ALMA MATER STUDIOURUM- UNIVERSITÀ DI BOLOGNA
SEDE DI CESENA
SECONDA FACOLTÀ DI INGEGNERIA CON SEDE A CESENA
CORSO DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

TITOLO DELLA TESI

Multicelled Behaviour in E. coli: Design and experimental characterization of an engineered library of hybrid promoters

Tesi in
Bioingegneria Molecolare e Cellulare LM

Relatore:
Prof. Ing. Stefano Severi

Correlatori:
Prof. Emanuele Domenico Giordano
Prof. Ing. Eric Klavins
Ing. Seunghee Shelly Jang

Presentata da:
Marilisa Cortesi

Sessione III
Anno Accademico 2011/2012

Contents

1 Introduction and Background 15
  1.1 Synthetic Biology 15
  1.2 Symmetry Breaking 17
  1.3 Molecular Biology 18
    1.3.1 DNA 18
    1.3.2 Protein Biosynthesis 20
  1.4 Natural regulatory systems 25
    1.4.1 E. coli lactose digesting system 25
    1.4.2 Luciferase enzyme complex 27

2 Leader Election project 31
  2.1 Circuit Design 31
    2.1.1 Leader Election problem 31
    2.1.2 The device 32
  2.2 Follower Module: a closer look 34
    2.2.1 The hybrid Lux-Lac promoter 35
    2.2.2 The Shine Dalgarno (RBS) sequence 40

3 Experimental Characterization 43
  3.1 Assembly reaction 43
  3.2 Single module Characterization 45
    3.2.1 Desired Behaviour 46
    3.2.2 Saturating concentrations 49
    3.2.3 Dose Response 52
    3.2.4 Fitting 53
  3.3 Interaction with the other components 57
    3.3.1 Supernatant assay 58
    3.3.2 Coculture assay: Saturating concentrations 60
    3.3.3 Coculture assay: Dose response 63

4 Materials and Methods 67
4.1 PCR, Polymerase Chain Reaction 67
4.2 DNA Electrophoresis 70
4.3 Gibson Assembly 71
4.4 Electroporation 71
4.5 Induction assay 72
4.6 Supernatant assay 73
4.7 Coculture assay 74

5 Conclusions 75
5.1 From sequence to gene expression 75
5.2 Regulated Promoters Engineering 77
5.3 Complex circuits realization 79

List of Figures

1.1 Nucleotide 19
1.2 Double helix DNA 20
1.3 Semiconservative Replication 21
1.4 Comparison between a strand of RNA and a strand of DNA . 22
1.5 Structure of an amino acid 23
1.6 Conversion table between codons and amino acids 24
1.7 Structure of a tRNA 25
1.8 Schematic structure of the Central Dogma of molecular biology 26
1.9 Structure of E.coli's lactose digesting system 27
1.10 Marine animals that maintain a symbiotic relation with *Vibrio Fischeri* 28
1.11 Luciferase enzyme complex 28
1.12 LuxI and LuxR communication system 29

2.1 Leader Election Circuit [13] 32
2.2 Leader Cells 34
2.3 Follower Cells 34
2.4 Follower Module 35
2.5 Promoters' architecture 35
2.6 A81 promoter 37
2.7 B72 promoter 38
2.8 Unregulated activity 39
2.9 Structure of the RBS region 40
2.10 Relation between the length of the spacer and the fluorescence level 41

3.1 Follower plasmid 44
3.2 Gibson assembly reaction 45
3.3 Oligo assembly structure 46
3.4 Flow cytometer 47
3.5 Reduced circuit used to characterize the Follower module . . . 47

3.6 AND gate 48
3.7 representation of the two indexes l and a for logic gates 49
3.8 RFP expression of the Follower module when induced with saturating concentrations of inducers 51
3.9 representation of the two indexes l and a for the reduced circuits 52
3.10 Dose response curve for the prototype promoter 53
3.11 Dose response curve for the first promoter 54
3.12 Dose response curve for the third promoter 55
3.13 Effect of the Hill coefficient on the shape of the curve 56
3.14 Fitting of the data of the system with the prototype promoter 57
3.15 Fitting of the data of the system with the first promoter . . . 58
3.16 Fitting of the data of the system with the third promoter . . . 59
3.17 Sender Cells 60
3.18 Results of the supernatant assay 61
3.19 Distribution of the GFP levels for the coculture assay 62
3.20 Results of the Coculture assay 62
3.21 Coculture assay: dose response promoter 1 64
3.22 Coculture assay: dose response promoter 3 65
3.23 Coculture assay: dose response prototype promoter 66

4.1 PCR reaction 68
4.2 Thermocycler 69
4.3 Scheme of the phases of the PCR reaction 69
4.4 Example of Electrophoresis in Agarose gel 70
4.5 Cuvettes for Electroporation 71

5.1 A transcription factor can bind two operators sites at the same time 78

“An experiment is a question
which science poses to Nature,
and a measurement is the reading
of Nature’s answer”.

Max Planck,

Scientific Autobiography and other papers (1949)

Sommario

La biologia sintetica ha recentemente avuto un enorme sviluppo, il numero di pubblicazioni in questo settore é aumentato esponenzialmente e sono state presentate applicazioni che spaziano dalla produzione di farmaci a quella di substrati bioenergetici fino alla catalisi industriale. Tuttavia la maggior parte delle applicazioni sono piuttosto semplici e non realizzano appieno le potenzialitá di questa disciplina.

Molte cause concorrono a limitare la complessitá dei sistemi fino ad ora implementati, come la ridotta caratterizzazione dei componenti o la notevole variabilitá dei sistemi biologici, ma uno dei piú importanti é l'incapacitá intrinseca della cellula di sopportare il carico metabolico introdotto da un circuito genetico sintetico di complessitá elevata.

L'obbiettivo del progetto, all'interno del quale questo lavoro si inserisce, é tentare di risolvere questo problema ingegnerizzando un comportamento multicellulare in cellule procariotiche; questo sistema porterebbe allo sviluppo di un comportamento cooperativo, che permetterebbe ad un insieme di batteri di realizzare funzioni complesse, non implementabili da un singolo individuo.

In particolare lo scopo é ottenere l'elezione di un Leader (Leader Election), questa procedura nasce nell'ambito della computazione distribuita e permette di individuare un singolo organismo come organizzatore e coordinatore di una serie di compiti assegnati all'intera popolazione, semplificando notevolmente la computazione. Oltre a implementare una funzionalitá basilare, utile in ogni applicazione che sfrutta la cooperazione tra procarioti, questo sistema potrebbe anche essere utilizzato per studi evoluzionistici che si propongono di indagare come gli organismi pluricellulari si sono evoluti a partire da sistemi unicellulari.

