermentable carbon source is not glucose but another complex sugar or compound, other specific reactions must occur to transform it before glycolysis takes place.

Engineered biological systems are promising solutions for biofuel production and synthetic biology can be used to produce such systems. Steps towards this goal have been recently reported in literature [22, 96, 97].

5.1.3 State of the art

The long-term goal of this project is the definition of an industrial-scale fermentation plant for the conversion of the lactose contained in cheese whey into ethanol. In particular, the work described here focuses on the design and construction of an engineered strain able to perform this conversion.

Among the naturally occurring microorganisms, the yeast *S. cerevisiae* is the most used catalyst in ethanol production plants because it is able to produce and tolerate ethanol concentrations higher than 150 g/l [98]. However, *S. cerevisiae* is not able to metabolize lactose, so it could only be used after a whey pre-treatment in which lactose is converted into glucose and galactose by the β-galactosidase enzyme [89]. Such additional step requires to purchase a purified enzyme with a non-negligible economic impact [99, 100]. The same problem also occurs for the fermentative soil bacterium *Zymomonas mobilis* (another industrially used ethanol producer), which is even unable to metabolize galactose [101]. In the bacteria kingdom, a number of microorganisms are able to consume lactose. However, none of them is able to yield ethanol as the main fermentation product [102]. Organic acids are usually obtained by such bacteria from sugar fermentation. *E. coli*, for example, produces a mixture of organic acids and ethanol, but the latter is produced at a low yield [103]. The yeasts *Kluyveromyces lactis* and *Kluyveromyces marxianus* are the only documented organisms able to directly convert lactose into industrially relevant ethanol amounts [98]. However, common drawbacks have been reported for them [98, 104]: incomplete sugar utilization and slow fermentation were present when too high lactose concentrations were added, thus preventing the production of high-concentration ethanol. Moreover, additional supplements (e.g. complex nutrients or lipids) have usually been necessary to obtain reasonable yields. Oxygen levels are also critical for the working of these strains and slight variations from the optimal levels may impair the overall performance of fermentation. In addition, the extent of such effects seems to be strain-dependent.

Despite some examples of industrial implementation in Ireland, New Zealand and the USA that use the above mentioned lactose-utilizing strains, the fermentation technology must be further improved in order to enhance the attractiveness of whey-to-ethanol bioprocesses, to reduce the reported drawbacks and to support cheese whey valorization worldwide [98]. Specifically, there is the need to develop microbial strains that ferment lactose to ethanol with very

high efficiency.

Genetically modified *S. cerevisiae* strains carrying β-galactosidase and lactose permease from *Kluyveromyces lactis* or secretable β-galactosidase from *Aspergillus niger* were constructed to mix the excellent lactose-utilization and ethanol production performances of two organisms [98]. On the other hand, genetically modified *E. coli* strains carrying the pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase II (*adhB*) genes from *Z. mobilis* were constructed to implement the efficient ethanol production capability in natural lactose-utilizing organisms [97]. In the latter case, an ethanologenic operon composed by *pdc-adhB* was incorporated into the genome of different *E. coli* strains. Screening and selection cycles brought to the isolation of recombinant clones able to ferment sugars into ethanol at high conversion yields [105].

These strains showed the desired functions, but their testing revealed industrially relevant problems: instability of the recombinant genes [104, 106], sub-optimal tuning of gene dosage [107] and large nutritional requirements to efficiently complete fermentations [108].

The design principles introduced in synthetic biology can play an important role in the optimization of such systems.

5.1.4 Specifications for a lactose-to-bioethanol system

The main system specifications for a lactose-to-ethanol producing strain, suitable for industry, are here summarized:

- the organism must effectively grow in cheese whey/permeate and ferment lactose with a minimum amount of supplements;
- no antibiotic must be used to maintain the foreign genes, because i) the cost associated to the process increases, ii) the fermentation broth could not be easily disposed of in the environment and iii) resistance genes should be avoided in large-scale plants to prevent the potential spreading of such resistances [18];

- the final strain should be able to ferment high concentrations of lactose into ethanol with a conversion yield close to the theoretical one (i.e. 0.54 grams of ethanol per gram of lactose), considering the upper limit of ethanol tolerance of the strain. Direct fermentation of whey or whey permeate to ethanol is generally not economically feasible because the ∼45 g/l of lactose result in low ethanol titre (<25 g/l), making the distillation process too expensive [98]. For these reasons, whey should be concentrated (e.g. by ultrafiltration and/or reverse osmosis) to reach a proper lactose concentration before starting fermentation;

- the introduced genetic devices should work with a minimum energy demand, in order to avoid starvation-like conditions in the cells in poor growth media and to limit the genetic instability of the device due to such conditions.

5.2 System design

5.2.1 Bottom-up engineering of bioethanol production

E. coli bacterium was used as a chassis to host the genetic program of bioethanol production because of the wide genetic parts toolbox available to engineer it. Here, according to the above mentioned specifications, an ethanol production pathway has been refactored, assembled and its elements tuned with the aim of optimizing the fermentation performance of the system. The pathway has been implemented using the two *Z. mobilis* genes described above: pdc and adhB, encoding pyruvate decarboxylase and alcohol dehydrogenase II respectively.

DNA synthesis technology and the Registry of Standard Biological Parts were used to optimize the characteristics of pdc and adhB: the codon usage has been optimized for *E. coli*, through DNA *de novo* synthesis, in order to introduce heterologous genes composed by the favourite codons of the host organism. A strong RBS was chosen from the Registry and ligated upstream

of the two genes. Codon usage optimization and strong RBS together can ensure a high translation rate per mRNA molecule. The ethanol yield and the metabolic burden can be calibrated using different input promoters for the two genes. Tab. 5.2 reports the essential BioBrick parts used to engineer the pathway.

The final version of the device should be controlled by constitutive promoter(s) in order to ensure an efficient and continuous ethanol production without any additional cost for inducer molecules. However, before assembling the final system, the strength of the promoters has to be calibrated and the assembly of the genes with a large number of BioBrick promoters can be avoided by using well characterized inducible promoters. These inducible systems can be used as user-controlled knobs for gene expression. In this study, the lux and tet regulated systems were used. Both systems have been shown to provide a homogeneous expression of the genes downstream of P_{lux} and P_{tetR} [54, 61, 109], thus avoiding *all-or-none* phenotype, i.e. the rising of two subpopulations (one uninduced and one fully induced) upon induction [110].

BioBrick	**Description**	**Source**
BBa_B0030	Strong RBS	Registry
BBa_K173016	*Z. mobilis* pdc gene codon-optimized for *E. coli*	Mr Gene DNA synthesis service
BBa_K173017	*Z. mobilis* adhB gene codon-optimized for *E. coli*	Mr Gene DNA synthesis service
BBa_B0015	Double transcriptional terminator	Registry
-	Constitutive promoters to be chosen	-
BBa_F2622 and BBa_R0040	HSL- and aTc-inducible promoters, used as genetic knobs to optimize gene dosage	Registry

Table 5.2: BioBrick parts for ethanol production in *E. coli*.

Fig. 5.1 shows how the engineered ethanol production pathway will look like: pyruvate metabolism is re-directed from organic acids production (e.g. lactate or acetate) to acetaldehyde/CO_2 production by pyruvate decarboxylase and acetaldehyde can be eventually transformed into ethanol by adhB. While no pdc analogous is present in wild type strains, *E. coli* already possesses native

alcohol dehydrogenases. However, they are present at low concentrations and are less efficient than adhB [111].

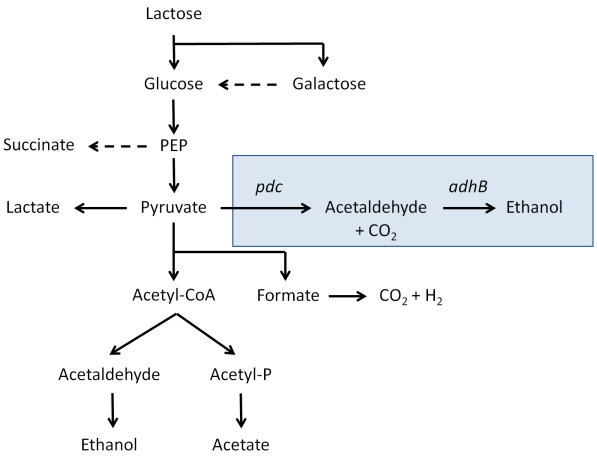


Figure 5.1: Schematic diagram of the wild type and engineered fermentation pathway of *E. coli* [103]. Solid and dashed arrows indicate a direct and indirect transformation respectively. PEP=phosphoenolpyruvate. The engineered pathway is highlighted and recombinant enzymes are indicated in italics. Wild type fermentation yields a mixture of lactate, acetate, succinate, formate and ethanol, while the engineered pathway diverts pyruvate metabolism towards ethanol as the main product with a theoretical yield of 0.54 grams of ethanol per gram of lactose.

Even if plasmids are an easy-to-handle solution for system optimization, the genetically encoded functions should be stably present in single copy in the final host genome, thus avoiding the propagation of plasmids, which usually require antibiotic selective pressure to be maintained.

The described strategy should provide a system with optimally regulated genes, in which the pathway enzymes are produced with a minimum transcriptional requirement in order to enable efficient ethanol production when the genetic program is present in single copy and to avoid the use of large amounts of supplementary nutrients to support gene expression at high levels. In fact, the extremely high nutritional requirement for ethanol genes transcription was one of the drawbacks of the constructed *E. coli* systems reported in literature [108]. If the transcriptional requirement is too high in single copy condition, multiple copies can be integrated in the genome. In the next section, a standard solution for the single copy incorporation of biological parts is

illustrated in order to facilitate the integration of BioBrick parts and devices in the chromosome of *E. coli*.

5.2.2 A standard integrative vector for *E. coli*

Several tools for *E. coli* have been proposed which exploit homologous recombination or site-specific recombination. In particular, homologous recombination can be used to insert the desired DNA fragment (called *passenger*) into a specific genomic locus that must show sufficient sequence homology with a second DNA fragment (called *guide*) used to target the locus [112]. Apart from genomic integrations, this technology can be used for general genome engineering, as gene knockout and mutation of native genes can be also carried out [113]. Integrative plasmids [114], linear PCR fragments [115] and also single-stranded DNA [116] can be used to perform the described tasks by means of unspecific endogenous or heterologous (e.g. the λRed system [115, 116]) protein machinery. 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 [117]. The process is mediated by a specific phage recombinase. This mechanism has been exploited for the development of integrative vectors that carry the attP site (guide) and the passenger of interest. A number of attP sites from native phages have been characterized and used, as well as the specific recombinases that target them into the specific attB site.

The gene expression machinery that mediates homologous or site-specific recombination can be carried on an easily curable *helper* plasmid transformed in the host strain [117].

In general, integrant clones are selected with an antibiotic resistance marker. It could 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 were also constructed [118].

When integrative plasmids are used, clones with a successful integration should be easily selected and these plasmids should also be easy to amplify *in vivo*. For this reason, 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 [113]. For example, the R6K replication origin can be used to propagate plasmid only when the pir or pir-116 gene is present in the host strain [117].

Temperature sensitive replication origins are also commonly used to cure helper plasmids.

Considering integrative plasmids, recently basic elements for integration have been refactored and adapted to the RFC21 BioBrick standard [18]. Site-specific recombination vectors based on the ϕ80 and P21 phages attP sites were constructed and validated for genome integration, as well as antibiotic marker excision. This effort was a demonstration that standard biological parts could highly benefit the construction of integrative systems. Anderson JC and co-workers provided several useful BioBrick parts for the construction of integrative systems, but no final standard and ready-to-use BioBrick compatible solution is available yet. This lack of solutions is analogous to the lack of a standard solution to engineer BioBrick-compatible vectors by using BioBrick parts, which lasted until the development of BBa_I51020, a base vector that can be easily specialized by using standard modules (e.g. to change replication origin or resistance marker) [47]. In this work, we face such problem in the integrative vector design and construction field, by providing a novel integrative base vector for *E. coli* which is compatible with BioBrick standard and can also be engineered with BioBrick parts to target the part of interest into the desired genome locus.

The structure of the designed vector, here named BBa_K300000, is shown in Fig. 5.2 and the parts notation is reported in Fig. 5.3. Most of its features have been inspired by the previous works reported in [18] and [47]. It can be considered as a base vector, which can be specialized to target the desired

integration site in the host genome. The default version of this backbone has the bacteriophage ϕ80 attP as guide. This vector enables multiple integrations in different positions of the same genome.

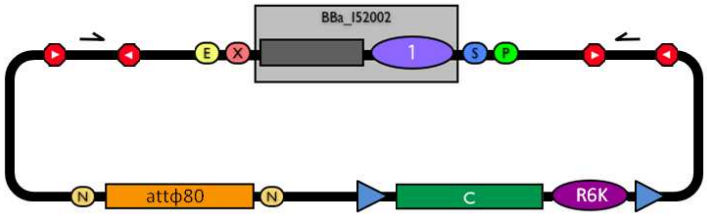


Figure 5.2: BioBrick integrative base vector BBa_K300000.

Part number	Function	Notation
BBa_G00000	BioBrick cloning site prefix	
BBa_G00001	BioBrick cloning site suffix	
BBa_P1016	*ccdB* positive selection marker	
BBa_I50022	minimal pUC-derived high copy replication origin	
BBa_J72001	FRT recombination site	
BBa_B0053 & BBa_B0054	forward transcriptional terminator	
BBa_B0055 & BBa_B0062	reverse transcriptional terminator	
BBa_G00100	forward sequencing primer annealing site (VF2)	
BBa_G00102	reverse sequencing primer annealing site (VR)	
BBa_B0045	NheI restriction site	
BBa_K300999	Chloramphenicol resistance marker (reverse orientation)	
BBa_K300998	R6K conditional replication origin	

Figure 5.3: Parts notation for BBa_K300000.

Vector engineering features:

1. The cloning site is compatible with the original BioBrick standard (RFC10), i.e. it is composed by the BioBrick prefix and suffix. The presence of illegal restriction sites XbaI in the FRT sites and SpeI in the default guide

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 BioBrick parts in the cloning site upon EcoRI-PstI digestion.

2. The two NheI restriction sites flanking the default integration guide sequence 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.

3. The default insert contains a *ccdB* toxin gene expression cassette and a pUC19-derived replication origin. The toxin is lethal for most *E. coli* strains and it is useful to prevent the growth of transformants containing the uncut plasmid contaminant DNA. For this reason, the default vector must be propagated in ccdB-tolerant strains, such as DB3.1. The pUC19-derived origin enables the propagation of this vector at high copy in the used ccdB-tolerant strain. Another version of this vector (not shown) was designed with an RFP-expression cassette driven by P_λ as the default insert. It gives a visible red-coloured phenotype to transformants containing the uncut plasmid contaminant DNA. Both versions were successfully used to obtain the results described in 5.3.

4. Like in many other standard vector backbones (e.g. the pSB**5 series in the Registry), 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.

Genome integration features:

1. The four transcriptional terminators flanking the cloning site ensure the transcriptional insulation of the integrated part from the adjacent genome sequences.

2. 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 integrations in the same strain, always using the same antibiotic resistance marker.

3. 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.

How to use it:

BBa_K300000 can be:

- propagated in *E. coli*;
- engineered to change the passenger and/or the integration guide;
- integrated into the desired locus of the host genome;
- used to perform the desired number of serial integrations in the same genome.

How to propagate it before performing genome integration:

This vector contains the default toxin-encoding insert, so it must be propagated in a ccdB-tolerant strain such as DB3.1. After the insertion of the desired BioBrick part in the cloning site, this vector does not contain a standard replication origin anymore, so it must be propagated in a pir^+ or $pir\text{-}116^+$ strain such as BW25141 or BW23474 that can replicate the R6K conditional origin. If using the version with the RFP cassette as the default insert, the plasmid must always be propagated in a pir^+ or $pir\text{-}116^+$ strain.

How to engineer it:

The DNA guide can be changed as shown in Fig. 5.4.

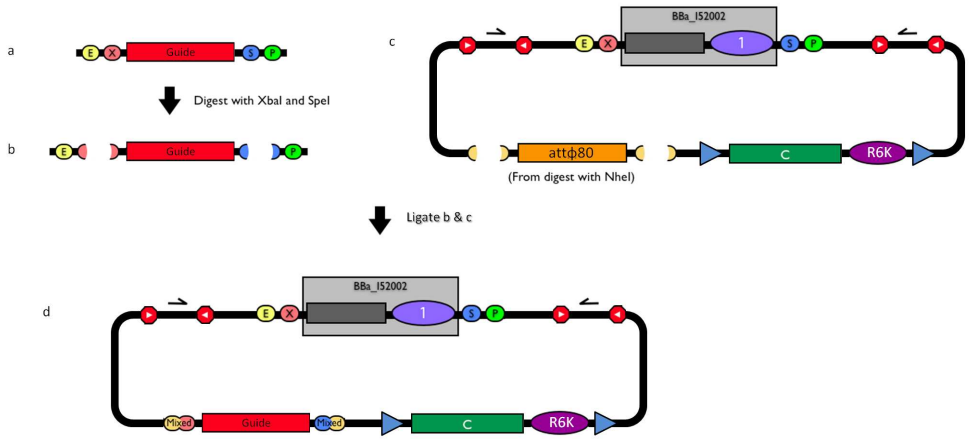


Figure 5.4: How to engineer the integrative base vector to assemble the desired DNA guide.

1. Be sure to have the desired guide in the RFC10 standard or a compatible one (Fig. 5.4a).
2. Digest the guide with XbaI-SpeI (Fig. 5.4b).
3. Digest the integrative base vector with NheI (Fig. 5.4c) and dephosphorylate the linearized vector to prevent re-ligation.
4. Ligate the digestion products (Fig. 5.4d). XbaI, SpeI and NheI all have compatible protruding ends. Note that the ligation is not directional, but the guide can work in both directions.
5. Transform the ligation in a ccdB-tolerant strain (if the insert is BBa_I52002) or a pir$^+$/pir-116$^+$ strain (if the insert is the RFP cassette) and screen the clone.

The DNA passenger can be changed as shown in Fig. 5.5.

1. Be sure to have the desired passenger in the RFC10 standard or a compatible one (Fig. 5.5a).
2. Digest the passenger with EcoRI-PstI (Fig. 5.5b).
3. Digest the integrative base vector with EcoRI-PstI (Fig. 5.5c).