Il lavoro qui presentato si riferisce alla progettazione e alla caratterizzazione sperimentale, di un componente del circuito genetico che risolve il problema del Leader Election. Questo modulo, formato da un promotore ibrido e da un gene, si attiva nelle cellule non-leader quando queste ricevono il segnale che un leader é emerso nella colonia.

L'accento, in particolare, é stato posto sul promotore ibrido, ne sono state realizzate differenti versioni, applicando regole euristiche riportate in letteratura [22], e la loro attivitá é stata valutata sperimentalmente.

L'obbiettivo della caratterizzazione sperimentale é stato stimare la risposta del circuito genetico all'introduzione, nell'ambiente cellulare, di particolari molecole, dette induttori, che possono essere considerate ingressi per il sistema in esame.

Il comportamento desiderato é assimilabile a quello di un AND logico, nel quale l'uscita, rappresentata in questo caso dalla intensitá del segnale lumi-

noso prodotto da una proteina fluorescente, ha valore uno solo in presenza di entrambi gli induttori. La robustezza e la stabilitá di questo comportamento sono state valutate modificando la concentrazione dei segnali di ingresso e costruendo curve di risposta. Da questi dati é stato possibile concludere che, i promotori ibridi considerati presentano una risposta simile a quella di un AND logico per un ampio range di concentrazioni degli induttori, anche se sono presenti notevoli differenze nel profilo di espressione dei vari costrutti; questo é dovuto al fatto che sia i segnali di ingresso che quello in uscita sono continui, e che quindi la similitudine con il comportamento di un circuito logico ha limitazioni evidenti.
Il modulo del circuito considerato in questa analisi ha un ruolo fondamentale nella realizzazione del sistema di comunicazione intercellulare, necessario per la realizzazione della cooperazione. Per questo motivo la seconda parte della caratterizzazione é stata incentrata sull'analisi del fenomeno di invio e ricezione del segnale; in particolare é stata testata l'interazione del modulo considerato fino a questo momento, e la parte del circuito resposabile dell'emissione del segnale chimico utilizzato come vettore dell'informazione.
Il comportamento desiderato rimane assimilabile a quello di un AND logico, dato che il segnale di uscita viene, anche in questo caso, determinato dall'attivitá dal promotore ibrido. I risultati sperimentali hanno dimostrato che il sistema presenta il comportamento corretto, pur mantenendo una variabilitá significativa tra i diversi costrutti; la costruzione delle curve di risposta, inoltre, ha evidenziato la presenza di vincoli piú stringenti sulla concentrazione degli induttori, necessaria per ottenere una netta separazione dei due livelli di espressione.
Nel capitolo conclusivo vengono poi analizzate le sequenze di DNA dei promotori ibridi, tentando di individuare quali elementi regolatori hanno un peso maggiore nella determinazione della forza espressiva dei costrutti analizzati, senza peró riuscire ad ottenere risultati definitivi. Infine vengono presentate alcune riflessioni generali sull'ingegnerizzazione di promotori e sulla costruzione di circuiti genetici complessi che mirano a ripredere brevemente i problemi descritti nell'introduzione e delineare possibili soluzioni.

- *Capitolo 1:* breve panoramica delle principali caratteristiche e applicazioni della biologia sintetica e introduzione dei concetti base della biologia molecolare. In questa sezione vengono anche descritti alcuni sistemi di regolazione genica naturali e viene introdotto il concetto di rottura della simmetria, entrambi questi elementi saranno utili alla comprensione dei capitoli successivi.

- *Capitolo 2:* qui viene descritto il problema dell'elezione di un leader, sia dal punto di vista computazionale che biologico e viene analizzato dettagliatamente il funzionamento del circuito scelto per implementare questa funzionalitá in E. coli. Infine viene riportata la progettazione della libreria di promotori ibridi, che verrá successivamente caratterizzata.

- *Capitolo 3:* realizzazione fisica e caratterizzazione della libreria di promotori ibridi. La caratterizzazione é stata suddivisa in due fasi che analizzano, rispettivamente, il comportamento in isolamento e in presenza di un altro componente del circuito. In quest'ultimo caso, in particolare, si é valutata l'interazione tra i due sistemi, implementando un sistema di comunicazione intercellulare.

- *Capitolo 4:* descrizione dettagliata delle tecniche e dei protocolli utilizzati per svolgere il progetto.

- *Capitolo 5:* riflessioni conclusive sull'analisi di sequenze di DNA per prevedere l'espressione genica, l'ingegnerizzazion di promotori e la costruzione di circuiti genetici sintetici complessi.

Abstract

Synthetic biology has recently had a great development, many papers have been published and many applications have been presented, spanning from the production of biopharmacheuticals to the synthesis of bioenergetic substrates or industrial catalysts. But, despite these advances, most of the applications are quite simple and don't fully exploit the potential of this discipline.

This limitation in complexity has many causes, like the incomplete characterization of some components, or the intrinsic variability of the biological systems, but one of the most important reasons is the incapability of the cell to sustain the additional metabolic burden introduced by a complex circuit. The objective of the project, of which this work is part, is trying to solve this problem through the engineering of a multicellular behaviour in prokaryotic cells. This system will introduce a cooperative behaviour that will allow to implement complex functionalities, that can't be obtained with a single cell. In particular the goal is to implement the Leader Election, this procedure has been firstly devised in the field of distributed computing, to identify the process that allow to identify a single process as organizer and coordinator of a series of tasks assigned to the whole population. The election of the Leader greatly simplifies the computation providing a centralized control. Furthermore this system may even be useful to evolutionary studies that aims to explain how complex organisms evolved from unicellular systems.

The work presented here describes, in particular, the design and the experimental characterization of a component of the circuit that solves the Leader Election problem. This module, composed of an hybrid promoter and a gene, is activated in the non-leader cells after receiving the signal that a leader is present in the colony.

The most important element, in this case, is the hybrid promoter, it has been realized in different versions, applying the heuristic rules stated in [22], and their activity has been experimentally tested. The objective of the experimental characterization was to test the response of the genetic circuit to the introduction, in the cellular environment, of particular molecules, inducers, that can be considered inputs of the system.

The desired behaviour is similar to the one of a logic AND gate in which the exit, represented by the luminous signal produced by a fluorescent protein, is one only in presence of both inducers. The robustness and the stability of this behaviour have been tested by changing the concentration of the input signals and building dose response curves. From these data it is possible to conclude that the analysed constructs have an AND-like behaviour over a wide range of inducers' concentrations, even if it is possible to identify many

differences in the expression profiles of the different constructs. This variability accounts for the fact that the input and the output signals are continuous, and so their binary representation isn't able to capture the complexity of the behaviour.

The module of the circuit that has been considered in this analysis has a fundamental role in the realization of the intercellular communication system that is necessary for the cooperative behaviour to take place. For this reason, the second phase of the characterization has been focused on the analysis of the signal transmission. In particular, the interaction between this element and the one that is responsible for emitting the chemical signal has been tested.

The desired behaviour is still similar to a logic AND, since, even in this case, the exit signal is determined by the hybrid promoter activity. The experimental results have demonstrated that the systems behave correctly, even if there is still a substantial variability between them. The dose response curves highlighted that stricter constrains on the inducers concentrations need to be imposed in order to obtain a clear separation between the two levels of expression.

In the conclusive chapter the DNA sequences of the hybrid promoters are analysed, trying to identify the regulatory elements that are most important for the determination of the gene expression. Given the available data it wasn't possible to draw definitive conclusions. In the end, few considerations on promoter engineering and complex circuits realization are presented. This section aims to briefly recall some of the problems outlined in the introduction and provide a few possible solutions.

Chapter 1

Introduction and Background

1.1 Synthetic Biology

> "Synthetic Biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems which display functions that do not exist in nature. This engineering perspective may be applied at all levels of the hierarchy of biological structures, from individual molecules to whole cells, tissues and organisms. In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way"
>
> High-level Expert Group European Commission

As reported in [4], engineered biological systems are being used to address a wide variety of society's needs. Examples includes the production of insulin and more than two hundreds others biopharmaceuticals, countless natural products, industrial catalysis, and bioenergy substrates such as sugars, ethanol, methane and hydrogen, in microbes, eukaryotic cells and higher organisms; and more exotic successes such as the use of cytokine expressing *Mycobacterium bovis* BCG as an effective treatment for certain forms of bladder cancer. But, even if Synthetic Biology could provide a toolbox of reusable genetic components [14] and many fields could benefit from its application, most of the state of art circuits are still quite simple and don't fully exploit the potential of Synthetic Biology [27].

As highlighted by different sources [14], [10], this limitation in complexity has many causes, first of all, many parts are undefined, the construction and characterization of synthetic systems can be tedious and unreliable, for this reason many of them have not been fully characterized. Their activity has not been tested and it is not clear how the experimental conditions influence the behaviour. Furthermore the circuitry is unpredictable so, even if the function of each part is known, they may not work as expected when put together and the exogenous DNA might even have unintended effects on the host.

Another fundamental feature of the biological systems is the ability to grow and evolve, but it has depended on those behaviours that frustrate predictability and might compromise the functioning of the synthetic circuits [4]. In addition, the molecular events inside the cells are prone to random fluctuations or noise and synthetic systems need to work regardless.

In order to overcome these difficulties some of the key elements of classical engineering: standardization, decoupling and abstraction are of prime importance [9].

As for the first factor, the biological engineering community has yet to develop formal, widely used, standards for most classes of basic biological functions, experimental measurements and system operations [10]. The iGem competition and the BioBrick Foundation have concerted many efforts in this direction, promoting generalized assembly and measuring techniques for genetic circuits.

Decoupling is the idea that it is useful to separate a complicated task into many simpler problems, that can be solved independently, such that the resulting work can eventually be combined to produce a functioning whole [10]. This concept can be incredibly powerful for the realization of complex genetic circuits, since it makes possible the realization of an elaborate behaviour through the composition of basic design units, that represent biological functions [9].

Abstraction is another fundamental concept to manage complexity, there are two main forms of abstraction considered in biological engineering [10]:

- The information describing biological functions might be organized across levels of abstraction using hierarchies. To be useful, abstraction hierarchies must allow individuals to work at any one level of complexity without regard for the details that define other levels, yet allow for the principle of exchange of limited information across levels.
- The devices that compose engineered biological systems should be designed and built so that they are simple to model and easily used in combination.

But, even if independently operating cells can perform tasks of varying complexity, more sophisticated coordinated behaviours can only be obtained with populations of communicating cells [9]. Two of the most prominent applications of this concept are the synchronized genetic clocks described in [26] and the pattern formation system reported in [23].

The emerging of this new field has opened a debate in the public opinion since, as reported in [28]: Synthetic Biology puts heavy pressure on many of the culturally entrenched distinctions and demarcations that are constitutive of our symbolic order. It shifts or blurs the boundaries between matter and information, life and non-life, nature and artefact, organic and inorganic, creator and creature, the evolved and the designed.

But the future promised by this discipline is full of incredible applications and, to use the words of Purnick and Weiss [19], Synthetic Biology has the potential to fabricate practical organisms that could clean hazardous waste in inaccessible places, to use plants to sense chemicals and respond accordingly, to produce clean fuel in an efficient and sustainable fashion, or to recognize and destroy tumours. [...] Our designs can potentially be more robust or efficient than systems that have been fashioned by evolution.

1.2 Symmetry Breaking

The word symmetry derives from the Greek word $\sigma\upsilon\mu\mu\epsilon\tau\rho\epsilon\iota\nu$ that literally means "measure together". In the course of history the concept of symmetry has been applied to a wide variety of fields ranging from mathematics to religion, from physics to geometry from literature to science and music. It generally convey a sense of harmonious or aesthetically pleasing proportionality and balance, such that it reflects beauty and perfection.

In physics symmetry means uniformity or invariance or the "existence of different viewpoints from which the system appears the same". Symmetry breaking is the process by which such uniformity is broken, or the number of points to the new invariance is reduced, to generate a more structured and improbable state [20]. This is a very important procedure because it accounts for the transition of a physical system to a different state and it plays a major role in pattern formation.

Symmetry breaking [...] is even a prevalent process in biology, because organismal survival depends critically on well defined structures and patterns at both microscopic and macroscopic level [20]. The main example of symmetry breaking in biology is morphogenesis, the process that causes an organism to develop its shape and in particular controls the organized spatial distribution of cells during the embryonic development.

As David Peat explains in his talk Meaning and Structure in Biology and Physics: Some Outstanding Questions [1], an increase in symmetry is strictly associated with a loss of genetic information. At first sight this is a curious relation between two different concepts, symmetry and information, but it implies that the more information an organism express or unfold, the more will its initial symmetry be broken and the more complex its structure will be. Considering the genome as the main information conveyor of a life form this principle creates a correlation between the size of the genetic material and the complexity of the system. So it is not surprising that important insights into symmetry breaking have come from studies of cell division, especially in the context of asymmetric division that lead to different cell fates. They state, in particular, that the origin of large scale asymmetry in biology often lies in asymmetries at lower scale and that of paramount importance are the physical and biochemical amplification systems that can break symmetry by turning local and transient signals or noise into stable and system level asymmetries [20].

The Leader Election problem, that will be described at the beginning of the following chapter, is a very clear and simple example of symmetry breaking. Starting from a genetically uniform group of elements, this procedure leads to the the separation of one component, now called leader, from the main population. In many applications the election of the leader induce a well defined reaction in the other elements.