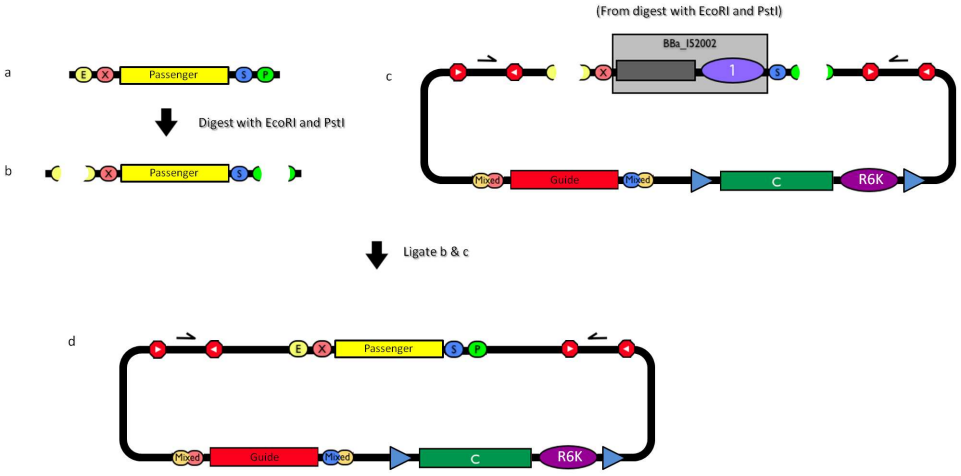


Figure 5.5: How to engineer the integrative base vector to assemble the desired DNA passenger.

4. Ligate the digestion products (Fig. 5.5d).

5. Transform the ligation in a pir$^+$/pir-116$^+$ strain. Transformants with the uncut plasmid contaminant DNA do not grow if the original insert contained the toxin-encoding cassette, otherwise they grow but they can be easily recognized as red colonies. Screen the clone.

How to perform genome integration:

Because the vector can be specialized with the desired DNA guide, the integration into the *E. coli* chromosome can exploit both bacteriophage attP-mediated integration or single-crossover homologous recombination. Detailed protocols about attP-mediated integration can be found in [18, 117], while detailed protocols about homologous recombination can be found in [112, 113].

How to perform multiple integrations in the same genome:

When this vector is integrated in the genome, the desired passenger should be maintained in the host, as well as the Chloramphenicol resistance marker and the R6K conditional-replication origin. The CmR and the R6K can be excised from the genome by exploiting the two FRT recombination sites that flank them. The Flp recombinase protein mediates this recombination event, so

it has to be expressed by a helper plasmid, such as pCP20 ([118]). This enables the sequential integration of several parts using the same antibiotic resistance marker, which must be eliminated every time before each integration step. Detailed protocols about such recombination can be found in [115, 118].

5.3 Results

5.3.1 Genomic integration

The BBa_K300000 integrative vector was successfully designed, constructed and used to integrate a number of proof-of-concept devices in the chromosome of two *E. coli* strains (MG1655 and MC1061).

To test the vector, RFP- or GFP-expression devices were integrated in the ϕ80 locus via the default guide, using BBa_J72008 as a heat-curable helper plasmid expressing the ϕ80 integrase (the protocol is reported in E.1).

A typical experiment (RFP- and GFP-constitutive cassettes in MG1655 and MC1061) gave integrations in the correct position (see Fig. 5.6) and loss of the helper plasmid (not shown) with 100% efficiency. However, 82% of the correct integrants showed multiple tandem copies of the passenger (see Fig. 5.7). When Chloramphenicol resistance and R6K origin were excised upon Flp recombinase action, 100% of the clones showed excision and the multiple copies always became a single copy (see Fig. 5.8). This happened because Flp mediated the recombination of *all* the FRT sites of the multiple tandem copies, until only a single FRT site was present in the ϕ80 locus, thus leaving only a single copy of the part, without antibiotic markers.

Up to now, the developed integrative vector has been exploited to integrate a total number of 15 BioBricks and a recently published work on nonlinearity analysis of gene expression as a function of DNA copy number extensively used it to construct the clones [54]. Among these experiments, 100% of the cases gave integrants in correct position with a successful loss of the helper

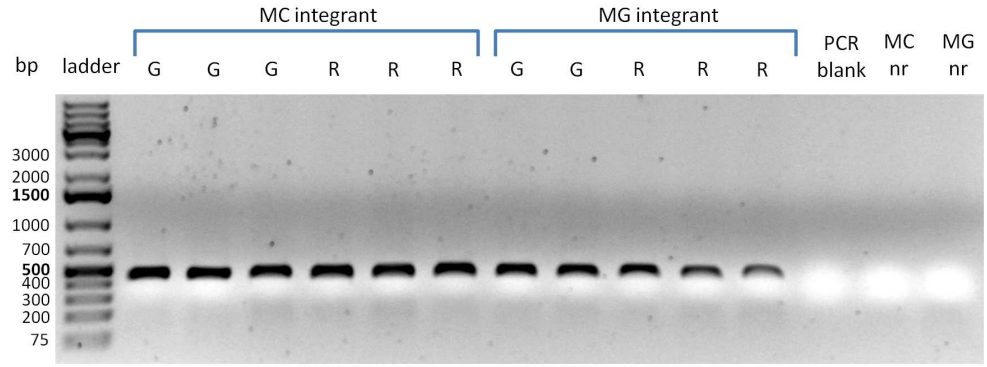


Figure 5.6: Colony PCR on integrant strains with primers P1-P2, annealing in opposite directions in the genome and in the integrative plasmid respectively (see Appendix E). If the integration position is correct, a 452-bp amplicon is produced. MC = MC1061 strain; MG = MG1655 strain; G = GFP cassette as passenger; R = RFP cassette as passenger. Control reactions on non-recombinant (nr) strains are also reported.

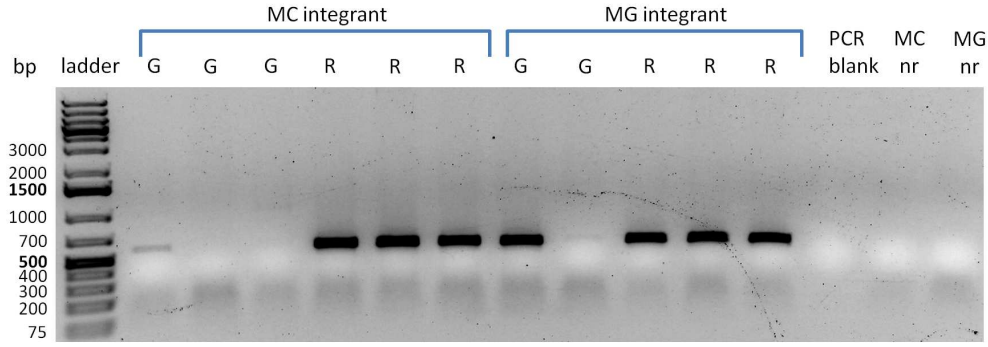


Figure 5.7: Colony PCR on integrant strains with primers P2-P3, both annealing in the integrative plasmid. Reaction produces a 666-bp amplicon if two or more tandem integrants are present in the genome (see Appendix E). MC = MC1061 strain; MG = MG1655 strain; G = GFP cassette as passenger; R = RFP cassette as passenger. Control reactions on non-recombinant (nr) strains are also reported.

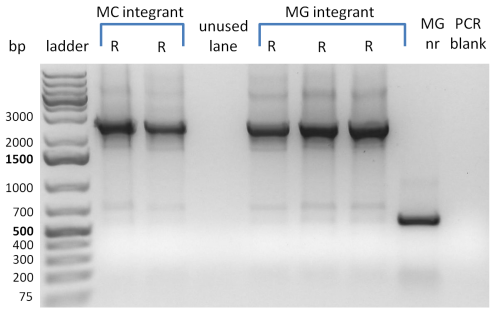


Figure 5.8: Colony PCR on integrant strains (after the FRT/Flp-mediated marker excision) with primers P1-P4, both annealing in the genome in opposite directions. Reaction produces a 2.3-Kbp amplicon if a single integrant of the passenger, without Chloramphenicol resistance or R6K origin, is present in the ϕ80 genomic locus (see Appendix E). MC = MC1061 strain; MG = MG1655 strain; R = RFP cassette as passenger. A control reaction on a non-recombinant (nr) strain is also reported.

plasmid. 73% were also successful in marker excision at the first attempt, while the remaining 27% yielded a correct FRT recombination only after two or more colony screening rounds. Finally, the majority of the recombinant strains obtained were analyzed by sequencing the integrated DNA fragment, which was correct in 100% of cases. When possible, phenotype was also assessed for the constructed clones. Only in two experiments the final clone could not work as expected, even if sequencing was correct. This could be explained by integration events in additional unwanted genomic positions, resulting in either multiple copies of the desired device or the disruption of native genes that were necessary to the device functioning (e.g. the lacI repressor). Such hypotheses have not been validated yet.

As the physical modularity of the vector allows to change the DNA guide, the steps described in Fig. 5.4 were also successfully performed to construct a vector that could target the aspA locus in the MG1655 genome through homologous recombination to integrate an RFP-constitutive expression cassette (data not shown).

5.3.2 Ethanol production

Ethanol production via regulated operon-structured genes (TOP10 strain, 100 g/l of glucose)

In order to carry out a preliminary characterization on the pdc and adhB genes, they were assembled in a promoterless synthetic operon (see Fig. 5.9) and incorporated into TOP10 *E. coli*. Being recA-deficient, this strain has a reduced occurrence of non-specific recombination in cloned DNA and it has already been used in literature for ethanol production [119, 120]. However, it is unable to utilize lactose. These preliminary experiments were carried out in LB medium (which does not contain significant amounts of fermentable sugars [121]) supplemented with a known amount of glucose as fermentable sugar, which is the easiest substrate to utilize for ethanol production in engineered *E. coli* [122].

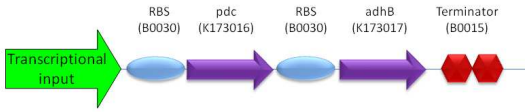


Figure 5.9: The ethanol-producing operon tested in this work. It is composed by pdc and adhB, both of them with a strong RBS upstream. BioBrick codes are reported in brackets.

Tab. 5.3 reports the promoters used to drive the expression of the operon and the plasmid copy number conditions tested.

BioBrick promoter or device	**Qualitative strength**	**Copy number***
P_{tetR} (BBa_R0040)	medium-strong	HC
BBa_J23118	medium	HC
BBa_J23106	weak	HC
HSL-inducible device (BBa_F2620), composed by the P_{lux} promoter and a luxR expression cassette driven by P_{tetR}	very strong (upon HSL induction)	HC and LC

*HC means a high copy plasmid (pSB1A2 or pSB1AK3), while LC means a low copy plasmid (pSB4C5).

Table 5.3: Promoters and plasmids used to express the pdc-adhB operon.

A qualitative analysis of the recombinant strains phenotype showed that:

- strains with the P_{tetR}, J23118 or J23106 promoters upstream of the operon gave very small colonies when plated on selective LB agar plates;

- strains with the HSL inducible system upstream of the operon in high copy plasmid could grow when not induced, but died when induced with >1 nM of HSL in both LB and LB + 20 g/l of glucose;

- strains with the HSL-inducible system upstream of the operon in low copy plasmid could grow when induced with 1 μM of HSL;

- bacterial density of strains with the HSL inducible system upstream of the operon in liquid cultures supplemented with 20 or 100 g/l of glucose was higher than the density of all the other tested strains grown in the same conditions.

The growth results suggest that a high metabolic burden affects the strains bearing an even weakly expressed ethanol-producing operon in high copy number: the P_{tetR}, J23106 or J23118 promoters gave very poor growth when used to constitutively express the operon and restriction analysis/sequencing highlighted large DNA mutations. The strain bearing an HSL-inducible pdc-adhB operon in high copy did not survive upon induction (1 nM of HSL), while it could survive in the same conditions when assembled in a low copy plasmid. In high copy number conditions, only the HSL-inducible pdc-adhB operon without HSL addition could survive. The observed high metabolic burden was not caused by fermentation products, as induced strains bearing pdc-adhB in high copy could not grow in both LB and LB supplemented with a fermentable sugar.

Considering the mentioned results, in the subsequent experiments the HSL-inducible system was used to drive the expression of the operon. In particular, no HSL was added in the high copy plasmid condition (thus trying to exploit the P_{lux} promoter leakage activity), while 1 μM of HSL was added in the low copy condition.

Quantitative experiments were performed to test the growth and fermentation of these recombinant strains.

Growth of strains bearing:

- HSL-inducible system upstream of the operon in high copy plasmid, uninduced;
- HSL-inducible system upstream of the operon in low copy plasmid, uninduced;
- HSL-inducible system upstream of the operon in low copy plasmid, induced with 1 μM of HSL;
- HSL-inducible system without the operon (BioBrick BBa_F2620) in high copy plasmid;
- a promoterless operon in high copy plasmid

was monitored in selective LB + 100 g/l of glucose in a microplate reader (see Fig. 5.10).

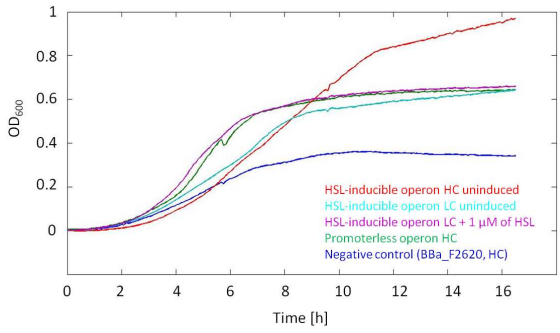


Figure 5.10: Growth curves of strains bearing the HSL-inducible ethanol-producing operon, grown in selective LB + 100 g/l of glucose in microplate reader. OD_{600} was used to monitor cell growth and it refers to the pathlength of a microplate well filled with 200 μl of liquid culture.

Ethanol fermentation capability of these strains in LB medium + 100 g/l of glucose was tested (see Appendix E for the detailed protocols). The main

fermentation results are illustrated in Fig. 5.11. Ethanol and pH were measured after 24 or 48 hours.

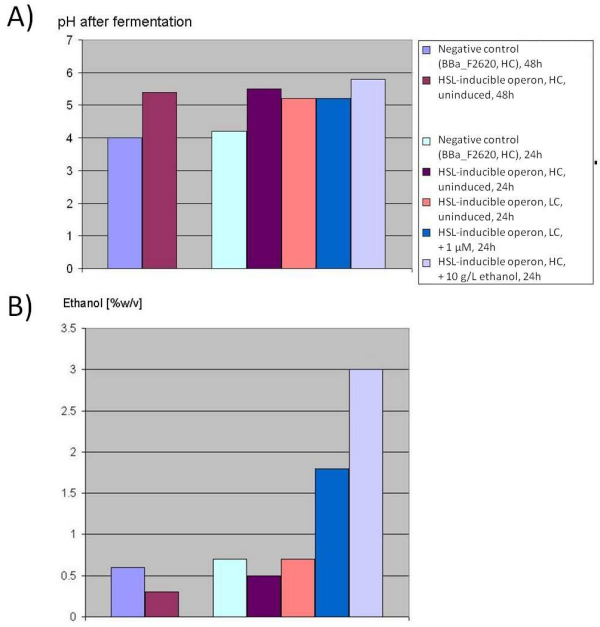


Figure 5.11: Final pH (A) and ethanol production (B) for strains bearing the HSL-inducible pdc-adhB operon, grown in selective LB medium + 100 g/l of glucose. Ethanol production is expressed in % weight/volume. When ethanol is added to the medium, the bar represents the *net* ethanol production due to fermentation.

The pH results show that strains expressing the pdc-adhB operon produce lower levels of organic acids than the negative control, which does not have pdc or adhB. It is surprising that also the pH of the promoterless pdc-adhB operon in high copy is higher than in the negative control. This could be due to a weak spurious transcription of pdc and adhB, amplified by the high copy number plasmid, which may re-direct part of pyruvate metabolism to ethanol and not to organic acids.

The results about dynamic growth (in terms of OD_{600}) are in accordance with the pH results: strains bearing the operon reach higher cell densities than the negative control, probably because the acidity of the medium is lower and it limits cell growth less than it happens in the negative control. No significant difference was noticed in the growth of promoterless operon in high

copy, uninduced and induced operon in low copy, further suggesting that in promoterless operon transcription could be activated in unspecific manner (i.e. without a promoter upstream).

Fermentation results showed that strains bearing expressed pdc-adhB operon in high copy plasmid produce less ethanol than the negative controls. On the other hand, if 10 g/l of ethanol are added to the medium this strain can produce 30 g/l of ethanol, which is about 60% of theoretical yield (in the hypothesis that all the glucose was consumed). This effect may be due to adhB behaviour, which was reported in literature to be enhanced by ethanol accumulation [123]. Moreover, it is possible that when starting from less than 10 g/l of ethanol in fermentation medium, ethanol production could be increased, because in this experiment the bacterial culture was exposed to 40 g/l of ethanol, which represents the survival threshold of TOP10 (data not shown) and so it is difficult for the strain to produce more than 30 g/l.

When induced with 1 μM, the strain bearing the inducible operon in low copy plasmid also yielded a higher ethanol concentration (18 g/l) than the negative control. This represents about 36% of theoretical yield.

In the found promising conditions for pdc-adhB operon, however, the theoretical yield was not reached and in the high copy context the addition of 10 g/l of ethanol was required and the constructed operon was toxic for TOP10 when expressed at even weak levels in a high copy plasmid.

In order to investigate which of the two genes was responsible of the toxicity, pdc and adhB were tested separately under the control of the HSL-inducible promoter in a high copy vector. Cultures were grown in LB or LB + 20 g/l of glucose for 16 hours and OD_{600} was used to monitor cell growth. The transcription of genes was tuned by different concentrations of HSL in the medium and results showed that pdc expression was highly toxic: even weakly induced cultures failed to grow and only the uninduced culture was turbid after overnight incubation (data not shown). On the other hand, adhB expression did not sig-

nificantly alter cell growth at any HSL concentration (data not shown). These results also confirmed that toxicity was caused by pdc itself and not by the produced acetaldehyde, as even pdc-expressing cultures in LB without any fermentable sugar could not grow.

pH was measured at the end of the experiment. As expected, uninduced cultures bearing pdc showed a much higher pH (5.8) than the negative control (4.5), confirming that the leakage activity of P_{lux} in high copy plasmid can express sufficiently high amounts of pdc to prevent the organic acid accumulation shown in the negative controls. Cultures with adhB, on the other hand, showed a pH comparable with the negative control, thus confirming that only pdc can re-direct pyruvate metabolism and not adhB.