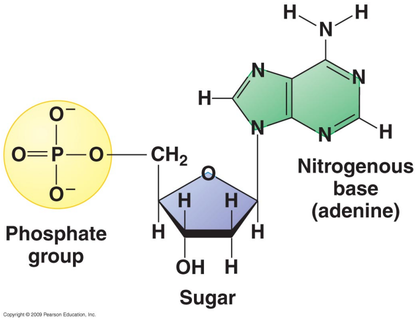
1.3 Molecular Biology

Synthetic biology employs the techniques developed in the molecular biology's field to implement the changes in the genetic material of the host cell necessary to realize the desired functions. In [6] molecular biology is defined as: "not so much a technique as an approach, an approach from the viewpoint of the so-called basic sciences with the leading idea of searching below the large-scale manifestations of classical biology for the corresponding molecular plan. It is concerned particularly with the forms of biological molecules and [...] is predominantly three dimensional and structural, which does not mean, however, that it is merely a refinement of morphology. It must, at the same time, inquire into genesis and function".

1.3.1 DNA

The DNA (Deoxyribonucleic acid) is the molecule in which is stored the information necessary to protein biosynthesis and the regulation of almost every cellular event. Its structure is modular and composed of a large number of elementary blocks, nucleotides, that in turn, are formed of a pentose sugar with a nitrogenous base covalently bound in position 1' and at least a phosphoric acid esterified in position 5' (Figure 1.1).

It was at the beginning of the 1950s that the structure and the function



Copyright © 2009 Pearson Education, Inc.

Figure 1.1: Nucleotide

of the DNA were unveiled allowing to define at molecular level its biological features:

- It codes the genetic information of a cell and it is able to pass it to the next generation.
- It directs and controls the entire cell cycle and the information that it contains.
- It can duplicate with very high precision so that, at every cell division, a complete and identical copy of the information is passed to every daughter cell.
- It can modify its structure to alter one or more of the codified instruction or make new ones appear.
- Using a specific system of translation of the message it makes the information available to the cell.

The DNA has a double helix structure in which two antiparallel strands fold around a common axis. The nitrogenous bases of the two filaments are specifically coupled, an adenine with a thymine and a cytosine with a guanine, and they keep together the two strands with multiple hydrogen bonds.

As shown in Figure 1.2 the structure of the DNA is similar to a winding staircase in which the bases are the steps and the frame is composed of an alternation of pentose sugar molecules and phosphate acid.

Given the complementarity between the two strands each one is a molecular mould for the other, this is very important in the duplication of the cell where the semi-conservative replication takes place. This process begins

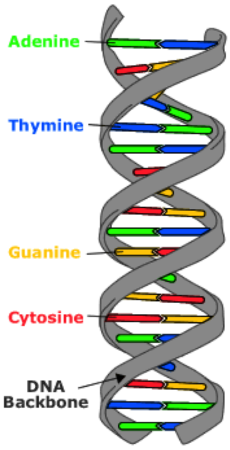


Figure 1.2: Double helix DNA

when the two filaments of the double strand are divided, this is possible because the hydrogen bonds are weak and reversible. At this point a molecular machinery, named DNA-polymerase, binds to one of the two filaments and creates the complementary strand from a pool of nucleotides available in the cytoplasm. This process is repeated on both the filaments leading to the construction of two copies of the entire genome, as represented in Figure 1.3.

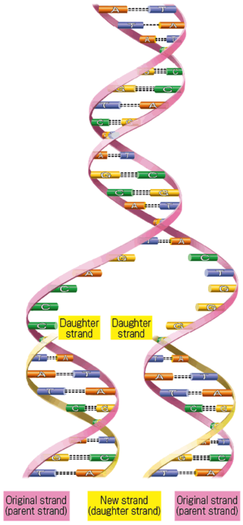
1.3.2 Protein Biosynthesis

Protein biosynthesis is one of the main and most complex activities of the cell and during its evolution it has acquired many features:

- It is very precise, because a change in the amino acids' sequence might cause the production of inactive or unstable molecules.
- It makes possible a chemical process that doesn't spontaneously happen.
- It has evolved a correspondence between nucleotides and amino acids.

The protein biosynthesis process is described in the **central dogma of molecular biology** by Francis H. Crick (Figure 1.8), it states that genetic information goes from DNA to RNA and then to proteins. The central dogma defines three main processes that preserve and pass on the genetic information:

- **Replication** is the production of new DNA identical to the old one.


Daughter strand
Daughter strand
Original strand (parent strand)
New strand (daughter strand)
Original strand (parent strand)

© CSLS/The University of Tokyo

Figure 1.3: Semiconservative Replication

- **Transcription** is the procedure by which the information is transmitted to RNA from DNA.
- **Translation** allows the genetic information to pass from RNA to proteins.

The first one is the Semi-conservative Replication earlier described, and allows the genetic information to pass to the daughter cell during cell division. The process of transcription, on the other hand, is realized by a protein complex named RNA-polymerase and copies the information conveyed by a particular gene on a fragment of RNA. The RNA molecule is very similar to the DNA (Figure 1.4) but its structure uses, as pentose sugar, the ribose instead of the deoxyribose and the thymine is substituted with uracil. Another important difference is that the RNA is always single-stranded and, beside information carrier, it has even other functions in the protein biosynthesis process.

The synthesis of the RNA molecule normally begins only in correspondence of specific sequences on the DNA template named promoters. A promoter is a sequence of about 40 bp that precedes the point at which transcription begins and that is recognized by the RNA-polymerase. In particular there are two regions, about 10 and 35 nucleotides before the point at which transcription begins, that allow the molecular machinery to bind the DNA and begin the transcription. Punctual mutations in these sequences modify

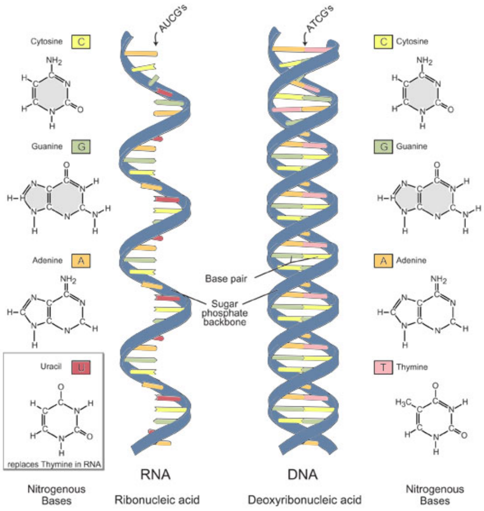

Cytosine C
Guanine G
Adenine A
Uracil U
replaces Thymine in RNA
AUCG's
ATCG's
Base pair
Sugar phosphate backbone
C Cytosine
G Guanine
A Adenine
T Thymine
Nitrogenous Bases
RNA
Ribonucleic acid
DNA
Deoxyribonucleic acid
Nitrogenous Bases

Image adapted from: National Human Genome Research Institute.
Talking Glossary of Genetic Terms. Available at: www.genome.gov/
Pages/Hyperion//DIR/VIP/Glossary/Illustration/rna.shtml.