Ethanol production with independently regulated genes (MG1655 Z1 strain, 40 g/l of lactose)

Although gene organization in a synthetic operon could be useful to preliminary characterize the system, as described in 5.2 it is desirable to optimize the expression of the individual genes because pdc and adhB may need different expression levels for the correct functioning (given the strong RBS upstream). For this reason, they were tested under the control of two independent promoters.

The final genetic system should be able to work in single genomic copy to ferment lactose into ethanol. In order to test the genes under similar conditions, independently regulated pdc and adhB were cloned in a low copy plasmid ($\sim$5 DNA copies per cell) and the resulting plasmid-based system was incorporated into the MG1655 Z1 *E. coli* strain, which naturally metabolizes lactose. HSL- and aTc-inducible systems (BioBricks BBa_F2622 and BBa_R0040) were used to regulate the expression of pdc and adhB respectively. BBa_R0040 is the P_{tetR} promoter, while BBa_F2622 contains a luxR expression cassette with P_{lux}. The used strain constitutively overexpresses the tetR gene, so the P_{tetR} promoter could be effectively tuned without adding a tetR gene in the plasmid.

First, the two inducible systems were characterized in terms of Relative Promoter Units (RPUs) (see Fig. 5.12). As shown in this figure, they can be tuned over a wide range of transcriptional activities, yielding maximum activities of 1 (tet) and 6.7 (lux) RPUs. Induction curves are extremely important in this work because when optimal promoters activity will be found for pdc and adhB expression, constitutive promoters with an equivalent activity will have to be used to regulate the pathway without exogenous inducer molecules.

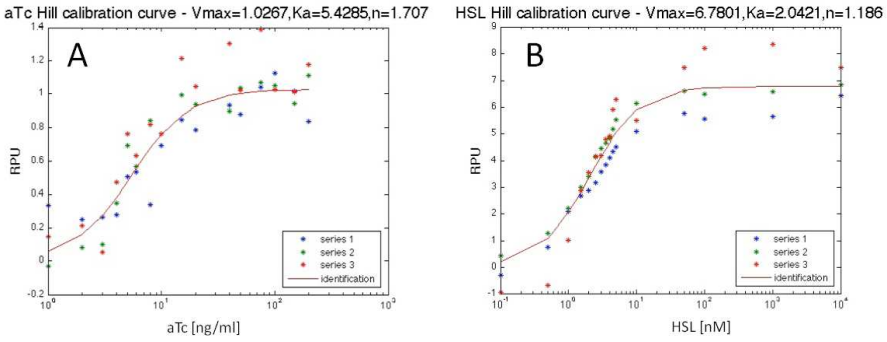


Figure 5.12: Induction curves of the two *genetic knobs* used to regulate pdc and adhB expression. The aTc-inducible system is the BBa_R0040 P_{tetR} promoter, while the HSL-inducible system is the BBa_F2622 device. The series represent RPU data from three independent clones assayed in LB via RFP measurements.

In order to test the double-regulated system (here called MGZ_{dbl}), P_{tetR} (driving adhB) was set to RPUs≈1 by adding 100 ng/ml of aTc, while P_{lux} (driving pdc) was set to RPUs≈0.5 and 1 by adding 0.25 nM and 0.5 nM of HSL respectively. Such promoter activities were chosen to express pdc and adhB at similar levels. Fermentation results after 24 hours are shown in Tab. 5.4.

While the negative control (strain lacking pdc/adhB) ferments only ∼0.5 g/l of lactose without producing any significant amount of ethanol, recombinant strains with induced pdc/adhB produce similar amounts of ethanol (∼6 g/l), consuming ∼30% of the lactose and approaching 100% of the theoretical yield. No lactic or acetic acid was detected for the two ethanol-producing strains. Surprisingly, even the negative control did not show detectable lactic or acetic acid. Its pH drop to 4.7 could be explained by the production of other acids (succinic and/or formic) during fermentation.

Strain	**pdc induction [nM of HSL]**	**Ethanol [g/l]**	**Residual lactose [g/l]**	**Lactic acid [g/l]**	**Acetic acid [g/l]**	**Final pH**	**Fermentation time [h]**
control	-	0	39.6	0	0	4.7	24
MGZ_{dbl}	0.5 (RPUs=1)	6.37	28.5	0	0	7	24
MGZ_{dbl}	0.25 (RPUs=0.5)	5.82	28.3	0	0	6.5	24
MGZ_{dbl}	0.25 (RPUs=0.5)	7.5	29.75	1.5	0.91	5.5	48
MGZ_{dbl}	0.25 (RPUs=0.5)	8	27.67	1.69	0.94	5.2	72
MGZ_{dbl}(*)	0.25 (RPUs=0.5)	11.6	5.44	0.52	0.98	6	48

Fermentation results for the ethanol-producing strain bearing the double-regulated system (MGZ_{dbl}), induced as reported. Fermentations were carried out in LB medium + 40 g/l of lactose (initial pH=6.8), except the one marked with (*), in which 200 mM of sodium phosphate buffer (pH 7.0) were supplemented (initial pH=7.0). P_{tetR} (driving adhB) was always set to RPUs≈1 by adding 100 ng/ml of aTc.

Table 5.4: Lactose-to-ethanol fermentation results.

Because the above illustrated fermentations gave similar results, the induction condition in which pdc expression was the lowest (RPUs≈0.5) was chosen for further studies in which fermentation was carried out for 48 and 72 hours (see Tab. 5.4). Results show that ethanol production is only slightly improved when compared to 24-hour experiments. Ethanol yields approached 100% again. Data also show that pH decreased over time and lactic and acetic acids were detected. The measured values are comparable to the typical organic acid concentrations of other ethanologenic strains reported in literature [105] and they are much lower than the theoretical values for lactic and acetic acids produced in wild type *E. coli* [103].

Even if ethanol yields were extremely high, only 30% of the available lactose was utilized by the strain in the conditions used above. In order to improve lactose utilization, the fermentation medium was supplemented with 200 mM of sodium phosphate buffer at pH=7.0. In fact, several published works report that acidification inhibits ethanol production in recombinant *E. coli*, which have an optimum fermentation pH between 6 and 7 [106, 112]. Fermentation results are reported in Tab. 5.4 for the ethanol-producing strain in the same induction conditions as above. Data after a 48-hour fermentation show that both ethanol final concentration and lactose consumption were improved

when compared to the previously studied conditions. In particular, 11.6 g/l of ethanol were produced and more than 86% of the available lactose was consumed, resulting in an ethanol yield of 62%.

5.4 Conclusions

The industrial problem of cheese whey valorization was faced by using the synthetic biology approach. In particular, metabolic engineering of lactose-to-ethanol fermentation was implemented in *E. coli*, a lactose-utilizing organism, to carry out the following main functions: i) reduce the pollutant load of whey by consuming its nutrient content and ii) convert lactose into bioethanol. Two genes were previously found to be necessary for the re-direction of pyruvate metabolism from mixed organic acids production to ethanol production as the main fermentation output. The main limitations of the up-to-date developed systems include genetic instability of the recombinant genes and insufficient expression levels when strain grows in poor media.

Here, the implementation of the pathway was conducted through synthetic biology concepts by optimizing the single genetic elements involved in the process: translation was maximized by optimizing the genes according to the host organism codon usage and then by assembling a strong RBS, coming from a library of pre-characterized parts, upstream of the two genes; finally, well-characterized regulatory modules were used to drive and tune the expression of the genes.

Such approach has provided promising bacterial systems that can convert lactose to ethanol with a conversion yield between 60% and 100% of the theoretical one. However, residual lactose was a major problem during fermentations, as only in a buffered environment (pH 7.0) consumption was reasonably high (∼86% of the initial lactose), but conversion yield was 62%. In the other experiments, only up to 30% of lactose was consumed.

Even if promising recombinant strains have been developed, a real opti-

mization process is still ongoing, to find the optimal inductions for pdc and adhB gene expression. When the optimal expression levels will be found for pdc and adhB, well-characterized constitutive promoters should be assembled to ensure the optimal expression strength without the addition of exogenous inducer molecules. This highlights the importance of devices characterization and modularity studies (as reported in Chapter 2), because the bottom-up optimization of the ethanol production pathway can only be done in a modular context, thus avoiding trial-and-error approaches. It is important to note, however, that the use of inducible systems in long experiments (24- to 72-hour fermentations) can hide noteworthy problems, in fact inducer molecules can spontaneously degrade. In the specific inducible systems used in this work, experiments (performed in different growth media) showed a ∼13-hour and ∼20-hour half life for HSL and aTc respectively (data not shown). Ethanol production could be improved by re-constituting over time the desired concentration of inducers, thus preventing the turn-off of the genetic knobs used.

Testing the systems in different bacterial strains is also a very important point, as here the MG1655 Z1 strain was used for lactose fermentation, but other strains could be more efficient to perform ethanol production. Crucial points of this project are the growth in whey/permeate with a minimum amount of expensive supplements and ethanol tolerance. The screening of different bacterial strains should also consider these key aspects to optimize the whole process. Preliminary experiments suggested that the used strains (TOP10 and MG1655 Z1) can effectively grow in filter-sterilized (0.1 μm) whey permeate, even if cell density is much lower than in LB medium (data not shown). Additional nutrients should be added to improve the cell yield.

Finally, a useful genetic tool for the integration of BioBrick parts into the desired genomic locus was developed, thus providing standard techniques to construct industrially-attractive stable strains without plasmids or antibiotic resistances.

In conclusion, the discussed project has contributed essential starting points

and promising results for the optimization of a very important metabolic pathway and its industrial utilization. Several BioBricks have been constructed, thus easily enabling a future re-use of the implemented genetic parts, e.g. by assembling the genes to different candidate promoters or by transferring the promising parts in different standard plasmids. Taken together, these aspects highlight the potential advantages of synthetic biology (e.g. physical standards and bottom-up engineering of modular regulated systems) to solve some of the current drawbacks that affect the metabolic engineering of ethanol production from lactose when traditional approaches are used.

Chapter 6

Engineering organisms for cellulose degradation with a library of standard biological parts

This chapter describes the activities performed towards the construction of a cellulose-degrading microorganism with the help of synthetic biology concepts. In particular, cellulolytic enzymes have been characterized in different organisms that can be potential hosts for an engineered cellulose degradation system for the large-scale production of biofuels or other valuable compounds. After an introduction about the cellulose degradation challenge (6.1), results about cellulase expression and secretion will be shown as a function of the used microbes, as well as the growth requirement characterization of some of the used organisms (6.2). Finally, conclusions and future perspectives will be presented (6.3).

6.1 Background

6.1.1 Ligno-cellulosic biomass as a source of sugars

Ligno-cellulosic biomass, i.e. the non-edible part of plants, is a highly interesting substrate for the production of fuels or other valuable compounds from renewable sources. Such interest is due to the large abundance of this biomass, which can be found in farms, paper industry wastes or otherwise in rapidly growing plants, cultivated with the specific aim of obtaining it. It accounts for 35-50% of plant dry weight and, in the majority of cases, it is composed by cellulose fibers embedded in a hemicellulose and lignin matrix [124]. Cellulose and hemicellulose can be hydrolyzed to release a variety of sugars that can be fermented by a number of microorganisms. In particular, hemicellulose ultimately releases a mixture of arabinose, galactose, mannose and xylose [125], while cellulose can mainly release cellobiose [23]. This chapter will focus only on cellulose.

Plant biomass has been defined as the only foreseeable sustainable source of fuels and materials available to humanity [124]. Cellulosic materials are particularly attractive in this context because of their relatively low cost and plentiful supply. The main technological bottleneck to more widespread utilization of this important resource is the unavailability of low-cost technologies for its treatment. Although naturally occurring microorganisms exist that perform the hydrolysis of cellulose with specific enzymes, none of them is industrially attractive because of their growth rate and none of them can naturally produce valuable compounds such as biofuels [125].

Nowadays, the production of fuels from ligno-cellulosic biomass is mainly performed through mechanical and enzymatic treatments that make the biomass release sugars, which can then be fermented by specific microorganisms to obtain the desired products. However, an organism able to carry out both hydrolysis and fermentation would make the entire process industrially more cost-effective.

Three classes of enzymes are necessary for cellulose hydrolysis. Although exceptions exist, the hydrolysis process mainly follows these steps:

- *endoglucanases* attack cellulose fiber cutting random bonds and releasing oligosaccharides of various lengths;
- *exoglucanases* (or *cellobiohydrolases*) move along the oligosaccharide chains and release cellobiose monomers;
- *β-glucosidases* hydrolyze one molecule of cellobiose into two molecules of glucose.

6.1.2 State of the art and specifications for cellulolytic organisms

Although the classes of enzymes mentioned in the previous section have already been identified, cellulose degradation by engineered microorganisms *in vivo* is still a major challenge. The cellulolytic enzymes must be exported outside the cells because cellulose cannot be imported. Apart from secretion, cellulases also have to be expressed at considerably high levels to achieve an efficient cellulose hydrolysis [23]. This represents a main limitation, which has prevented the realization of industrially attractive organisms. Pretreatment with ionic liquids or removal of hydrolase inhibitors are promising processes that can lower the enzyme need for an efficient cellulose hydrolysis, but currently such technologies are not economically convenient [126]. Moreover, the existence of a wide number of cellulases encoded in cellulolytic bacteria genomes suggests that a synergistic activity of several enzymes of the same class is necessary to effectively degrade cellulose [124]. Different combinations of such enzymes may also be differently specific to attack cellulose from various sources or compositions [23].

Synthetic biology offers the ability to combine large numbers of biological components and for this reason it can be a useful tool to discover the synergy

among cellulases with the final goal of creating a novel customized cellulolytic organism. A necessary step in this ambitious project is the characterization of the recombinant cellulases *in vivo*, with particular attention to the secretion capability as a function of the host strain.

The sugars resulting from hydrolyzed cellulose can then be converted into the compounds of interest, such as biofuels, by the same organism. Ethanol has been widely studied as a possible final product from ligno-cellulosic biomass. Several recombinant hosts, such as *S. cerevisiae* [127], *E. coli* [128] and *Klebsiella oxytoca* [129], have been engineered to express specific heterologous cellulases and convert different cellulosic substrates into ethanol. The ideal microorganism for cellulose to ethanol conversion should be able to produce at least 40 g/l of ethanol with a $>$ 1g/l/h productivity and the conversion yield should be at least 90% of the theoretical one [103]. The production of other biofuels, such as biodiesel, butanol and pinene (precursor of a potential jet fuel) has also been reported to be successfully carried out by recombinant *E. coli* grown on plant biomass [126].

The choice of the host strain plays a crucial role in the project, as it might be important not only for the secretion, but also for the growth requirements: in fact an industrially-attractive living system for cellulose degradation and large-scale fermentation should be able to accomplish its functions in a growth medium containing a limited amount of complex nutrients, which may account for an important expense in a future industrial process. In addition, the engineering of microbes which already have the cellobiose utilization capability has also been considered in this project in order to avoid the incorporation of recombinant β-glucosidase enzymes.

6.2 Results

Here, two widely studied cellulases have been used in the whole project for the study of expression and secretion in different hosts. The two genes, both from

the cellulolytic gram-positive bacterium *Cellulomonas fimi*, are *cenA* (encoding an endoglucanase) and *cex* (encoding an exoglucanase). A signal peptide sequence is present at the N-terminal of both the proteins that have been reported to pass the inner membrane to reach the periplasm in gram-negative bacteria. However, some proteins with a periplasmic signal peptide can also pass through the outer membrane to be secreted in the growth medium, probably for unspecific leakage through the membrane, but the process has not been completely understood [130, 131]. For this reason, testing the cellulases in different chassis is needed to find the best organism for biomass degradation. To this aim, specific protocols have been developed and tested in different potentially useful host strains.

6.2.1 Cellulase expression and secretion

The cenA and cex genes were tested under the control of the *E. coli* wild type lac promoter (P_{lac}) in a high copy plasmid in three different microorganisms, here called *A*, *B* and *C*, where *A* and *B* are widely used *E. coli* strains (so they cannot utilize cellobiose), while *C* is a cellobiose-utilizing gram-negative bacterium (see Appendix F for strain denomination and detailed enzyme characterization protocols). Another cellobiose-utilizing organism, called *D*, has also been considered for some of the tests. *C* and *D* are non-standard chassis for engineered recombinant systems. The quantitative methods developed for this study and the knowledge about the behaviour of the non-standard expression chassis used may be helpful for the future characterization of recombinant organisms bearing different combinations of cellulases.

To test the cellulase expression capability and the secretion yielded, two versions of the expression plasmids were constructed: one with the cellulase alone and one with the cellulase in operon with xylE gene. xylE encodes a reporter protein which can be easily detected through a colorimetric assay (using Catechol as chromogenic substrate). Because it is known that XylE is a cytoplasmic protein [132], it was used as a control for secretion assays. In fact,

cellulase extracellular activity can be due either to cellulase export outside the cell or to partial cell lysis and the extracellular detection of XylE can be useful to separate these contributions.

XylE was successfully expressed in all the organisms, giving higher enzyme levels for A and B when compared to C (Fig. 6.1A). This could be due to several factors, for example the P_{lac} and/or the RBS upstream of xylE could be weaker in C when compared to A and B or the plasmid copy number could be responsible for the lower per-cell expression of xylE. Fig. 6.1B shows the secretion of XylE in all the tested chassis and results confirmed that the protein is cytoplasmic for the very low percentage of secretion (<1.3%).

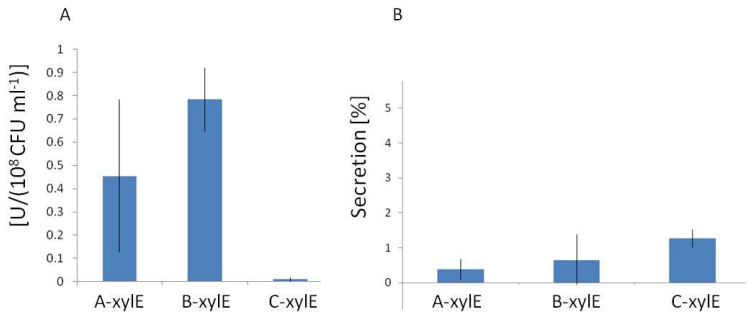


Figure 6.1: Expression (panel A) and secretion (panel B) of xylE in recombinant strains. Error bars represent the standard deviation computed on at least two independent experiments.

Fig. 6.2 shows the intracellular expression of cenA and cex in exponential and stationary growth phase of the microorganisms. Even if the B chassis yields a higher cellulase expression in almost all cases, it is important to note that growth rate and final cell density are, on average, lower than it happens for A and C (see Tab. 6.1).

Fig. 6.3 shows the extracellular enzyme level after overnight growth. Chassis A yields the highest extracellular level for both enzymes.

From intracellular and extracellular activity data it was possible to calculate the secretion capability of all the chassis in exponential and stationary phases. Fig. 6.4 summarizes secretion results, also considering the cytoplasmic control when available.

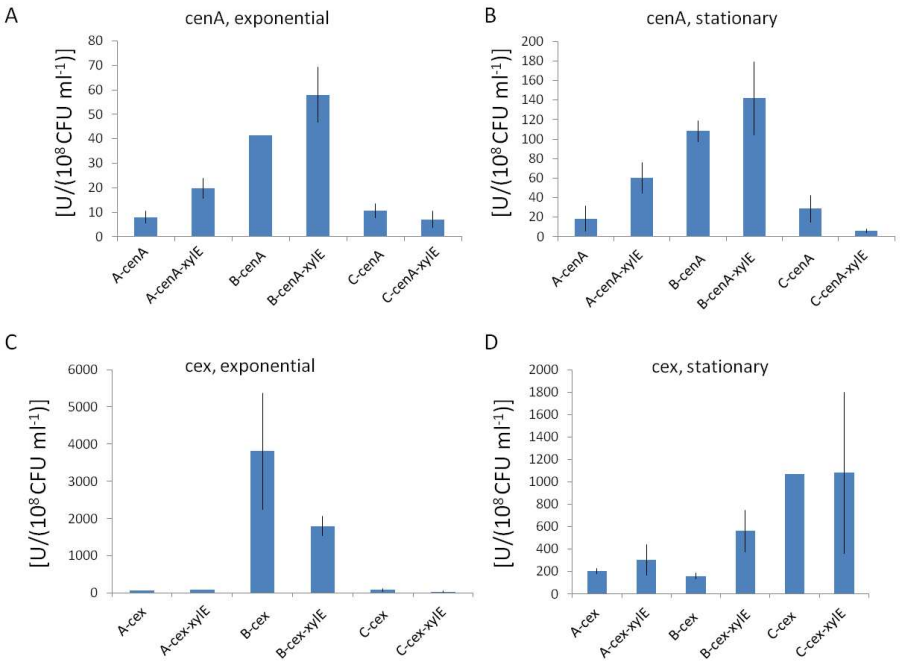


Figure 6.2: cenA (panels A and B) and cex (panels C and D) intracellular activity in exponential and stationary phase. Error bars represent the standard deviation computed on at least two independent experiments.

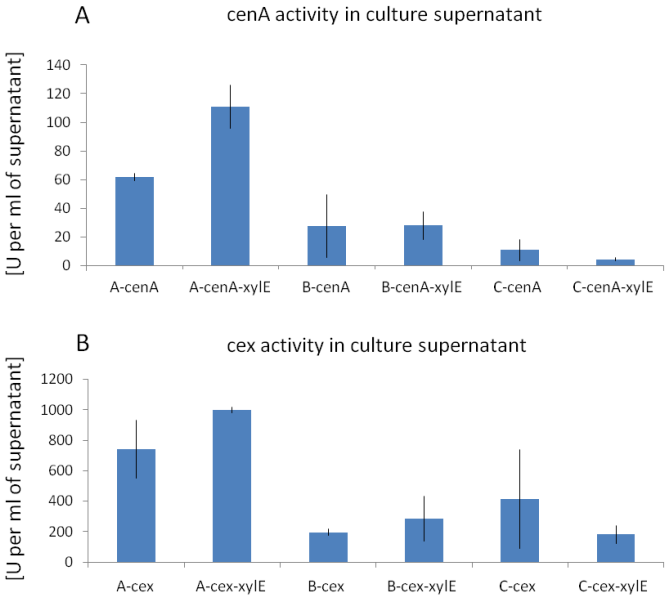


Figure 6.3: cenA (panel A) and cex (panel B) activity in overnight supernatants. Error bars represent the standard deviation computed on at least two independent experiments.

Recombinant strain	Mean doubling time [min]	Mean CFUs reached per ml of culture
A-cenA	62	$3 \cdot 10^9$
A-cenA-xylE	123	$8 \cdot 10^8$
B-cenA	81	$3 \cdot 10^8$
B-cenA-xylE	110	$2 \cdot 10^8$
C-cenA	58	$1 \cdot 10^9$
C-cenA-xylE	55	$1 \cdot 10^9$
A-cex	88	$5 \cdot 10^9$
A-cex-xylE	111	$9 \cdot 10^8$
B-cex	115	$2 \cdot 10^9$
B-cex-xylE	193	$1 \cdot 10^8$
C-cex	66	$4 \cdot 10^8$
C-cex-xylE	45	$7 \cdot 10^8$

Table 6.1: Growth rate and final cell density of cellulase-expressing recombinant strains. CFUs represent the number of recombinant cells per ml of culture.

Results show that no significant cellulase secretion occurred in the exponential phase: even if some strains exported up to 7% of cenA, secretion data were poorly reproducible, while the only expression systems that exhibited a cex secretion above 5% also had a significant amount of cytoplasmic control outside the cells (>3%, much higher than controls without cellulases, see Fig. 6.1B), probably due to partial cell lysis. This hypothesis is supported by the very low growth rate of the cex-bearing strains exhibiting such phenomenon, which could have a high metabolic burden. On the other hand, in stationary phase all the chassis showed a consistent amount of secreted cellulases, up to 20% for cenA and up to 32% for cex, with a low amount of xylE in the supernatant. Surprisingly, in both exponential and stationary phases the addition of xylE seems to affect the expression and the secretion of cellulases in some of the tested systems.

The cellulase secretion results are consistent with some previous works in which the 'leakage' of the enzymes from periplasm to growth medium happened in stationary phase but not in exponential phase [130]. Taken together, these results suggest that chassis *A* is a better secretion machine than the other two tested organisms for expression, secretion and cell growth capabilities.

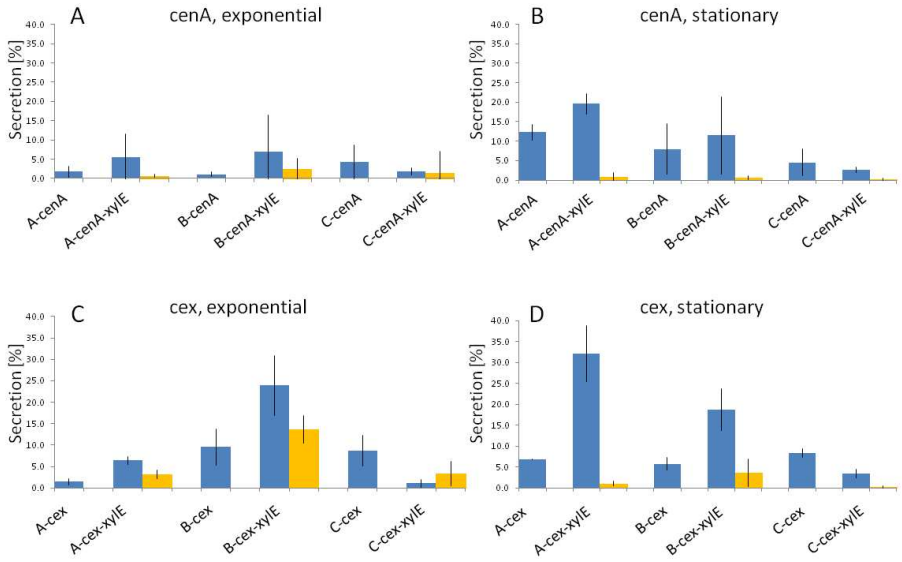


Figure 6.4: Secretion of cenA (panels A and B) and cex (panels C and D). Blue bars: cellulase; orange bars: xylE cytoplasmic control. Error bars represent the standard deviation computed on at least two independent experiments.

During the experiments, an unspecific exoglucanase activity was found for cenA (see Fig. 6.5). According to cenA and cex expression, such unspecific activity is about 200-fold lower than the activity of cex. However, this should be taken into account when performing plate assays to screen exoglucanases with the MUC fluorogenic substrate (see Appendix F).

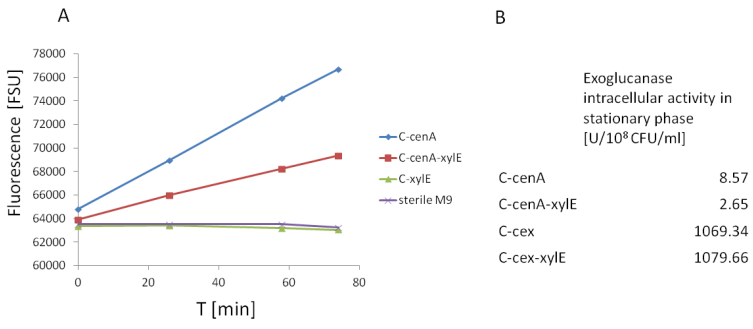


Figure 6.5: Exoglucanase activity of cenA. Fluorescence measurements over time with the MUC assay for C-cenA, C-cenA-xylE and the negative controls C-xylE and M9 supplemented medium (A). Exoglucanase activity of C-cenA and C-cenA-xylE in comparison with C-cex and C-cex-xylE.

6.2.2 Growth requirements of cellobiose-utilizing chassis

As they are non-standard host organisms for the construction of recombinant systems, *C* and *D* chassis were tested for plasmid maintenance in presence of antibiotic. Tab. 6.2 reports the plasmid maintenance data for *C* and *D* bearing some of the expression constructs used in this study.

Recombinant strain	Recombinants in the population [%]	Growth phase
C-cenA	99 (3%)	Exponential
C-cenA-xylE	58 (46%)	Exponential
C-cex	64 (78%)	Exponential
C-cex-xylE	98 (12%)	Exponential
C-cenA	73 (46%)	Stationary
C-cenA-xylE	59 (44%)	Stationary
C-cex	18 (>100%)	Stationary
C-cex-xylE	61 (61%)	Stationary
D-cenA	13 (32%)	Exponential
D-cex	46 (54%)	Exponential
D-cenA	5 (19%)	Stationary
D-cex	21 (3%)	Stationary

Table 6.2: Plasmid maintenance in the cellobiose-utilizing strains used in this study, in presence of antibiotic. All the plasmids had the pUC19 replication origin. Coefficients of variation are reported in brackets.

Data show that plasmid loss progressively occurred in both chassis in the exponential and stationary phases. On average, *C* strain maintained plasmids better than *D*. Plasmid loss seemed to be construct-dependent, with the expression plasmid of cex and of cenA giving the most dramatic losses in *C* and *D* strains respectively. In particular, the extensive loss of cenA in *D* impaired the execution of reliable secretion measurements, as only a very small part of the cell population expressed cenA and the overall amount of this protein was not enough to be detected by the AZO-CMC assay (data not shown). On the other hand, the MUC assay for cex activity was much more sensitive and this protein could be effectively measured even when extensive plasmid loss occurred.

Because only plasmids with pUC19 replication origin and Ampicillin resis-

tance were assayed, different origins and antibiotic resistances should be tested for these studied strains to find the best conditions for construct incorporation.

No significant plasmid loss could be detected for the two *E. coli* strains *A* and *B* (data not shown).

All the presented experiments were performed in the M9 medium supplemented with glycerol and complex nutrients (casamino acids). Apart from cellulase secretion, the investigation of the cheapest growth medium composition that supports an efficient cell growth is necessary.

For this reason, the nutritional requirements of the two promising cellobiose-utilizing strains *C* and *D* were tested, mainly focusing on growth in minimal media with cellobiose as the only carbon source. Results are shown in Tab. 6.3

Data suggest that:

- *C* and *D* strains have different nutritional requirements, as *C* can clearly grow in minimal media with cellobiose as the only carbon source, while *D* seems to be unable (after 36 hours in liquid culture and after 6 days on plate);

- thiamine and trace elements B appear to be helpful to *C* growth;

- both *C* and *D* can grow in minimal media with glucose as the only carbon source;

- both *C* and *D* show β-glucosidase activity (MUG plates), but *D* produces much smaller amounts of acids than *C* during cellobiose fermentation (MacConkey plates). This should be considered when trying to engineer ethanol production in such strains, as low amounts of acids may also mean low ethanol yields from cellobiose. In this case, *C* could be a better host than *D*.

Chassis	**Liquid (L) or plate (P) medium**	**Medium**	**Results**
C	L	M9cellobiose	OD_{600}=0.51 at 24h ; OD_{600}=0.87 at 36h
C	L	M9(thiamine)cellobiose	OD_{600}=0.55 at 24h ; OD_{600}=1.3 at 36h
C	L	M9(thiamine, trace elements B)cellobiose	OD_{600}=0.85 at 24h ; OD_{600}=3.0 at 36h
C	P	M9(thiamine)cellobiose	colonies start to appear visibly after 3 days
C	P	M9glucose	colonies start to appear visibly after 1-2 days
C	P	LB(MUG)	fluorescent colonies
C	P	MacConkey(cellobiose)	pink colonies
C	P	MacConkey(glucose)	strongly pink colonies
D	L	M9cellobiose	OD_{600}=0.01 at 24h ; OD_{600}=0.03 at 36h
D	L	M9(thiamine)cellobiose	OD_{600}=0.16 at 24h ; OD_{600}=0.16 at 36h
D	L	M9(thiamine, trace elements B)cellobiose	OD_{600}=0.07 at 24h ; OD_{600}=0.09 at 36h
D	P	M9(thiamine)cellobiose	colonies are still very small after 6 days, almost undetectable
D	P	M9glucose	colonies start to appear visibly after 1-2 days
D	P	LB+MUG	fluorescent colonies
D	P	MacConkey(cellobiose)	weakly pink colonies (much weaker colour than *C*, slightly higher than the *E. coli* cellobiose negative control)
D	P	MacConkey(glucose)	strongly pink colonies

Table 6.3: Cellobiose-utilizing chassis characterization in minimal and supplemented media.

6.3 Conclusions

The study described in this chapter had the ultimate ambitious goal of developing a recombinant system for cellulose degradation, which can ultimately release easy-to-utilize sugars for the production of valuable compounds, such as biofuels, from renewable sources. As no naturally occurring organisms possess the required capabilities for cellulose hydrolysis and high-value products formation for industrial purposes, the currently working industrial plants for cellulose-to-biofuels conversion rely on biomass hydrolysis via purified enzymes and a subsequent fermentation step in which a microorganism converts the released sugars into the biofuel. The construction of engineered systems is necessary to lower the costs of the whole process, in particular to avoid the

purchase of purified enzymes.

The main challenges in re-engineering cellulose degradation by cellulolytic enzymes are the necessary high-level expression and secretion of the enzymes and their still not well understood synergy. In such framework, physical standardization introduced by synthetic biology enables the assembly of several combinations of cellulolytic enzymes that can then be characterized in the desired expression chassis.

The carried out work was focused on the quantitative characterization of two cellulolytic enzymes (endoglucanase cenA and exoglucanase cex) *in vivo* and provided quantitative data for them when produced by standard and non-standard host organisms. A reporter gene (xylE), which encodes a cytoplasmic protein, was co-expressed with the cellulase as a control in some of the constructs to test if cellulase export in the growth medium was due to secretion or unwanted partial cell lysis. BioBricks were used to construct the desired expression systems.

Results showed successful expression of cellulases in all the chassis. No significant secretion occurred during the exponential growth phase, while it successfully occurred in stationary phase in all the tested microorganisms. Some of the strains expressing cex showed a notably high xylE activity in the growth medium (especially during the exponential phase), suggesting that cex affected the integrity of the host cell. This metabolic burden status was also confirmed by the doubling times of cex-expressing strains that were higher than the ones of cenA-expressing strains.

Importantly, in some cases both expression and secretion were significantly different when strains expressed the cellulase alone or in combination with xylE. This introduces a considerable limit in the prediction of cellulase secretion capability of candidate host strains, as combinations of expressed enzymes could alter the overall extracellular performances. The used high copy number vectors could contribute to such unpredictability, as they have been reported to introduce nonlinearities in the output of biological systems [54, 60].

According to expression, secretion and cell growth, a best-performing chassis was identified (strain *A*). However, other factors, such as growth conditions and medium composition, could alter the tested capabilities. For example, secretion capabilities were significantly different when a subset of the expression strains were tested at 30°C instead of 37°C (data not shown).

The non-standard cellobiose-utilizing used strains (*C* and *D*) were also characterized in terms of plasmid loss, growth requirements and organic acids production during cellobiose fermentation. These data can be considered for the future selection of the expression chassis. Most importantly, these strains should also be tested when grown on cellulosic substrates and expressing cellulolytic enzymes. All these data can be useful to define the cheapest medium composition for a future biomass conversion process.

In this work, the cellulolytic enzymes possessed a natural N-terminal signal sequence that transports the enzymes in the periplasm for a subsequent secretion. The engineering of such secretion signal represents a different approach that has been followed by other research groups to optimize the cellulase export [96, 126].

In conclusion, quantitative methods have been developed and cellulase expression and secretion have been characterized. A lot of work still has to be done to engineer an effective cellulose-degrading strain. Here, steps towards this goal have been performed by providing quantitative data on individual cellulase activity in different microbial hosts, including non-standard promising cellobiose-utilizing strains. BioBricks can support the rapid construction of a large number of these expression systems to test a wider collection of enzymes and their synergy.

Chapter 7

Overall conclusions

The bottom-up programming of living organisms to implement novel user-defined biological capabilities is one of the main goals of synthetic biology. Engineering plays a crucial role in the rational design and construction of these systems, as principles like standardization, modularity and predictability of biological parts are considered key aspects. Currently, predominant problems connected with the construction of even simple synthetic biological systems are the unavailability of well-characterized parts and the unpredictability of the genetic circuitry behaviour when assembled and incorporated in living cells. Steps towards the solution of such problems have been contributed in this thesis.

In Chapter 2 the modularity of promoters, herein considered as representative transcription-based components, has been studied in *E. coli* by means of *ad-hoc* constructed model systems of increasing complexity. Results demonstrated that promoters activity (relative to a standard reference promoter) can vary when they are individually measured via different reporter devices (up to 22%), when they are moved to form a two-expression-cassette system (up to 35%) and when used to drive the expression of another device in a functionally-interconnected circuit (up to 44%). This study also elucidated crucial points that can support the bottom-up composition of biological systems: i) not all

the studied promoters are affected by context-dependent activity variation; ii) even when biological components do not functionally interact, their physical context (e.g. the flanking sequences) can compromise their predictability; iii) even if flanking sequences are the main variability source, promoters activity nonlinear change was also observed when surrounding sequences were kept constant; iv) these effects are strain-dependent. An analogy for context-dependent variations in the engineering world is the mutual inductance that could prevent the complete decoupling of different electrical circuits. However, while this effect can be measured and taken into account in electronic engineering, in biological circuits nonlinear and not fully understood context-dependent changes occur. The measurement of such context-dependent variability can provide an estimation of the modularity boundaries that biological engineers should consider when designing a new system from the bottom-up. However, it is important to note that several simplistic design features affect the constructed model systems, e.g. all the promoters had a very similar transcription start site and so, given a downstream device, the produced mRNA molecules are all the same; the growth rate is similar for all the plasmids used in this study for any given bacterial strain. Deviations from these and other conditions may result in higher variability in parts activity among different contexts.