Figure 1.4: Comparison between a strand of RNA and a strand of DNA

the affinity of the RNA-polymerase for the promoter and this technique is very commonly used to modulate the amount of RNA produced and, indirectly, the protein synthesis.

Another way to control protein biosynthesis is through the use of regulated promoters, these promoters contain, beside the -10 and -35 boxes, other regulatory sequences that tie the transcription initiation to the presence, or absence, of specific molecules in the environment. These molecules, named transcription factors, bind to the DNA, in correspondence of additional regulatory sites and modify the probability of production of the RNA, acting on the affinity between the RNA-polymerase and the DNA. This regulation is very specific, and strictly dependent on the transcription factor employed, but it is even more flexible than the previous method, since its dynamic range is much wider and its activity more predictable: a transcription factor might be either an activator or a repressor, while the effect of a punctual mutation is much more complex to predict.

The transcription then proceeds copying the DNA template with a speed of about 20-50 nt/s. In order for this process to end and the RNA to be released, the RNA polymerase needs a stop signal. There are many kind of terminators but the most common, especially in bacteria, is a DNA sequence rich in guanines and cytosines followed by a region with predominance of adenines and thymines. This pattern makes the connection between the

DNA and the RNA-polymerase unstable and this leads to the interruption of the transcription and the release of the RNA.
Once the RNA is in the cytoplasm the translation can begin, it is performed by another molecular machinery, composed of proteins and ribosomal RNA (rRNA). The name of this complex is ribosome and it realizes the conversion between nucleotides and amino acids. The amino acids are organic molecules made from amine, carboxylic acid and a side chain specific to each amino acid, represented with R in Figure 1.5.

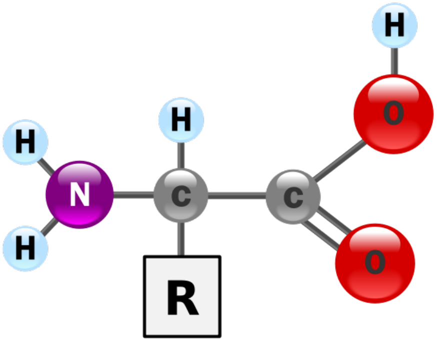
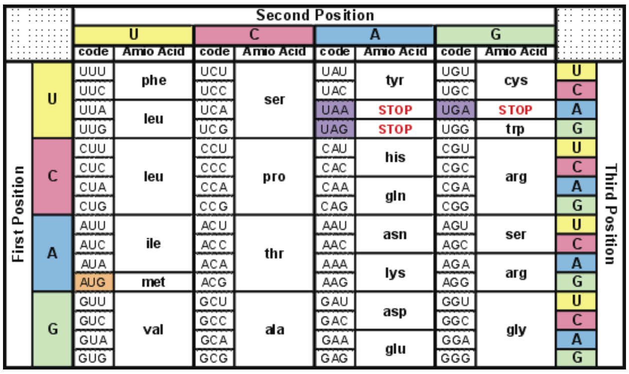


Figure 1.5: Structure of an amino acid

As the RNA-polymerase, the ribosome recognize and binds a part of the RNA sequence, the ribosome binding site or RBS, this is very important in order for the ribosome to interpret correctly the information coded into the RNA.
Only twenty of the almost five hundred known amino acids are used to build proteins and so, in order to have a unique correspondence with the nucleotides, it is necessary to consider a ratio 1:3 between amino acids and nucleotides. But this encoding isn't univocal since, considering four nucleotides, there are 64 possible triplets. One of these triplets, AUG, codes for methionine but is even used to signal the translation initiation. Three other triplets, UAA, UAG and UGA, aren't associated to any amino acid because they mark the end of the translation, causing the disconnection of the ribosome from the RNA. The other sixty triplets, or codons, map nucleotides into amino acids with up to four codons coding for the same amino acid (Figure 1.6).

The correct positioning of the ribosome on the RNA is essential, a shift in the reading frame fundamentally alter the information since the triplets will be interpreted wrongly, leading to the construction of an incorrect sequence



First Position		Second Position								Third Position
		U		C		A		G		
		code	Amio Acid	code	Amio Acid	code	Amio Acid	code	Amio Acid	
	U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
		UUC		UCC		UAC		UGC		C
		UUA	leu	UCA		UAA	STOP	UGA	STOP	A
		UUG		UCG		UAG	STOP	UGG	trp	G
	C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
		CUC		CCC		CAC		CGC		C
		CUA		CCA		CAA	gln	CGA		A
		CUG		CCG		CAG		CGG		G
	A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
		AUC		ACC		AAC		AGC		C
		AUA		ACA		AAA	lys	AGA	arg	A
		AUG	met	ACG		AAG		AGG		G
	G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
		GUC		GCC		GAC		GGC		C
		GUA		GCA		GAA	glu	GGA		A
		GUG		GCG		GAG		GGG		G

Figure 1.6: Conversion table between codons and amino acids

of amino acids. This protein will probably be non-functional, thus a waste of cell's energy. For this reason, a wrong reading frame usually contains stop codons very early in its sequence, causing a premature interruption of the translation.

The ribosome binds the RNA with a strength related to the RBS' sequence and then it positions itself in correspondence of the start codon. The translation is technically realized by another RNA, the transfer RNA or tRNA (Figure 1.7), that, together with the RNA and the ribosome, forms the initiation complex and acts as molecular adapter binding a triplet on one side and the corresponding amino acid on the other side.

As the ribosome shifts on the RNA, other tRNAs make contact with the following codons and the corresponding amino acids are connected with a peptide bond. When the ribosome reach a stop codon, the initiation complex is dismantled due to the action of release factors, and the various subunits are ready to initiate another translation.

At this point, the information contained in the DNA has been transferred to a protein, as stated by the central dogma of molecular biology (Figure 1.8), but in many cases the amino acids' sequence needs to be subjected to structural and chemical modifications before it is fully functional [3].

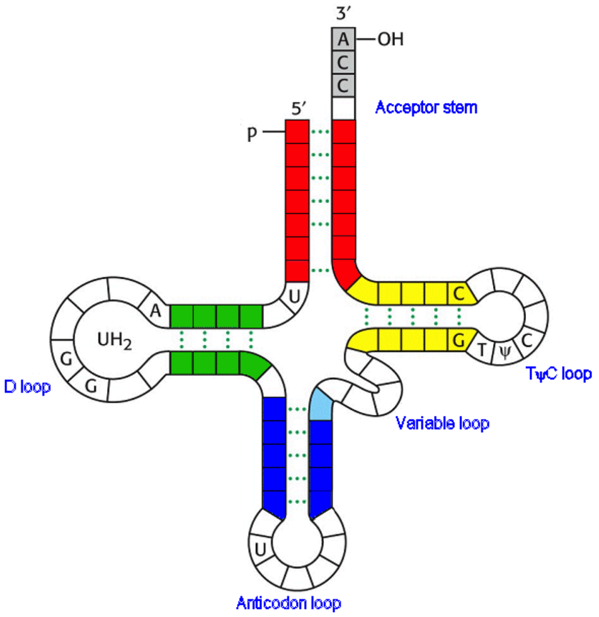


Figure 1.7: Structure of a tRNA

1.4 Natural regulatory systems

In the remainder of this chapter are described two regulatory systems that are naturally found in prokaryotes. These networks have been studied very extensively and are part of many applications of Synthetic Biology.

1.4.1 E. coli lactose digesting system

This system has an extreme importance in the biological research, it has been studied since the beginning of the 1940s and it has been incredibly useful in clarifying some basic notions like the concept of operon [11] and the carbon sources utilization by the enteric bacteria, like E.coli.

An operon is cluster of genes under the control of the same promoter, this structure has the advantage of coupling the transcription of the genes involved and so is very useful whenever it is necessary to have a closely related expression for different genetic components.

The Lac operon, in particular, regulates the lactose metabolism in prokaryotes and is fundamental for the optimization of carbon sources utilization. Bacterial cells preferably utilize monosaccharides, like glucose or fructose, but they are even able of digest disaccharides, if they are readily accessible

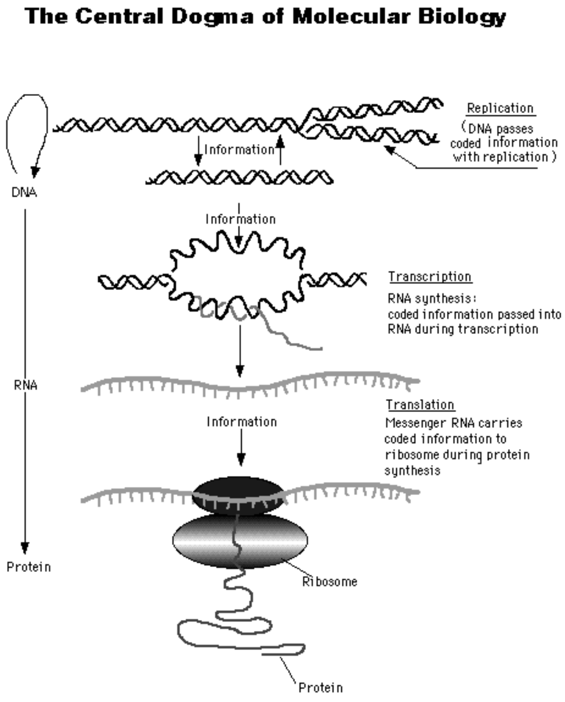

The Central Dogma of Molecular Biology
Replication
(DNA passes coded information with replication)
Information
DNA
Information
Transcription
RNA synthesis: coded information passed into RNA during transcription
RNA
Translation
Messenger RNA carries coded information to ribosome during protein synthesis
Information
Protein
Ribosome
Protein

Figure 1.8: Schematic structure of the Central Dogma of molecular biology

in the environment and the carbon sources of election are unavailable.
The structure of the Lactose digesting system is reported in Figure 1.9, its main element is the operon, on the right hand side, but even the LacI gene has a fundamental activity.

LacI is constitutively produced and acts as a repressor for the Lac promoter, its activity limits the production of the lactose digesting proteins to the background level, reducing the metabolic burden on the cell when this system is not needed.
If lactose is present in the environment a small fraction of it can diffuse into the cell and it is degraded by the β-galactosidase enzyme. A side product of this degradation, allolactose, can bind the LacI protein making it unable to bind the Lac promoter and act as a repressor. So the presence of lactose into the intracellular environment triggers the activation of the operon, it is composed of three genes, LacZ, LacY and LacA. The first one codes for the enzyme β-galactosidase that is able to cleave lactose into glucose and galactose. The expression of the LacY gene leads to the production of the β-galactoside permease, a membrane-bound transport protein that facilitates

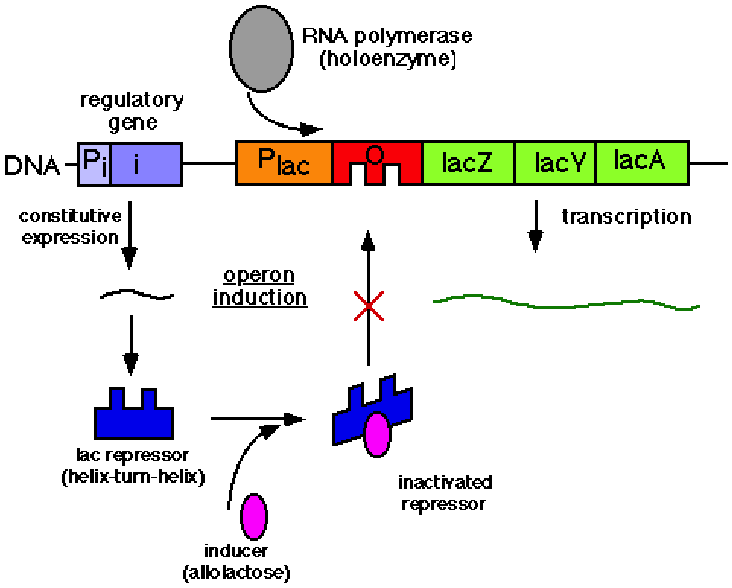

RNA polymerase
(holoenzyme)
regulatory gene
DNA
Pi
i
Plac
O
lacZ
lacY
lacA
constitutive expression
transcription
operon induction
lac repressor
(helix-turn-helix)
inactivated repressor
inducer
(allolactose)

Figure 1.9: Structure of E.coli's lactose digesting system

the import of lactose into the cell. The last gene, LacA, encodes another enzyme, the β-galactoside transacetylase, that chemically modifies the lactose molecule, even if its function is not completely understood.

This system is between the most studied and analyzed, that's probably why its regulatory elements have been widely employed in Molecular Biology and Synthetic Biology.