In the philosophy of conducting extensive quantitative studies on biological components, Chapter 3 describes the characterization of a bacterial lysis device that can be controlled by the desired inducible system. A datasheet has been produced for this biological component to support its re-use by future designers. Static and dynamic characteristics elucidated that lysis entity can be actually tuned, but the response time of the device was induction-dependent in a nonlinear fashion. The device has also been demonstrated to be compatible with other *E. coli* strains, growth media or input devices. Failure rate is a very important parameter in the prediction and re-use of a biological part; it has been measured during continual growth of bacterial populations bearing the device and results showed that it is highly strain-dependent.

Chapter 4 reports a mathematical modelling study that supports the de-

sign and construction of proof-of-concept genetic circuits. In particular, the multiplexer and demultiplexer logic functions have been designed to work in a biological chassis and models have been defined to study their quantitative behaviour. Model parameters were identified from the experimental data produced by *ad-hoc* constructed biological devices and critical parameters were found. Crosstalk between components and RBS strength gave the major impact to systems correct functioning.

Finally, Chapter 5 and 6 illustrate the industrially-relevant research made towards the conversion of waste materials into useful compounds by means of synthetic biology key concepts.

Cheese whey is a lactose-rich waste of dairy industries that is classified as an environmental pollutant. *E. coli*, a lactose-utilizing bacterium, has been engineered with an ethanol-production pathway to carry out the conversion of lactose into a biofuel, with the simultaneous disposal of a waste material. The pathway involves the activity of two enzymes, whose expression has to be optimized. A gene-expression platform, including the enzyme-encoding genes under the control of two inducible promoters, has been constructed, incorporated in an *E. coli* strain and tested. Even if lactose-to-ethanol fermentation has been successfully demonstrated, the optimization process still has to be completed. Because the final engineered system is to be used in an industrial framework, several traits will have to be present: fermentation should be carried out with a minimal amount of expensive supplemental nutrients, genetic stability of the biological system must be ensured and no antibiotic resistances must be present in order to avoid the use of antibiotics or the spreading of resistance genes. To support the realization of some of these traits, in this thesis a standard genetic tool for genome engineering is also presented and validated. It is an integrative vector that can be engineered to incorporate BioBrick-compatible biological parts in the genome of the chassis of interest. The vector can also be specialized to target the desired genomic locus.

The other industrially-relevant research is focused on the ambitious goal

of ligno-cellulosic biomass degradation, thus enabling the conversion of the released sugars into the desired compounds, e.g. biofuels. Endoglucanases and exoglucanases attack cellulose to ultimately release the sugar cellobiose, while β-glucosidases convert cellobiose into glucose. The main challenges in the construction of an engineered cellulose-degrading system are the necessary secretion of cellulolytic enzymes for cellulose attack and the not fully understood synergy among the different enzyme species. Synthetic biology offers the ability to combine large numbers of biological components and for this reason it can be a useful tool to discover the synergy among cellulases with the final goal of creating a novel customized cellulolytic organism. Here, different chassis have been used to quantify the expression and secretion of two cellulolytic enzymes. Cellobiose-utilizing non-standard expression chassis have also been characterized and used in order to ultimately avoid the incorporation of recombinant β-glucosidases in a final system. The obtained results highlighted many difficulties in the rational design of such systems, as expression and secretion capabilities were highly context-dependent and different chassis yield different, unpredictable behaviour, just like it was found in the modularity and lysis device studies.

Taken together, the research studies carried out in this thesis have touched many relevant aspects of the quantitative characterization of biological parts and devices: functional modularity, static/dynamic characteristics, failure rate, compatibility, crosstalk between components and chassis-dependent effects are all important features that have to be fully characterized in different case studies to support the rational choice of biological components in the future.

Although great expectations are placed on synthetic biology, this promising field is just at the beginning of its life. Several basic research steps still have to be completed to enable the bottom-up design of customized biological systems, in the same linear manner as electronic or hydraulic circuits are composed from a set of well-characterized parts from a catalogue. Engineered strains for industrial applications would significantly benefit from standard genetic tools that support their development.

Appendix A

Nomenclature of BioBricks

The BioBrick parts in the Registry are identified by a univocal code [47]:

BBa_XN

where the bold-type characters are constant (BB=BioBrick; a=alpha version), while the other characters are variable. X is a letter that indicates the BioBrick class according to Fig. A.1, while N is a numeric code, given by the inventor of the part.

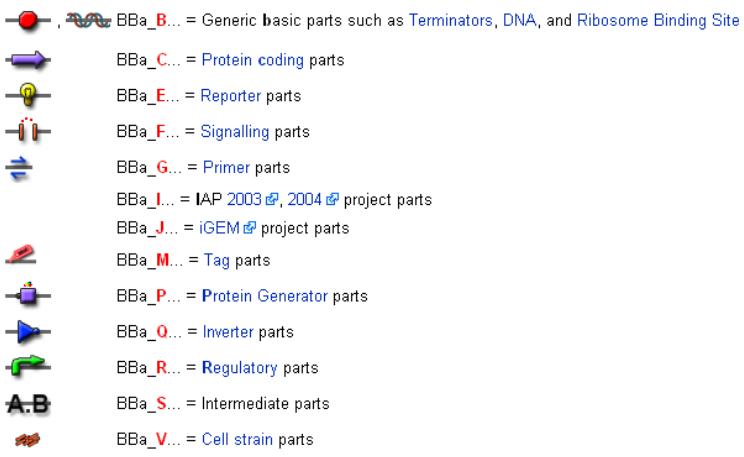


Figure A.1: Identification letters for BioBrick parts. The letter for parts produced in iGEM projects before 2007 is 'J', while it is 'I' for parts produced in 2007 and 'K' for parts after 2007. Reference: [34]

All the BioBricks are contained in specific plasmids that enable the stan-

dard assembly of parts. The standardization is ensured by the structure of prefix and suffix sequences, as described in the previous sections. Given a specific BioBrick standard, despite plasmids must share a common prefix/suffix structure, other parameters can vary, as the replication origin or the antibiotic resistance marker. The nomenclature of BioBrick plasmids is [47]:

pSBNXM

where the bold-type characters are constant (p=plasmid; SB=synthetic biology), while the other characters are variable.

N is a number that indicates the replication origin. This parameter is strictly related to the plasmid copy number, i.e. the number of plasmid DNA copies that are maintained in a recombinant cell. Fig. A.2 reports the replication origins available in the Registry.

Number	Replication origin	Copy number	Purpose
1	modified pMB1 derived from pUC19	500-700	Easy plasmid DNA purification
2	F and P1 lytic derived from pSCANS-1-BNL	1-2 inducible to high copy	Inducible copy number
3	p15A derived from pMR101	10-12	Multi-plasmid engineered systems
4	rep101, repA derived from pSC101	~5	Small cell to cell copy number variation
5	derived from F plasmid	1-2	Improved plasmid stability
6	pMB1 derived from pBR322	15-20	Multi-plasmid engineered systems

Figure A.2: Nomenclature of BioBrick replication origins. Reference: [34].

X is a code that indicates the antibiotic resistance(s) encoded in the plasmid. If more than one resistance is encoded, all the codes are indicated in alphabetical order (e.g. Ampicillin and Kanamycin resistances are indicated as AK). Fig. A.3 reports the antibiotic codes.

M is a number that indicates the version of the plasmid. Different versions can include different signal/regulatory sequences, as reported in Fig. A.4.

So, according to the project specifications, it is possible to choose the suitable plasmid. For example, if the artificial biological system produces a protein that causes toxicity at high concentrations, a low to medium copy number is more convenient than a high copy plasmid.

Code	Antibiotic
A	ampicillin
C	chloramphenicol
E	erythromycin
G	gentamycin
K	kanamycin
N	neomycin
Na	nalidixic acid
R	rifampicin
S	spectinomycin
St	streptomycin
T	tetracycline
Tm	trimethoprim
Z	zeocin

Figure A.3: Nomenclature of antibiotic resistances for BioBrick plasmids. Reference: [34].

Number	Key features	Purpose	Example	Designer
0	absent or incomplete BioBrick cloning site		pSB2K0	Brookhaven National Lab
1	complete BioBrick cloning site (BCS)	assembly of BioBrick parts	pSB4A1	Reshma Shetty
2	5' terminator and BCS	transcriptional insulation of vector upstream of cloned BioBrick part	pSB1A2	Tom Knight
3	5' terminator and BCS and 3' terminator	transcriptional insulation of vector downstream of cloned BioBrick part	pSB1AC3	Reshma Shetty & Tom Knight
4	pSB2K3-derived vector free of many restriction sites	Genome refactoring[1]	pSB2K4	Leon Chan
5	constructed from BioBrick base vector	standardized BioBrick vector design	pSB4K5	Reshma Shetty
6	Reserved for some vector versions that the Berkeley iGEM team has made			
7	BCS flanked by terminator BBa_B0015	transcriptional insulation of cloned BioBrick part	pSB1A7	Karmella Haynes
8	Unassigned			
9	Unassigned			
10	Screening plasmid v1.0	characterization of single input, single output transcriptional devices	pSB1A10	Josh Michener & Jason Kelly
11	Reserved for a version of Jason's measurement plasmid			
12	Reserved for a version of Jason's measurement plasmid			
13	Reserved for a version of Reshma's measurement plasmid			
20-29	Reserved for a version of Berkeley iGEM team's plasmids			

Figure A.4: Codes identifying the version of BioBrick plasmids. Reference: [34].

Appendix B

Methods and supplementary information for Chapter 2

B.1 Methods

B.1.1 Plasmid construction and cloning

BioBrick Standard Assembly 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) [33]. 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, except for pSB4C5(BBa_I52002) which was propagated in chemically competent DB3.1 (Invitrogen), a ccdB toxin-tolerant strain. Ampicillin (100 μg/ml), Chloramphenicol (12.5 μg/ml) or Kanamycin (50 μg/ml) were added as required.

TOP10 and KRX (Promega) strains were used as chassis for all quantitative experiments. Enzymes, purification kits and competent cells were used accord-

ing to manufacturer's instructions. All the plasmids realized in this study were assembled from basic or composite parts from the Registry (see Tab. B.1). BioBrick physical parts were from the MIT 2009 or 2010 DNA Distribution [34]. 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 μg/ml of Chloramphenicol to the growth media. 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.

BioBrick	**Description**
BBa_B0031	weak RBS
BBa_B0032	medium-weak RBS
BBa_E0240	GFP reporter device with BBa_B0032 RBS upstream and BBa_B0015 transcriptional terminator downstream
BBa_F2622	HSL-inducible device (P_{lux} promoter and luxR expression cassette driven by P_{LlacO1})
BBa_I13501	RFP with BBa_B0015 transcriptional terminator downstream
BBa_I13507	RFP reporter device with BBa_B0034 RBS upstream and BBa_B0015 transcriptional terminator downstream
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_P0140	Promoterless tetR expression device with BBa_B0031 RBS upstream and BBa_B0015 terminator downstream
BBa_P0440	Promoterless tetR expression device with BBa_B0034 RBS upstream and BBa_B0015 terminator downstream
BBa_R0011	P_{LlacO1} promoter
BBa_R0040	P_{tetR} promoter
BBa_R0051	P_R promoter from lambda phage

Table B.1: List of the existing BioBricks used as a starting point for all the devices constructed in this study.

B.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. Bacteria were diluted 1:500 in 2 ml of fresh selective M9 supplemented medium and grown for 6 hours under the same conditions as before. When required, properly diluted isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma Aldrich, #I1284) or N-3-oxohexanoyl-L-homoserine lactone (HSL, Sigma Aldrich, #K3007) was added to the liquid culture after 3 hours from the 1:500-dilution to reach the desired final concentration of inducers. 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. In sections 2.2.1 and 2.2.2, strains with P_{LlacO1} were always induced

with 1 mM of IPTG.

B.1.3 Data analysis

Data were analyzed as described in [37] to obtain RPUs. 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 (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 $\overline{S}_{cell}$. The RPU value of a promoter was computed as $\frac{\overline{S}_{cell,\varphi}}{\overline{S}_{cell,ref}}$, where φ is the culture bearing the promoter of interest and *ref* is the reference culture.

The induction curves of the regulated input devices and the input-output transfer curves of the logic inverter were fitted with the Hill function $Y = \delta + V_{max} \cdot \left(\frac{I^n}{K_m^n + I^n}\right)$ or $Y = \delta + V_{max} \cdot \left(1 - \frac{I^n}{K_m^n + I^n}\right)$, where I can be an inducer molecule concentration or an RPU input, δ+V_{max} is the maximum RPU output, K_m is the input that yields Y=δ+V_{max}/2, δ is the basic RPU activity of the promoter in the OFF state and n is the Hill coefficient. All the data were processed either with Microsoft Excel or with the MATLAB 2007b suite (MathWorks, Natick, MA). In particular, the fitting of the Hill functions was performed through the MATLAB *lsqnonlin* routine which implements the least squares method.

All the coefficients of variation were corrected for small samples: for N samples, $CV_{corrected} = CV \cdot \left(1 + \frac{1}{4N}\right)$. Hypothesis tests were performed via MATLAB.

When assessing the statistical difference among the mean promoter activity values of a group, ANOVA test was performed. If a difference was detected in a group, individual t-tests were performed to compare the mean values of the group members to identify statistically different sub-groups. 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 in a 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.

B.2 Additional results

B.2.1 Crosstalk estimation between GFP and RFP spectra

Crosstalk between fluorescent proteins spectra might be an issue regarding systems containing both GFP- and RFP-expressing devices. In order to validate whether this phenomenon was significantly present with the acquisition system used in this study (i.e. the Infinite F200 microplate reader, Tecan), at least in the tested conditions for the combined gene expression cassettes, an *ad-hoc* experiment was performed. TOP10 recombinant cultures bearing J23118 with GFP32, (the only GFP-expressing construct in the combined cassettes study), P_{LlacO1} with RFP34 and P_{lacIQ} with RFP34 (the strongest and the weakest RFP-expressing constructs), all in the low copy vector pSB4C5, were grown as described in B.1. Cultures bearing P_{LlacO1} with RFP34 were induced with 1 mM of IPTG. All the cultures were assayed in the microplate reader (gain=80) as described in B.1. In this experiment, red fluorescence was measured for J23118 with GFP32 and P_{lacIQ} with RFP34, while green fluorescence was measured for J23118 with GFP and P_{LlacO1} with RFP34.

After a proper background subtraction (see B.1, Data analysis section), the $100 \cdot \frac{\overline{S}_{cell,J23118,GFP32}}{\overline{S}_{cell,PlacIQ,RFP34}}$ (red fluorescence) and the $100 \cdot \frac{\overline{S}_{cell,PLlacO1,RFP34}}{\overline{S}_{cell,J23118,GFP32}}$ (green fluorescence) ratios, estimating the maximum percent contribution that GFP and RFP could give to red and green fluorescence measurements respectively,

were computed. No detectable RFP contribution could be observed in the green fluorescence acquisitions, while the resulting ratio was 1.4% when red fluorescence of GFP-expressing cells was acquired. This can be considered as reasonably low crosstalk value.

B.2.2 Preliminary design of the interconnected system

The first design of the interconnected system included a logic inverter with a strong RBS (BioBrick BBa_B0034) upstream of the tetR gene. Such system, however, always remained in the OFF state even when the uninduced P_{lux} promoter, whose basic activity was very low, was assembled upstream. Only the promoterless logic inverter gave a high output when tested (data not shown). This result is consistent with previous findings in which such logic inverter was tested in similar conditions [41]. In order to construct a logic inverter that could switch from the ON state to the OFF state in a range of RPUs between 0.05 and 2, which is exhibited by a number of easy-to-retrieve promoters from the Registry of Standard Biological Parts [37], the RBS upstream of tetR was changed. A much weaker candidate was chosen (BioBrick BBa_B0031). It gave the expected effect (see 2.2.3), as the switch point occurred when the input RPUs were $\sim$0.14.

The low copy vector condition was chosen to characterize the system because such condition can give more reliable results than in high copy vectors [54, 60]. Moreover, attempts in cloning the interconnected circuit in a high copy vector (pSB1A2, which has a pUC19-derived replication origin) gave no successful transformants, suggesting that one (or more) of the modules causes a high metabolic burden when present in >100 DNA copies.

A set of four synthetic constitutive promoters of different strengths was chosen as the INPUT1. All of them are 35-bp long and share a common structure [133]. Inducible lacI- and luxR-regulated promoters were chosen as INPUT2 and INPUT3 respectively. Single-cell analysis reported in [54, 61, 110] showed that these two systems produce a homogeneous response in an induced

cell population in presence of sub-saturating concentrations of IPTG or HSL. KRX *E. coli* strain was used in this study because it overexpresses the LacI repressor through a lacI expression cassette with the lacIq mutation, carried in the F plasmid. This allows the tight transcriptional control of lacI-regulated promoters without including a lacI gene in the circuit.

B.2.3 Characterization of individual promoters in a high copy vector

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 [54]. This was confirmed by comparing the activity of promoters characterized via the same reporter device in low copy and high copy vectors (see Fig. B.1): the RPUs 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. 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). 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.

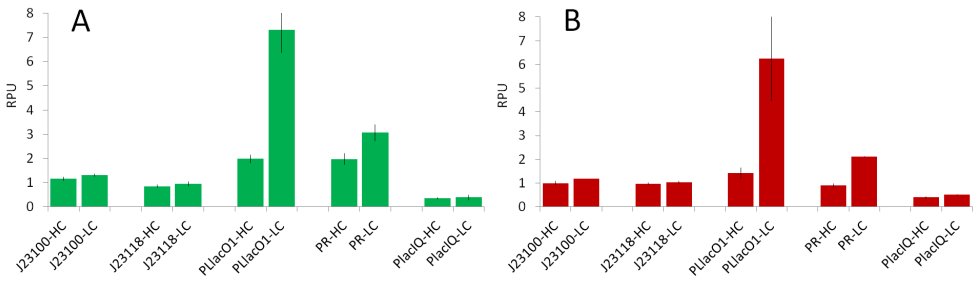


Figure B.1: Measured RPU values for individual promoters characterized in a high or low copy vector, via GFP32 (panel A) and RFP34 (panel B) reporter devices. Error bars represent the standard deviation of the mean activity computed on three clones. Strains with P_{LlacO1} were induced with 1 mM of IPTG.