1.4.2 Luciferase enzyme complex

This system drives the quorum sensing regulated luminescence and has been found in a marine bacteria, *Vibrio fischeri*, in the early 1970s. This environmental sensing system allows bacteria to monitor their own population density and emit light when the bacterial concentration is over a certain threshold. *Vibrio fischeri* can survive free in the ocean but this system is activated only when a bacterial colony maintains a symbiotic relation with marine fishes or squids and the cellular density is high. The Australian Pinecone Fish (left panel of Figure 1.10) uses the light produced by the bacteria to hunt in the coral reef during the night, while the Hawaiian Bobtail Squid (right panel of Figure 1.10) has the light organ within the mantle cavity and, by emitting a light that matches the one that filters through the water, makes itself invisible to the predators below [30].

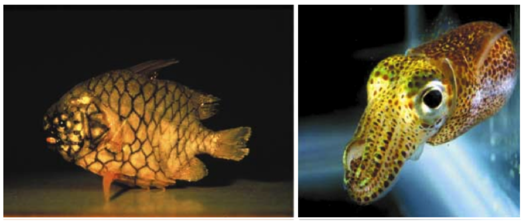


Figure 1.10: Marine animals that maintain a symbiotic relation with *Vibrio Fischeri*

The natural Luciferase enzyme complex is quite elaborated, it is about 8 kbp long and it is composed of two operons containing a total of eight genes (Figure 1.11).

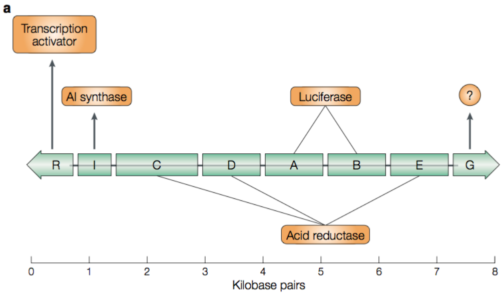


Figure 1.11: Luciferase enzyme complex

The first two genes on the left, LuxI and LuxR, are constitutively transcribed, the former is an enzyme and uses a couple of substrates from the lipids' production pathway to produce the small molecule AHL that can freely diffuse across the cell membrane and accumulates in the environment (Figure 1.12).

When the concentration of the autoinducer reaches the critical concentration, the interaction with LuxR becomes significant. This activation makes LuxR able to induce both the luminescence and the Lux I genes, leading to the production of the luminous signal [29], [30].
This system has been imported in E. coli and other bacterial strains and used for many interesting applications from the realization of synchronized clocks [26] to the implementation of logic gates [5] or of a system to maintain constant the population size, using quorum sensing regulated killing [16].

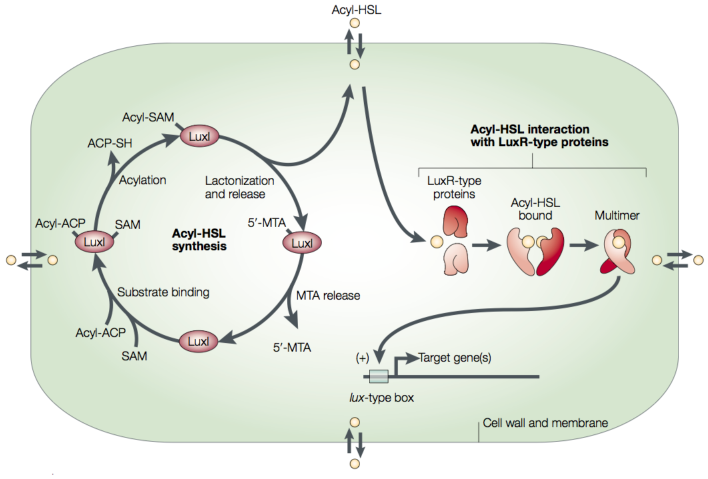

Acyl-HSL
Acyl-SAM
LuxI
ACP-SH
Acylation
Lactonization
and release
Acyl-HSL interaction
with LuxR-type proteins
LuxR-type
proteins
Acyl-HSL
bound
Multimer
Acyl-ACP
SAM
5′-MTA
Acyl-HSL
synthesis
Substrate binding
MTA release
Acyl-ACP
SAM
5′-MTA
(+)
Target gene(s)
lux-type box
Cell wall and membrane

Figure 1.12: LuxI and LuxR communication system

Even if the Luciferase enzyme complex is the most commonly used, probably due to its early discovery, many bacterial strains have similar systems, as described in [7], [18].

Chapter 2

Leader Election project

2.1 Circuit Design

2.1.1 Leader Election problem

The Leader Election is a well known process in distributed computing, and leads to the identification of a single process as the organizer, coordinator, initiator or sequencer of some task distributed among several computers (nodes). The presence of a Leader greatly simplifies the process synchronization providing a centralized controller [24].

Being able to implement the same kind of behaviour in E. coli would provide a basic tool that can be very useful in every application requiring a coordinate action of bacterial cells, like in pattern formation or in the production of new biomaterials. On the scientific research front, engineering a multicellular response in bacteria might give insights into the evolution of eukaryotics from prokaryotes, widen the spectrum of realizable genetic circuits and, with it, their functionalities.

The two main features necessary for the realization of this behaviour are the Leader choosing mechanism and the system of communication. In the computational field the latter is generally defined by the connections between the nodes, while the former can be addressed in many ways even if it's possible to identify two main criteria:

- **Extrema finding:** the leader election is based on global priority.
- **Preference-based:** processes in the group can vote for leader based on personal preference like locality or reliability.

Unfortunately these rules cannot be applied in the biological Leader Election because the cells can't uniquely identify each other. For this reason it

was necessary to find another way to the leader election.
The most straightforward solution might be to use two different bacterial strains for Leader and non Leader cells, this would inevitably increase the number of Leaders since there isn't a reliable method to stop the growth of the bacterial cells without interfering with protein production. Moreover, since the outcome is dependent on the initial condition, it would be better to introduce the minimum number of perturbations. Thus it was chosen to elect the Leader using the stochasticity of gene expression. In a bacterial culture, even if the individuals are genetically identical, the protein expression levels are, generally, quite variable depending on various factors like the polymerases available at a given time or the number of copies of the plasmid. Amplifying this variability can lead to the identification of a few cells, ideally one, as Leaders.
As for the intracellular communication the choice has fallen on the luciferase enzyme complex, that controls the quorum sensing in *Vibrio fischeri* [29], but has been proven functional even in E. coli. This system of communication between cells is quite simple and it's been widely studied and characterized both in its original environment and in other organisms [30],[18],[7]. Moreover the Lux based system has already been used for different applications in synthetic genetic circuits [16],[26],[5].

2.1.2 The device

The Leader Election circuit is represented in Figure 2.1.