Appendix C

Methods and supplementary information for Chapter 3

C.1 Methods

C.1.1 Plasmids

All the plasmids used in this study are listed in Tab. C.1.

pSB1A2 and BBa_J61002 are high copy number vectors with a pUC19-derived pMB1 replication origin (∼100-300 molecules per cell) and Ampicillin resistance marker, while pSB4C5 is a low copy number vector with a pSC101 replication origin (∼5 molecules per cell) and Chloramphenicol resistance marker [47]. The full description of vectors and inserts, including their nucleotide sequence, can be found in the BioBrick individual pages in the Registry of Standard Biological Parts web site.

C.1.2 Cloning methods

Chemically competent TOP10 *E. coli* (Invitrogen) were routinely used both for cloning and for quantitative experiments. Chemically competent DB3.1

Plasmids

Name	BioBrick code	Description
pLC-ccdB	pSB4C5(BBa_I52002)[a]	pUC19-derived pMB1 replication origin and ccdB toxin constitutive expression cassette in low copy vector
pHC-T4Lys$^-$	pSB1A2(BBa_K112808)[a]	promoterless lysis device in high copy vector
pHC-HSL	pSB1A2(BBa_F2620)[b]	HSL-inducible promoter in high copy vector
pHC-RFP	BBa_J61002(BBa_J23118)[a]	RFP constitutive expression cassette
pHC-Heat	pSB1A2(BBa_K098995)[a]	heat-inducible promoter in high copy vector
pHC-T4LysHSL	pSB1A2(BBa_K173015)[c]	HSL-inducible lysis device in high copy vector
pLC-HSL	pSB4C5(BBa_F2620)[c]	HSL-inducible promoter in low copy vector
pLC-T4Lys$^-$	pSB4C5(BBa_K112808)[c]	promoterless lysis device in low copy vector
pLC-T4LysHSL	pSB4C5(BBa_K173015)[c]	HSL-inducible lysis device in low copy vector
pLC-T4LysHeat	pSB4C5(BBa_J107014)[c]	heat-inducible lysis device in low copy vector

[a]: taken from the Registry DNA Distribution 2009;
[b]: given by the iGEM Headquarters, Massachusetts Institute of Technology, Cambridge, USA;
[c]: constructed in this study;

Table C.1: Plasmids used in this study. The prefix pHC- indicates a high copy vector backbone (pSB1A2 or BBa_J61002), while pLC- indicates a low copy vector backbone (pSB4C5).

E. coli (Invitrogen) were used to propagate pLC-ccdB plasmid, containing a ccdB expression cassette, which is toxic for TOP10 but not for DB3.1. DH5α (Invitrogen) and MG1655 (purchased from CGSC, Yale University, USA) were only used for quantitative experiments. TOP10, DH5α and DB3.1 were heat-shock transformed according to manufacturer's protocol. MG1655 were made chemically competent with the protocol described in [134] and they were heat-shock transformed at 42°C with the required plasmid. All the strains were routinely grown at 37°C in selective LB medium [134] with Ampicillin (100 μg/ml) or Chloramphenicol (12.5 μg/ml) to propagate plasmids, except for pLC-T4LysHeat, which was grown at 30°C to avoid heat-induction of lysis genes. Plasmids have been purified through QIAprep Spin Miniprep kit (Qiagen) from 5-ml overnight cultures. For every plasmid, 750 μl of culture were mixed with 250 μl of sterile 80% glycerol for long-term storage at -80°C. DNA was digested with EcoRI, XbaI, SpeI or PstI according to BioBrick Standard Assembly procedure [33] and isolated from 1% agarose gel by Gel Extraction Kit (Roche). Cloning of parts was assessed by T4 Ligase and ligation products

were heated at 65°C for 10 min to inactivate T4 Ligase before proceeding with heat-shock transformation. All the enzymes were purchased from Roche and used according to manufacturer's protocol.

TOP10-rfp-lys strain (see Section 3.2.2), which bears two plasmids, pHC-RFP and pLC-T4LysHSL, was built up with the following procedure: competent TOP10 were transformed with pLC-T4LysHSL, then transformed cells were grown in selective LB and made chemically competent again with the protocol described in [134] and heat-shock transformed at 42°C with pHC-RFP. Co-transformants were selected using LB medium with Ampicillin at 100 μg/ml and Chloramphenicol at 12.5 μg/ml.

LB and M9 supplemented medium (11.28 g/l M9 salts, 1 mM thiamine hydrochloride, 2 mM MgSO4, 0.1 mM CaCl2, 0.2% casamino acids and 0.4% vol/vol glycerol as carbon source) [134] were used in quantitative experiments.

DNA sequencing was performed through BMR Genomics (Padova, Italy) DNA analysis service.

C.1.3 Plasmid construction

pSB4C5 low copy vector bulk, obtained from pLC-ccdB after EcoRI-PstI cut, was ligated to the inserts of pHC-T4Lys$^-$ and pHC-HSL cut with EcoRI-PstI to yield pLC-T4Lys$^-$ and pLC-HSL respectively. pHC-T4LysHSL was constructed by assembling the insert of pHC-HSL cut with EcoRI-SpeI to pHC-T4Lys$^-$ cut with EcoRI-XbaI. pLC-T4LysHSL was constructed by assembling the insert of pHC-T4Lys$^-$ cut with XbaI-PstI to pLC-HSL cut with SpeI-PstI. Finally, pLC-T4LysHeat was constructed by assembling the insert of pHC-Heat cut with EcoRI-SpeI to pLC-T4Lys$^-$ cut with EcoRI-XbaI.

C.1.4 Lysis assays

Unless otherwise noted, 5 μl of bacteria bearing pLC-T4LysHSL and pLC-T4Lys$^-$ 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 was 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.

For lysis assays on pLC-T4LysHeat, the cultures were grown at 30°C instead of 37°C and an automatic temperature shift from 30°C to 42°C was used to induce lysis in the microplate reader instead of HSL.

C.1.5 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(OD600(t))* over time. Then, the exponential phase was identified by visual inspection as the linear region of *ln(OD600(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(OD600(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 $100 * (1 - min_{OD600}/ref_{OD600})$, where min_{OD600} is the minimum OD_{600} reached by the culture and ref_{OD600} is the OD_{600} immediately before the density drop caused by cell lysis. The rise time was computed as the time required for the $100 * (1 - OD_{600}(t)/ref_{OD600})$ signal to rise from 10% to 90% of the lysis entity.

C.1.6 Protein release assays

5 μl of TOP10-rfp-lys, TOP10 bearing pHC-RFP and TOP10 bearing pLC-T4Lys$^-$ glycerol stocks were inoculated in 5 ml of selective LB medium and grown at 37°C, 220 rpm overnight. TOP10-rfp-lys and TOP10 with pHC-RFP cultures were diluted 1:100 into six 15-ml tubes containing 5 ml of pre-warmed selective LB medium. TOP10 with pLC-T4Lys$^-$ culture was diluted 1:100 in one 15-ml tube with 5 ml of pre-warmed selective LB. The 13 resulting cultures were grown to an OD_{600} of about 0.55 (exponential phase). Three of the six replicates of TOP10-rfp-lys and TOP10 with pHC-RFP were induced with HSL at a final concentration of 100 nM and all the 13 tubes were incubated under the same conditions as before for 125 min. Every 25 min, the OD_{600} was measured with the NanoDrop ND-1000 and a 300-μl aliquot was taken from each culture. The 300-μl samples were centrifuged at 11000 rpm in a table top centrifuge and the fluorescence of 200 μl of the supernatant was measured in a microplate with the Infinite F200 reader using: 535 nm excitation filter, 620 nm emission filter, excitation bandwidth = 25 nm, emission bandwidth = 20 nm, gain = 55, number of flashes = 25, integration time = 20 μs, top reading. The raw fluorescence measurements of TOP10-rfp-lys and TOP10 bearing pHC-

RFP (induced and uninduced) were normalized by subtracting the background fluorescence value of pLC-T4Lys$^-$ cells at the same time points.

To measure the percentage of the RFP released in the medium after 125 minutes, TOP10-rfp-lys, TOP10 bearing pHC-RFP and TOP10 bearing pLC-T4Lys$^-$ were grown and induced as described above, in 3-ml cultures. At t = 125 min, the cultures were centrifuged (4000 rpm, 4°C for 10 min) and supernatants and pellets were processed to measure the amount of intracellular and extracellular RFP. Pellets were resuspended in 200 μl of sterile LB broth and 200 μl of 2X lysis buffer (25 mM of Tris-HCl at pH 8.0, 8% SDS) were added to the resuspended pellets [135]. The resuspended pellets with the lysis buffer were left at room temperature for 15 min, then transferred in a 1.5-ml tube and centrifuged at 13000 rpm for 15 min in a table top centrifuge to spin down the cell debris. 200 μl of these supernatants were transferred in a microplate and the RFP fluorescence was measured as described above, to evaluate the intracellular RFP amount. To evaluate the extracellular RFP concentration, the growth media were treated as follows: 100 μl of media were mixed with 100 μl of 2X lysis buffer, left at room temperature for 15 min, transferred in a microplate, then the fluorescence was measured in the same way as the lysed pellets. The 2X lysis buffer was added to reproduce the same conditions as the pellets.

The raw intracellular and extracellular RFP measurements were normalized by subtracting the background fluorescence of pLC-T4Lys$^-$ lysed pellet and supernatant respectively. Finally, the mean RFP values of lysed pellets and supernatants of TOP10-rfp-lys and TOP10 bearing pHC-RFP (induced and uninduced) were corrected by the total culture volume, so the percentage of extracellular RFP molecules was computed for each culture.

C.1.7 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 path-

length 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.

C.1.8 Evolutionary stability characterization

Bacteria bearing pLC-T4LysHSL were propagated for 100 generations as described in [76] without adding HSL and every 10 generations a lysis assay was performed to test the stability of the lysis phenotype. Stability was assessed by measuring lysis entity in each assay. In particular, in order to achieve 100 generations, the culture was diluted 1:1000 every 24 hours, thus yielding about 10 generations per day ($\log_2 1000 = 9.97$) [76]. Every day, an aliquot was taken from the propagated culture and lysis was assayed as described above. This experiment was performed in triplicate for each tested strain.

C.1.9 Analysis of mutants

To analyze mutants, a lysis assay was performed on TOP10 bearing pLC-T4LysHSL induced with HSL 10 nM and when the lysed and re-grown cells reached an $OD_{600} = 0.22$ they were diluted 1:1000 in fresh selective LB medium in a new 96-well microplate, incubated in the Infinite F200 reader and induced again with HSL 10 nM when they reached an OD_{600} of 0.35 (exponential phase) to check if they could lyse again. The re-grown cells were also streaked on selective LB agar, then 2 single colonies were inoculated in 5 ml of selective LB and let grow overnight (37°C, 220 rpm). Plasmid DNA was extracted from the 5-ml overnight cultures and analyzed by restriction enzyme digestion/electrophoresis and DNA sequencing using primers VF2 (TGCCACCTGACGTCTAAGAA), VR (ATTACCGCCTTTGAGTGAGC) and C0062VF (GAATGTTTAGCGTGGGCATG). This procedure was carried out starting from 5 μl of pLC-T4LysHSL glycerol stock and also from single colonies isolated from this stock.

C.2 Additional results

C.2.1 Sequence analysis of the used BioBrick parts

BBa_K112808 sequencing, performed on the DNA submitted to the Registry, showed discrepancies when compared to its submitted nucleotide sequence: there was a silent point mutation in the holin gene (c70a); the transcriptional terminator at the end of the part was actually BBa_K112710 and not BBa_B0010; the alignment showed several gaps in the DNA scars between the basic parts. However, the regulatory parts and the amino acid sequences of the translated genes were correct and the lysis assays showed that BBa_K112808 is fully functional. BBa_K098995 showed discrepancies as well: the DNA scar between the cIts gene and its RBS was not present; there were also two point mutations in the cIts gene (g198a and g351a of cIts) resulting in two amino acid substitutions (A67T and E117K respectively), but they do not appear to affect the heat-induction capability of this part, as shown in the main text.

C.2.2 Characterization of the HSL-inducible lysis device in high-copy plasmid

Lysis was assayed in a 96-well microplate by measuring the OD_{600} dynamics of TOP10 bearing HSL-inducible lysis device pHC-T4LysHSL, induced with HSL and uninduced. Induced and uninduced TOP10 bearing the promoterless lysis device pHC-T4Lys$^-$ were used as negative controls. On 10 experiments carried out in separate days, bacteria did not lyse upon induction in half of them (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. TOP10 bearing the HSL-inducible lysis device pHC-T4LysHSL 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 P_{lux} promoter present in high copy number, while the negative control pHC-T4Lys$^-$ grew faster (doubling time of 45 ± 4.7 min). In pHC-T4LysHSL cultures 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 in Tab. C.2.

TOP10 with pHC-T4LysHSL

Growth phase:	**Exponential**	**Early stationary**	**Late stationary**
Lysis entity [%]	58.6±7.5[b]	42.2±9.9[c]	0[a]
Lysis delay after induction [min]	15[b]	15[c]	-
Doubling time [min]	69±6[b]		
Doubling time when lysis failed [min]	47.4±2.1[b]		
Doubling time of negative control [min]	45±4.7[d]		

[a]: measured on 3 independent experiments
[b]: measured on 5 independent experiments
[c]: measured on 2 independent experiments
[d]: measured on 10 independent experiments

Table C.2: Quantitative characterization of TOP10 bearing the HSL-inducible lysis device in high copy plasmid pHC-T4LysHSL grown at 37°C in microplate. Induction was carried out with HSL 100 nM. Mean lysis entity and lysis delay after the induction are reported together with their standard error for the different growth phases. The doubling time of the uninduced lysis device and its negative control are reported too. The doubling time of the induced cultures which failed lysis is also shown.

Appendix D

Methods and supplementary information for Chapter 4

D.1 Methods

D.1.1 Cloning and parts construction

BioBrick parts from the MIT Registry of Standard Biological Parts were used to construct all the plasmids. They were assembled according to the BioBrick Standard Assembly. In particular, parts from MIT Spring 2008 DNA Distribution, spotted on filter paper, were resuspended in warm (55°C) TE buffer 10:1, pH 8.0, and used to transform the TOP10 competent of *E. coli* (Invitrogen) by heat-shock, according to manufacturer's protocol. Plasmids were purified using a QIAprep Spin Miniprep kit (Qiagen) from 5-ml overnight cultures. For every plasmid, 800 μl of culture were mixed with 200 μl of sterile glycerol for long-term storage at -80°C. Parts or vector bulks were obtained through EcoRI, XbaI, SpeI, PstI (Roche) digestion of 1 μg of DNA for 3 h at 37°C and isolated from 1% agarose gel using a DNA Gel Extraction Kit (Roche). Cloning of parts was assessed by T4 Ligase (Roche) activity on XbaI-PstI inserts downstream of SpeI-PstI parts or EcoRI-SpeI inserts upstream of EcoRI-XbaI parts.

Ligations were performed overnight at 16°C in 10 μl with a 1:6 molar ratio (bulk/insert) and heated at 65°C for 10 min before proceeding to heat-shock transformation of TOP10.

D.1.2 Fluorescence assays

1-5 μl of recombinant bacteria were recovered from thawed glycerol stocks and grown in 1-5 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, Ampicillin 100 μg/ml) for about 20 h at 37°C and 220 rpm shaking. Cultures were diluted 1:100 in 1-5 ml of fresh selective M9 medium and incubated for 5 h under the same conditions as before. 200-μl aliquots were transferred into a flat-bottomed 96-well microplate (Greiner) and assayed for 3 h with an automatically repeating protocol with the Infinite F200 microplate reader (Tecan): fluorescence (excitation at 485 nm, emission at 540 nm, gain = 50 for GFP; and excitation at 535 nm, emission at 620 nm, gain = 50 for RFP) and absorbance (600 nm) were measured every 5 min. The temperature was kept constant at 37°C and every 5 min the cultures were shaken (3 mm amplitude, linear shaking) for 15 s. Induction experiments were performed analogously, except that cultures were induced in the microplate wells with 2 μL of $3OC_6HSL$ (Sigma Aldrich #K3007) to yield the desired final concentration.

The absorbance of 200 μL of sterile M9 and the autofluorescence of a TOP10 culture without GFP or RFP were measured to estimate the absorbance and fluorescence backgrounds, respectively. M9 and TOP10 background values were subtracted from all absorbance and fluorescence raw measurements to obtain values proportional to the per-well cell count and number of fluorescent molecules.

The fluorescent reporter protein synthesis rate per cell time series of each culture was computed as the numeric time derivative of the fluorescence values divided by absorbance. This time series was averaged over a 2-h period

during the bacterial exponential growth phase and, when dealing with induced cultures, after about 50 min from the induction time [40].

D.1.3 $3OC_6HSL$ production assay

Bacteria with the RBS-luxI test parts were recovered from thawed glycerol stocks and were grown in 1-5 ml of selective M9-supplemented medium for about 20 h at 37°C and 220 rpm shaking. Cultures were diluted 1:100 in 1-5 ml of fresh selective M9 medium and incubated for 8 h under the same conditions as before. Cultures were centrifuged (7000 rpm, 10 min) and the supernatant was filtered (0.2 μm) to remove all the bacterial cells. Supernatants were stored at -20°C until further use.

TOP10 with BioBrick BBa_T9002 in pSB1A3 high-copy vector were used as a $3OC_6HSL$ biosensor, which synthesizes GFP as a function of $3OC_6HSL$ concentration [40]. TOP10 with BBa_T9002 were recovered from thawed glycerol stocks and were grown in 1-5 μl of M9-supplemented medium for about 20 h. The culture was diluted 1:100 in a proper volume of fresh selective M9 medium and, after 5 h, 180-μl aliquots were transferred to a 96-well microplate. Wells were induced with 20 μl of the supernatants obtained before and the BBa_T9002 culture was assayed, as described above, in the microplate reader for 3 h. Standard calibration curves were obtained by inducing the wells with 20 μl of M9 medium supplemented with $3OC_6HSL$ at different autoinducer concentrations. When the measured concentrations were higher than the biosensor upper detection limit, the supernatants were diluted properly with fresh M9 medium to a final volume of 20 μl, which was used to induce BBa_T9002 wells. Measurements were performed in triplicate. Background values for absorbance and fluorescence were subtracted as described above. The average GFP synthesis rate per cell was computed over a 3-h period to indirectly measure the $3OC_6HSL$ concentration.

D.2 Mathematical model

Ordinary differential equations with Hill functions were used to model the mux and demux genetic circuits. Equations and model parameters definition are reported below, as well as the parameter estimation procedures performed on experimental data expressly acquired as a part of this study. The model was implemented in MATLAB (MathWorks, Natick, MA) routines and all the simulations reported herein were performed using the MATLAB 2010a suite. Note that RFP was considered as the mux single output (GOI), while RFP and GFP were considered as the two outputs of demux (GOI0 and GOI1, respectively).

The following differential equations with Hill functions were used to model the mux and demux genetic circuits, including $3OC_6HSL$ (H), $3OC_{12}HSL$ (P), cI (C), LasR (L), LuxR, LuxR-autoinducer complex (XH), LasR-autoinducer complex (LP), RFP (R) and GFP (G) species (see Parameter Table D.1, reported below, for parameter definition), under the assumption that the intra- and extra-cellular autoinducer concentrations were the same.

Mux output was RFP (under the control of both P_{lux} and P_{las}), while demux outputs were RFP and GFP, respectively. If not differently stated, the equations are valid for both mux and demux. Crosstalk between lux and las systems was also considered, in fact the unspecific activation of transcription factors X and L by P and H respectively and the unspecific activation of P_{lux} or P_{las} by LP and XH respectively were taken into account.

$$\frac{d[H]}{dt} = \alpha_H(P_A)\cdot N - \gamma_H \cdot H \qquad (for\ mux) \tag{D.1}$$

$$\frac{d[P]}{dt} = \alpha_P(P_B)\cdot N - \gamma_P \cdot P \qquad (for\ mux) \tag{D.2}$$

$$\frac{d[H]}{dt} = \alpha_H(P_I)\cdot N - \gamma_H \cdot H \qquad (for\ demux) \tag{D.3}$$

$$\frac{d[P]}{dt} = \alpha_P(P_I)\cdot N - \gamma_P \cdot P \qquad (for\ demux) \tag{D.4}$$

$$\frac{d[C]}{dt} = P_S - (\gamma_C + K - \frac{K\cdot N}{N_{max}})\cdot C \tag{D.5}$$

$$\frac{d[L]}{dt} = P_S - (\gamma_L + K - \frac{K \cdot N}{N_{max}}) \cdot L \tag{D.6}$$

$$\frac{d[X]}{dt} = \alpha_\lambda \cdot \left(\delta_\lambda + \frac{1-\delta_\lambda}{1+\left(\frac{C}{K_C}\right)^{n_C}} \right) - (\gamma_X + K - \frac{K \cdot N}{N_{max}}) \cdot X \tag{D.7}$$

$$XH = \frac{X}{1+\left(\frac{K_H}{H+u \cdot P}\right)^{n_H}} \tag{D.8}$$

$$LP = \frac{L}{1+\left(\frac{K_P}{P+v \cdot H}\right)^{n_P}} \tag{D.9}$$

$$\frac{d[R]}{dt} = \alpha_X \cdot \left(\delta_X + \frac{1-\delta_X}{1+\left(\frac{K_H}{XH+w \cdot LP}\right)^{n_X}} \right) + \\ + \alpha_L \cdot \left(\delta_L + \frac{1-\delta_L}{1+\left(\frac{K_L}{LP+z \cdot XH}\right)^{n_L}} \right) - (K - \frac{K \cdot N}{N_{max}}) \cdot R \qquad (for\ mux) \tag{D.10}$$

$$\frac{d[R]}{dt} = \alpha_X \cdot \left(\delta_X + \frac{1-\delta_X}{1+\left(\frac{K_H}{XH+w \cdot LP}\right)^{n_X}} \right) - (K - \frac{K \cdot N}{N_{max}}) \cdot R \qquad (for\ demux) \tag{D.11}$$

$$\frac{d[G]}{dt} = \alpha_L \cdot \left(\delta_L + \frac{1-\delta_L}{1+\left(\frac{K_L}{LP+z \cdot XH}\right)^{n_L}} \right) - (K - \frac{K \cdot N}{N_{max}}) \cdot R \qquad (for\ demux) \tag{D.12}$$

$$\frac{d[N]}{dt} = N \cdot \left(K - \frac{K \cdot N}{N_{max}} \right) \tag{D.13}$$

H and P concentrations were expressed in nM, while N was expressed in cells per well. RFP synthesis rate per cell produced by the promoters was used to approximate protein synthesis rates. As a consequence, all the protein concentration levels C, X, L, XH, LP and R were expressed in arbitrary units of RFP (AU_R) per cell, except G which was expressed in arbitrary units of GFP (AU_G) per cell.

Experimental data have been used to estimate some model parameters. Bacterial growth parameters K and N_{max} (Eq.D.13) were estimated respectively by computing the growth rate of *E. coli* in M9 supplemented medium and by plating out properly diluted bacteria in stationary phase.

The $\alpha_H(P_A)$ function was estimated from the numerical solution of Eq.D.1 at t=8 hours and the experimental data shown in Fig. 4.7C. For this analysis, the initial value H(0) was set to $\frac{N_{max} \cdot \alpha_H(P_A)}{100 \cdot \alpha_H}$ to take into account the $3OC_6HSL$ produced by the overnight culture before the 1:100 dilution.

K_X was estimated by fitting the experimental data shown in Fig. 4.8

(dashed line) with Eq.D.7, D.8 and D.10 at the steady state, considering C=0.

α_X and δ_X were estimated respectively as the maximum RFP synthesis rate reached in Fig. 4.8 (dashed line) and as the minimum divided by the maximum rate in the same curve.

For the other model parameters the literature references are reported in the parameter table (D.1). If not differently indicated, parameter values reported in this table were used in all the simulations reported and discussed in this work.

Parameter definition and units.

If no reference is present, the parameter value was estimated in this study.

Parameter	Description	Value and reference	Units
A	Protein synthesis rate driven by mux input A promoter	variable	AU_R min^{-1} $cell^{-1}$
B	Protein synthesis rate driven by mux input B promoter	variable	AU_R min^{-1} $cell^{-1}$
I	Protein synthesis rate driven by demux input promoter	variable	AU_R min^{-1} $cell^{-1}$
S	Protein synthesis rate driven by mux or demux selector promoter	variable	AU_R min^{-1} $cell^{-1}$
K	Bacterial growth rate in M9	0.0116	min^{-1}
N_{max}	Maximum number of cells in a saturated culture	$2 \cdot 10^9$	cells per well
$\alpha_H(input)$	$3OC_6HSL$ synthesis rate as a function of the input promoter activity	$3 \cdot 10^{-12} \cdot input^2 + 4 \cdot 10^{-12} \cdot input$	nM of $3OC_6HSL$ min^{-1} $cell^{-1}$
$\alpha_P(input)$	$3OC_{12}HSL$ synthesis rate as a function of the input promoter activity	$3 \cdot 10^{-12} \cdot input^2 + 4 \cdot 10^{-12} \cdot input^a$	nM of $3OC_{12}HSL$ min^{-1} $cell^{-1}$
α_λ	P_λ maximum synthesis rate	28.54 [internal unpublished results]	AU_R min^{-1} $cell^{-1}$
α_X	P_{lux} maximum synthesis rate	100	AU_R min^{-1} $cell^{-1}$
α_L	P_{las} maximum synthesis rate	7.5[b]	AU_R min^{-1} $cell^{-1}$
δ_λ	Basal activity of P_λ	$3.75 \cdot 10^{-4}$ [136]	-
δ_X	Basal activity of P_{lux}	0.01	-
δ_L	Basal activity of P_{las}	0.01 [26]	-
K_H	Dissociation constant of $3OC_6HSL$-LuxR	553[c] [137]	nM
K_P	Dissociation constant of $3OC_{12}HSL$-LasR	200 [26]	nM
K_C	Dissociation constant of cI-P_λ	6[d]	AU_R min^{-1} $cell^{-1}$

K_X	Dissociation constant of $3OC_6HSL$-LuxR complex and P_{lux}	0.11	AU_R min^{-1} $cell^{-1}$
K_L	Dissociation constant of $3OC_{12}HSL$-LasR complex and P_{las}	0.2[c] [26]	AU_R min^{-1} $cell^{-1}$
n_H	Hill cooperativity constant of $3OC_6HSL$-LuxR	2 [86]	-
n_P	Hill cooperativity constant of $3OC_{12}HSL$-LasR	1.4 [26]	-
n_C	Hill cooperativity constant of cI-P_λ	2 [138]	-
n_X	Hill cooperativity constant of $3OC_6HSL$-LuxR complex and P_{lux}	1 [139]	-
n_L	Hill cooperativity constant of $3OC_{12}HSL$-LasR complex and P_{las}	1 [139]	-
γ_H	$3OC_6HSL$ degradation rate at pH 7.0	0.01 [140]	min^{-1}
γ_P	$3OC_{12}HSL$ degradation rate at pH 7.0	0.01[a]	min^{-1}
γ_C	cI degradation rate	0.0173 [141]	min^{-1}
γ_X	LuxR degradation rate	0.06 [137]	min^{-1}
γ_L	LasR degradation rate	0.0173 [141]	min^{-1}
u	Relative weight for unspecific activation of X by P	0	-
v	Relative weight for unspecific activation of L by H	0	-
w	Relative weight for unspecific activation of P_{lux} by LP	0	-
z	Relative weight for unspecific activation of P_{las} by XH	0	-

[a] $\alpha_P(input)$ and γ_P were set equal to $\alpha_H(input)$ and γ_H respectively in order to perform simulations, as no experimental data are available to estimate these parameters in the conditions required in this work.

[b] Estimated considering that the tetR promoter (P_{tetR}) is 4.3-fold stronger than P_{las} at full induction [26] and the P_{tetR} RFP synthesis rate per cell was measured to be ∼32 AU_R min^{-1} $cell^{-1}$ from internal unpublished results of our lab. The resulting value is 7.5 AU_R min^{-1} $cell^{-1}$ (or 75 AU_G min^{-1} $cell^{-1}$ because the GFP output is approximately 10-fold higher than the RFP).

[c] Computed considering an *E. coli* cell volume of 10^{-15} liters.

[d] Data submitted to the Registry (part BBa_Q04510 web page) show the behaviour of P_λ as a function of promoters upstream of cI with a strong RBS. Considering the relative strength of these promoters (reported in the BBa_J23100 web page), a strength below 0.4 (∼41% of BBa_J23114 activity) but higher than 0.066 (∼6.6% of BBa_J23114 activity) AU_R min^{-1} $cell^{-1}$ is required to switch P_λ on. A switch value $P_{S,switch}$=0.17 has been hypothesized for P_S, so, considering equation D.5 at the steady state in exponential growth phase, $K_c = C_{switch} = P_{S,switch}/(\gamma_C + K)$.

[e] Adjusted in order to yield the same switch point as P_{las} transfer function reported in [26] as a function of $3OC_{12}HSL$. Equations D.6, D.9 and D.12 were used considering the steady state. In [26], recombinant bacteria were grown in LB, the lasR gene did not have the LVA tag and it was constitutively expressed by a promoter about 11-fold weaker than J23101 (data submitted to the Registry, BBa_J23117 web page). So, considering Eq. D.6 in exponential growth phase, L was set at the constant value of $26.6/(11 \cdot K)$ during the adjustment. In this case K was set at 0.015 (doubling time of 45 min) to take into account the higher growth rate of *E. coli* in LB than in M9.

Table D.1: Parameter table for mux and demux mathematical model.

D.3 Additional results

D.3.1 NOT gate repression capability when coupled to the luxR-P_{lux} system

NOT gate coupled to the luxR-P_{lux} system was studied by simulating the RFP synthesis rate per cell at the steady state as a function of $3OC_6HSL$ (H) and the activity of the promoter (P_S) upstream of cI gene. cI (C) and LuxR (X) proteins were considered at the steady state, so $C = \frac{P_S}{\gamma_C}$ and $X = \frac{\alpha_\lambda}{\gamma_X} \cdot \left(\delta_\lambda + \frac{1-\delta_\lambda}{1+\left(\frac{C}{K_C}\right)^{n_C}} \right)$. Then, XH and RFP synthesis rate per cell can be computed as $X \cdot \left(\frac{1}{1+\left(\frac{K_H}{H}\right)^{n_H}} \right)$ and $\alpha_X \cdot \left(\delta_X + \frac{1-\delta_X}{1+\left(\frac{K_X}{XH}\right)^{n_X}} \right)$ respectively.

A simulation of this system is reported and discussed in the main text (Fig. 4.9C). Here, a further simulation is reported decreasing α_λ from 28.54 (value estimated for the real biological part) to 3, in order to decrease the LuxR steady-state value by reducing its maximum synthesis rate per cell. As

expected, Fig. D.1 shows a better repression capability of P_S at high H values, when compared to Fig. 4.9C.

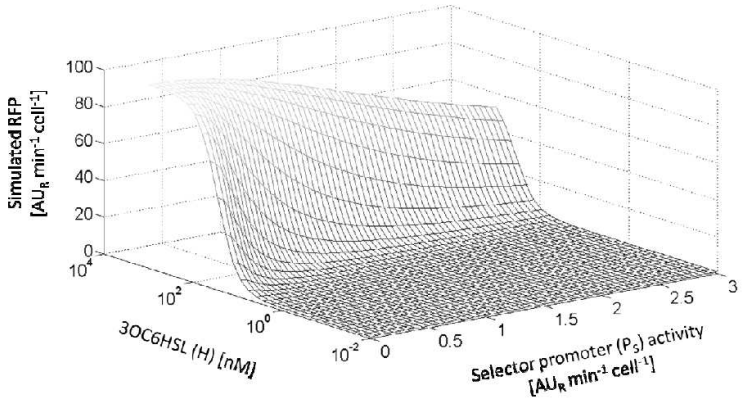


Figure D.1: Simulated RFP output of the NOT gate coupled with the luxR-P_{lux} system, as a function of 3OC$_6$HSL and of the activity of P_S promoter expressing cI gene. In this simulation, the maximum activity of P_λ promoter (α_λ) was set to 3.

D.3.2 Crosstalk analysis

Crosstalk effect between lux and las systems was studied as a function of crosstalk parameters u, v, w and z (see D.2 for parameter definition), which, for simplicity, were varied together even though they could assume independent values. Mux and demux overall performances, in terms of fold change between the expected ON and OFF output states, are reported in Fig. D.2.

The simulations were performed considering ON state = 3 AU$_R$ min^{-1} cell^{-1} and OFF state = 0.03 AU$_R$ min^{-1} cell^{-1} for all the input/selector promoters. Simulation results depicted that crosstalk can impair the two devices behaviour. In fact, the fold change between ON and OFF output states became close to 1, i.e. there is no difference between ON and OFF output states, increasing the values of crosstalk parameters. This did not occur for GFP in demux, which maintained a fold change higher than 7 for all the spanned crosstalk parameters values, even if it showed a decreasing trend. Finally, it is noteworthy that for a range of crosstalk parameters values between 0 and 0.1,

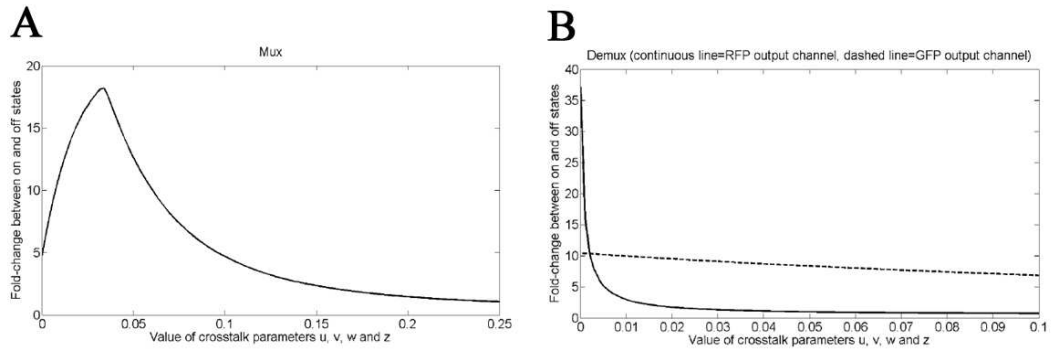


Figure D.2: Simulated fold change between the expected ON and OFF output states of mux RFP output and demux RFP (continuous line) and GFP (dashed line) output channels, as a function of the crosstalk parameters u, v, w and z when they are varied together.

mux seems to work better than in absence of crosstalk (i.e. when u, v, w and z = 0), as there was a higher difference between ON and OFF output states. This result was unexpected and is probably due to the unspecific activation of the strong promoter P_{lux} by LasR in the activated form, which 'amplified' the ON state output values of mux and gave a higher fold change between ON and OFF states.

Appendix E

Methods and supplementary information for Chapter 5

The contents of this appendix are confidential.

Appendix F

Methods and supplementary information for Chapter 6

The contents of this appendix are confidential.

Bibliography

[1] Pasotti L and Magni P. BioBrick, Standard Assembly e registro delle parti biologiche standard. In *Biologia Sintetica*, pages 179–204. Patron Ed., Bologna, Italy, 2010.

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

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

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

[5] Jha A. From the cells up. *The Guardian*, March 10, March 2005.

[6] Serrano L. Synthetic biology: promises and challenges. *Molecular Systems Biology*, 3(158), 2007.

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

[8] Kwok R. Five hard truths for synthetic biology. *Nature*, 463:288–290, 2010.

[9] Carlson R. The changing economics of DNA synthesis. *Nature Biotechnology*, 27:1091–1094, 2009.

[10] Cello J, Paul A, and Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, 297(5583):1016–1018, 2002.

[11] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison III CA, Smith HO, and Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329(5987):52–56, 2010.

[12] Gibson DG. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research*, 37(20):6984–6990, 2009.

[13] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, and Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6:343–345, 2009.

[14] Mansy SS and Szostak JW. Thermostability of model protocell membranes. *PNAS*, 105(36):13351–13355, 2008.

[15] Stano P and Luisi PL. Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chemical Communications*, 46(21):3639–3653, 2010.

[16] Voigt CA. Genetic parts to program bacteria. *Current Opinion in Biotechnology*, 17(5):548–557, 2006.

[17] Kaznessis YN. Models for synthetic biology. *BMC Systems Biology*, 1(47), 2007.

[18] Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, and Keasling JD. BglBricks: a flexible standard for biological part assembly. *Journal of Biological Engineering*, 4(1), 2010.

[19] Anderson JC, Clarke EJ, Arkin AP, and Voigt CA. Environmentally controlled invasion of cancer cells by engineered bacteria. *Journal of Molecular Biology*, 355(4):619–627, 2006.

[20] Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, and Voigt CA. Engineering the *Salmonella* type III secretion system to export spider silk monomers. *Molecular Systems Biology*, 5(309), 2009.

[21] Aldaye FA, Senapedis WT, Silver PA, and Way JC. A structurally tunable DNA-based extracellular matrix. *Journal of the American Chemical Society*, 132(42):14727–14729, 2010.

[22] Savage DF, Way J, and Silver PA. Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chemical Biology*, 3(1):13–16, 2008.

[23] French CE. Synthetic biology and biomass conversion: a match made in heaven? *Journal of the Royal Society Interface*, 6(Suppl 4):S547–S558, 2009.

[24] Clarke EJ and Voigt CA. Characterization of combinatorial patterns generated by multiple two-component sensors in *E. coli* that respond to many stimuli. *Biotechnology and Bioengineering*, 108(3):666–675, 2011.

[25] Anderson JC, Voigt CA, and Arkin AP. Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, 3(133), 2007.

[26] Tamsir A, Tabor JJ, and Voigt CA. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469:212–215, 2010.

[27] Regot S, Macia J, Conde N, Furukawa K, Kjellen J, Peeters T, Hohmann S, Nadal ED, Posas F, and Solé R. Distributed biological computation with multicellular engineered networks. *Nature*, 469:207–211, 2010.

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

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

[30] Gardner TS, Cantor CR, and Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000.

[31] Ro DK, Paradise EM, Ouellet M, Fisher K, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, and Keasling JD. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440:940–943, 2006.

[32] Endy D. Foundations for engineering biology. *Nature*, 438:449–453, 2005.

[33] Knight TF. Idempotent vector design for standard assembly of biobricks. Technical report, MIT DSpace [http://web.mit.edu/synbio/release/docs/biobricks.pdf], 2003.

[34] MIT. Registry of Standard Biological Parts. [http://partsregistry.org].

[35] International Genetically Engineered Machine (iGEM) [http://www.igem.org].

[36] de Mora K, Joshi N, Balint BL, Ward FB, Elfick A, and French CE. A pH-based biosensor for detection of arsenic in drinking water. *Analytical and Bioanalytical Chemistry*, 400(4):1031–1039, 2011.

[37] Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, de Mora K, Glieberman AL, Monie DD, and Endy D. Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *Journal of Biological Engineering*, 3(4), 2009.

[38] Miller JH. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1972.

[39] Canton B. *Engineering the interface between cellular chassis and synthetic biological systems*. PhD thesis, Massachusetts Institute of Technology, 2008.

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

[41] Kelly JR. *Tools and reference standards supporting the engineering and evolution of synthetic biological systems*. PhD thesis, Massachusetts Institute of Technology, 2008.

[42] Pollack A. Custom-made microbes at your service. *New York Times*, January 17, 2006.

[43] SYNBIOSAFE project. [http://www.synbiosafe.eu].

[44] Lutz R and Bujard H. 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.

[45] NCBI. The genetic codes. [http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi], last update: July 07, 2010.

[46] Salis HM, Mirsky E, and Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10):946–950, 2009.

[47] Shetty RP, Endy D, and Knight TF. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, 2(5), 2008.

[48] BioBricks Foundation (BBF). [http://biobricks.org].

[49] BBF RFC process. [http://openwetware.org/wiki/The_BioBricks_Foundation:RFC].

[50] Davis JH, Rubin AJ, and Sauer RT. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research*, 39(3):1131–1141, 2011.

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

[52] Heidelberg iGEM Team 2009. Spybricks - a starter kit for synthetic biology in mammalian cells. [http://2009.igem.org/Team:Heidelberg].

[53] Ellis T, Wang X, and Collins JJ. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology*, 27(5):465–471, 2009.

[54] Zucca S, Pasotti L, Mazzini G, Cusella De Angelis MG, and Magni P. Characterization of an inducible promoter in different DNA copy number conditions. *BMC Bioinformatics (accepted)*, 2012.

[55] Sauro HM. Modularity defined. *Molecular Systems Biology*, 4(166), 2008.

[56] Del Vecchio D, Ninfa AJ, and Sontag ED. Modular cell biology: retroactivity and insulation. *Molecular Systems Biology*, 4(161), 2008.

[57] Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, and Collins JJ. Programmable cells: interfacing natural and engineered gene networks. *PNAS*, 101(22):8414–8419, 2004.

[58] Yokobayashi Y, Weiss R, and Arnold FH. Directed evolution of a genetic circuit. *PNAS*, 99(26):16587–16591, 2002.

[59] Cox III RS, Dunlop MJ, and Elowitz MB. A synthetic three-color scaffold for monitoring genetic regulation and noise. *Journal of Biological Engineering*, 4(10), 2010.

[60] Hajimorad M, Gray PR, and Keasling JD. A framework and model system to investigate linear system behavior in *Escherichia coli*. *Journal of Biological Engineering*, 5(3), 2011.

[61] Wang B, Kitney RI, Joly N, and Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2(508), 2011.

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

[63] Kim KH and Sauro HM. Fan-out in gene regulatory networks. *Journal of Biological Engineering*, 4(16), 2010.

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

[65] Young R. Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews*, 56(3):430–481, 1992.

[66] Wang IN, Smith DL, and Young R. HOLINS: the protein clocks of bacteriophage infections. *Annual Review of Microbiology*, 54:799–825, 2000.

[67] Grundling A, Manson MD, and Young R. Holins kill without warning. *PNAS*, 98(16):9348–9352, 2001.

[68] Calendar R. *The bacteriophages*. Oxford University Press, US, 2006.

[69] Morita M, Asami K, Tanji Y, and Unno H. Programmed *Escherichia coli* cell lysis by expression of cloned T4 phage lysis genes. *Biotechnology Progress*, 17(3):573–576, 2001.

[70] Mergulhao FJM, Summers DK, and Monteiro GA. Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, 23:177–202, 2005.

[71] Tanji Y, Asami K, Xing X, and Unno H. Controlled expression of lysis genes encoded in T4 phage for the gentle disruption of *Escherichia coli* cells. *Journal of Fermentation and Bioengineering*, 85(1):74–78, 1998.

[72] Yun J, Park J, Park N, Kang S, and Ryu S. Development of a novel vector system for programmed cell lysis in *Escherichia coli*. *Journal of Microbiology and Biotechnology*, 17(7):1162–1168, 2007.

[73] Cai Z, Xu W, Xue R, and Lin Z. Facile, reagentless and in situ release of *Escherichia coli* intracellular enzymes by heat-inducible autolytic vector for high-throughput screening. *Protein Engineering, Design and Selection*, 21(11):681–687, 2008.

[74] UC Berkeley iGEM Team 2008. Clonebots. [http://2008.igem.org/Team:UC_Berkeley].

[75] Ramanculov E and Young R. An ancient player unmasked: T4 rI encodes a *t*-specific antiholin. *Molecular Microbiology*, 41(3):575–583, 2001.

[76] Sleight SC, Bartley BA, Lieviant JA, and Sauro HM. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, 4(12), 2010.

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

[78] Pasotti L and Magni P. Circuiti sintetici I - Operatori logici. In *Biologia Sintetica*, pages 207–227. Patron Ed., Bologna, Italy, 2010.

[79] Morris Mano M and Kime CR. *Logic and computer design fundamentals*. Prentice Hall, Upper Saddle River, New Jersey, 2010.

[80] Stallings W. *Computer organization and architecture: designing for performance*. Prentice Hall, Upper Saddle River, New Jersey, 2003.

[81] Stevens AM and Greenberg EP. Quorum sensing in *Vibrio fischeri*: essential elements for activation of the luminescence genes. *Journal of Bacteriology*, 179(2):557–562, 1997.

[82] Pesci EC, Pearson JC, Seed PC, and Iglewski BH. Regulation of las and rhl quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 179(10):3127–3132, 1997.

[83] Fussenegger M. Synthetic biology: synchronized bacterial clocks. *Nature*, 463:301–302, 2010.

[84] Brenner K, Karig DK, Weiss R, and Arnold FH. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *PNAS*, 104(44):17300–17304, 2007.

[85] Albus AM, Pesci EC, Runyen-Janecky LJ, West SE, and Iglewski BH. Vfr controls quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 179(12):3928–3935, 1997.

[86] Balagaddè FK, Song H, Ozaki J, Collins CH, Barnet M, Arnold FH, Quake SR, and You L. A synthetic *Escherichia coli* predator-prey ecosystem. *Molecular Systems Biology*, 4(187), 2008.

[87] Ptashne M. *A genetic switch: phage lambda revisited*. Cold Spring Harbor Laboratory Press, 2004.

[88] Siso MIG. The biotechnological utilization of cheese whey: a review. *Bioresource Technology*, 57(1):1–11, 1996.

[89] Bansal S, Oberoi HS, Dhillon GS, and Patil RT. Production of β-galactosidase by *Kluyveromyces marxianus* MTCC 1388 using whey and effect of four different methods of enzyme extraction on β-galactosidase activity. *Indian Journal of Microbiology*, 48(3):337–341, 2008.

[90] Agenzia per l'innovazione e l'internazionalizzazione delle imprese (AINT). Tecnologie disponibili per il trattamento di siero di latte e reflui dell'industria lattiero casearia e localizzazione degli impianti in Puglia. Technical report, 2007.

[91] Banchelli L. Recupero dei costituenti organici del siero del latte e bioconversione in proteine unicellulari da lievito alimentare. Technical report, OGM Food and Biopollution certification, 2002.

[92] Kosikowski FV. Whey utilization and whey products. *Journal of Dairy Science*, 62:1149–1160, 1979.

[93] ISTAT. [http://www.istat.it/agricoltura/datiagri/latte/elelat.html], 2011.

[94] OECD-FAO. Agricultural Outlook 2008-2017 Highlights. [http://www.agri-outlook.org/dataoecd/54/15/40715381.pdf], 2008.

[95] Smithers GW. Whey and whey proteins - from 'gutter-to-gold'. *International Dairy Journal*, 18:695–704, 2008.

[96] Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, and Keasling JD. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, 463:559–563, 2010.

[97] Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, and Ingram LO. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. *Journal of Biomedicine and Biotechnology*, 2010(761042), 2010.

[98] Guimaraes PMR, Teixeira JA, and Domingues L. Fermentation of lactose to bioethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnology Advances*, 28:375–384, 2010.

[99] Asraf SS and Gunasekaran P. Current trends of β-galactosidase research and application. In *Current Rearch, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*. A. Mendez-Vilas Ed., 2010.

[100] de Glutz FN. *Fuel bioethanol production from whey permeate*. PhD thesis, Ecole Polytechnique Federale de Lausanne, 2009.

[101] Yanase H, Kotani T, Yasuda M, Matsuzawa A, and Tonomura K. Metabolism of galactose in *Zymomonas mobilis*. *Applied Microbiology and Biotechnology*, 35(3):364–368, 1991.

[102] Leite AR, Guimaraes WV, de Araujo EF, and Silva DO. Fermentation of sweet whey by recombinant *Escherichia coli* KO11. *Brazilian Journal of Microbiology*, 31:212–215, 2000.

[103] Dien BS, Cotta MA, and Jeffries TW. Bacteria engineered for fuel ethanol production: current status. *Applied Microbiology and Biotechnology*, 63:258–266, 2003.

[104] Guimaraes WV, Dudey GL, and Ingram LO. Fermentation of sweet whey by ethanologenic *Escherichia coli*. *Biotechnology and Bioengineering*, 40:41–45, 1992.

[105] Ohta K, Beall DS, Mejia JP, Shanmugam KT, and Ingram LO. Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding Pyruvate Decarboxylase and Alcohol Dehydrogenase II. *Applied and Environmental Microbiology*, 57(4):893–900, 1991.

[106] Lawford HG and Rousseau JD. Loss of ethanologenicity in *Escherichia coli* B recombinants pLOI297 and KO11 during growth in the absence of antibiotics. *Biotechnology Letters*, 17(7):751–756, 1995.

[107] Guimaraes PMR, Francois J, Parrou JL, Teixeira JA, and Domingues L. Adaptive evolution of a lactose-consuming *Saccharomyces cerevisiae* recombinant. *Applied and Environmental Microbiology*, 74(6):1748–1756, 2008.

[108] Martinez A, York SW, Yomano LP, Pineda VL, Davis FC, Shelton JC, and Ingram LO. Biosynthetic burden and plasmid burden limit expression of chromosomally integrated heterologous genes (pdc, adhB) in *Escherichia coli*. *Biotechnology Progress*, 15(5):891–897, 1999.

[109] Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, and Collins JJ. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7):841–847, 2004.

[110] Khlebnikov A, Risa O, Skaug T, Carrier TA, and Keasling JD. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *Journal of Bacteriology*, 182(24):7029–7034, 2000.

[111] Ingram LO, Conway T, Clark DP, Sewell GW, and Preston JF. Genetic engineering of ethanol production in *Escherichia coli*. *Applied and Environmental Microbiology*, 53(10):2420–2425, 1987.

[112] Martinez-Morales F, Borges AC, Martinez A, Shanmugam KT, and Ingram LO. 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.

[113] Posfai G, Koob MD, Kirkpatrick HA, and Blattner FR. 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.

[114] Datta S, Costantino N, and Court DL. A set of recombineering plasmids for gram-negative bacteria. *Gene*, 379:109–115, 2006.

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

[116] Ellis HM, Yu D, DiTizio T, and Court DL. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *PNAS*, 98(12):6742–6746, 2001.

[117] Haldimann A and Wanner B. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *Journal of Bacteriology*, 183(21):6386–6393, 2001.

[118] Cherepanov PP and Wackernagel W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, 158(1):9–14, 1995.

[119] Kalscheuer R, Stolting T, and Steinbuchel A. Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology*, 152:2529–2536, 2006.

[120] Chen J, Zhang W, Tan L, Wang Y, and He G. Optimization of metabolic pathways for bioconversion of lignocellulose to ethanol through genetic engineering. *Biotechnology Advances*, 27, 2009.

[121] Sezonov G, Joseleau-Petit D, and D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology*, 189(23):8746–8749, 2007.

[122] Alterthum F and Ingram LO. Efficient ethanol production from glucose, lactose, and xylose by recombinant *Escherichia coli*. *Applied and Environmental Microbiology*, 55(8):1943–1948, 1989.

[123] Neale AD, Scopes RK, Kelly JM, and Wettenhall RE. The two alcohol dehydrogenases of *Zymomonas mobilis*. Purification by differential dye ligand chromatography, molecular characterisation and physiological roles. *European Journal of Biochemistry*, 154(1):119–124, 1986.

[124] Lynd LR, Weimer PJ, van Zyl WH, and Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 66(3):506–577, 2002.

[125] Asghari A, Bothast RJ, Doran JB, and Ingram LO. Ethanol production from hemicellulose hydrolysates of agricultural residues using genetically engineered *Escherichia coli* strain KO11. *Journal of Industrial Microbiology*, 16:42–47, 1996.

[126] Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Lee TS, Tullman-Ercek D, Voigt CA, Simmons BA, and Keasling JD. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *PNAS*, 108(50):19949–19954, 2011.

[127] Cho KM and Yoo YJ. Novel SSF process for ethanol production from microcrystalline cellulose using the delta-integrated recombinant yeast, *Saccharomyces cerevisiae* L2612 delta GC. *Journal of Microbiology and Biotechnology*, 9:340–345, 1999.

[128] Moniruzzaman M, Lai X, York SW, and Ingram LO. Isolation and molecular characterization of high-performance cellobiose-fermenting spontaneous mutants of ethanologenic *Escherichia coli* KO11 containing the *Klebsiella oxytoca* casAB operon. *Applied and Environmental Microbiology*, 63(12):4633–4637, 1997.

[129] Wood BE and Ingram LO. Ethanol production from cellobiose, amorphous cellulose, and crystalline cellulose by recombinant *Klebsiella oxytoca* containing chromosomally integrated *Zymomonas mobilis* genes for ethanol production and plasmids expressing

thermostable cellulase genes from *Clostridium thermocellum*. *Applied and Environmental Microbiology*, 58(7):2103–2110, 1992.

[130] Guo Z, Arfman N, Ong E, Gilkes NR, Kilburn DG, Warren RAJ, and Miller Jr RC. Leakage of *Cellulomonas fimi* cellulases from *Escherichia coli*. *FEMS Microbiology Letters*, 49:279–283, 1988.

[131] Filloux A. The underlying mechanisms of type II protein secretion. *Biochimica et Biophysica Acta*, 1694(1-3):163–179, 2004.

[132] Winkler J, Eltis LD, Dwyer DF, and Rohde M. Tetrameric structure and cellular location of catechol 2,3-dioxygenase. *Archives of Microbiology*, 163(1):65–69, 1995.

[133] Anderson JC. J231xx promoter collection. [http://partsregistry.org/Part:BBa_J23100].

[134] Sambrook J and Russell DW. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, 3rd edition, 2001.

[135] Kumar JK, Tabor S, and Richardson CC. Proteomic analysis of thioredoxin-targeted proteins in *Escherichia coli*. *PNAS*, 101(11):3759–3764, 2004.

[136] Braun D, Basu S, and Weiss R. Parameter estimation for two synthetic gene networks: a case study. In *ICASSP 5*, pages 769–772. IEEE, 2005.

[137] Goryachev AB, Toh DJ, and Lee T. Systems analysis of a quorum sensing network: design constraints imposed by the functional requirements, network topology and kinetic constants. *BioSystems*, 83:178–187, 2006.

[138] Basu S, Gerchman Y, Collins CH, Arnold FH, and Weiss R. A synthetic multicellular system for programmed pattern formation. *Nature*, 434:1130–1134, 2005.

[139] Urbanowski ML, Lostroh CP, and Greenberg EP. Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein. *Journal of Bacteriology*, 186:631–637, 2004.

[140] You L, Cox III RS, Weiss R, and Arnold FH. Programmed population control by cell-cell communication and regulated killing. *Nature*, 428:868–871, 2004.

[141] Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, and Molin S. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied Environmental Microbiology*, 64(6):2240–2246, 1998.

[142] Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, and Shao Y. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462, 1997.

[143] Chung CT, Niemela SL, and Miller RH. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *PNAS*, 86:2172–2175, 1989.

[144] Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, and Court DL. An efficient recombination system for chromosome engineering in *Escherichia coli*. *PNAS*, 97(11):5978–5983, 2000.