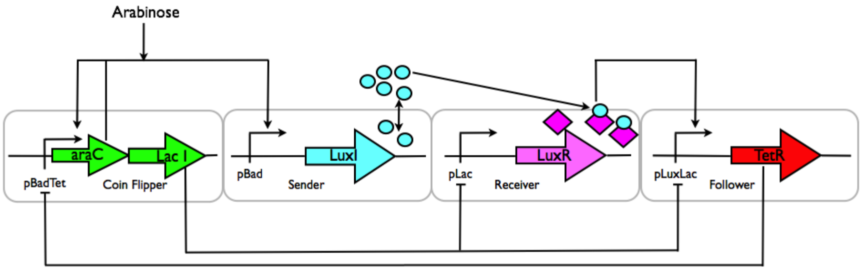


Figure 2.1: Leader Election Circuit [13]

The system is composed of four interacting modules: Coin Flipper, Sender, Receiver and Follower. This design allows the realization of complex circuits as each module can be built and characterized almost independently from the others. The **Coin Flipper** module is the part of the circuit responsible for the election of the Leader, it is composed of an operon regulated by an hybrid promoter. An hybrid promoter is a promoter that is sensitive to two different transcriptional factors, they can be both repressors or activators or one repressor and one activator. In this case the expression of the two genes is down-regulated by TetR, that is a repressor, and up-regulated by the combination of AraC and Arabinose. In a small number of cells, due to the stochasticity of gene expression, there will be a slightly higher concentration of AraC. In these cells, the Arabinose in the media will trigger the positive feedback, turning them into Leaders. Since the goal of the Leader Election project is to obtain the smallest possible number of Leaders, ideally one, it is very important to calibrate both the concentration of Arabinose and the promoter to obtain the correct behaviour. The activation of the hybrid promoter leads to the production of AraC and LacI, the first one is necessary to maintain the positive feedback, while the second one represses the Receiver and Follower modules.

Once a cell is in the Leader state it activates the **Sender** module. The pBAD promoter, that regulates the expression of LuxI, is activated in presence of Arabinose and an higher concentration of AraC. LuxI codes for an enzyme, the AHL synthase, that catalyzes the production of the small molecule AHL from some intermediates of the lipids' production. Since the autoinducer can freely cross the cell membrane and diffuses very quickly, it is readily accessible to the neighbouring cells. The non-Leader cells will sense the signal through the **Receiver** module, the AHL can indeed bind the LuxR protein and become an activator for the Follower module.

The Receiver block is active in all the cells in the initial condition, this is fundamental for the signal to be received but, at the same time, might reduce the stability of the Leader state, since these cells can sense their own signal and respond to it. In order to maintain a clear separation between the two states the repression of the Lac promoter must be very tight and the LuxR degradation must be fast, these two conditions ensure the quick turn off of the signal receiving system in the Leader cells.

The activation of the **Follower module** is driven by the signal and it causes the cells to enter in the Follower state. This element contains the second hybrid promoter of the circuit, even in this case the two operator sites have opposite effects on the gene expression, Lac I acts as repressor while the combination of LuxR and AHL activates the protein production. This promoter drives the production of TetR, that is a repressor for the Coin flipper module, and prevents the Follower cells from becoming Leaders. In the current design the Tet R protein is fused to the red fluorescent protein (RFP) making it easy to identify this class of cells. The Leader cells can be recognized in the same way, since the Coin flipper operon contains even the coding sequence for the green fluorescent protein (GFP).

So the main goal is, starting from an homogeneous population, to obtain two distinct groups of cells, Leaders and Followers, they both are transformed with the same construct but only part of it is active and the genes that are produced define the category in which the cell fall. In the Leader cells the Coin Flipper and the Sender modules are on, as in Figure 2.2, while in the Follower cells the Receiver and the Follower modules are active (Figure 2.3).

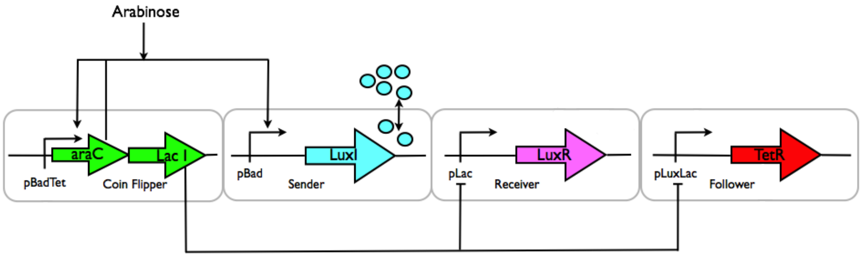

Arabinose
araC
Lac I
pBadTet
Coin Flipper
LuxI
pBad
Sender
LuxR
pLac
Receiver
TetR
pLuxLac
Follower

Figure 2.2: Leader Cells

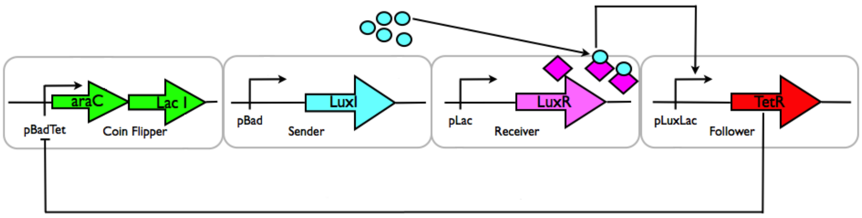

araC
Lac I
pBadTet
Coin Flipper
LuxI
pBad
Sender
LuxR
pLac
Receiver
TetR
pLuxLac
Follower

Figure 2.3: Follower Cells

2.2 Follower Module: a closer look

Besides analysing the different components in isolation, the modularity of the system makes it really easy to build many versions (libraries) of the main components. This diversification can reduce the time necessary for the composition of the final circuit providing, for each element, many alternatives with slightly different activities. The construction of libraries of components can also give insights into the robustness of the behaviour and the effects of each regulatory element on gene expression.

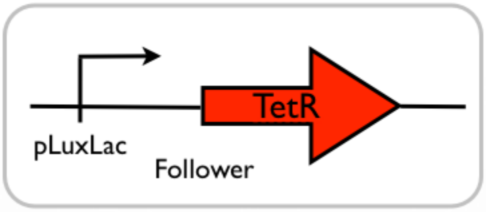


Figure 2.4: Follower Module

2.2.1 The hybrid Lux-Lac promoter

The main component of the Follower module (Figure 2.4) is the hybrid promoter. It needs to be very sensitive to both inducers because it is responsible for the stability of the Follower state and it needs to be triggered by a very low concentration of the signal. Promoter's engineering can be complex because the interaction of each transcription factor with the DNA is specific and not always completely understood.

In [22] the realization and characterization of a huge library of regulated promoters is described. Four transcription factors are considered, two activators, AraC and LuxR, and two repressors, LacI and TetR. The promoters were built combinatorially by mixing in a single reaction three classes of synthetic duplex DNA sequences with compatible 5' cohesive ends. These units correspond to the Distal, Core and Proximal regions in Figure 2.5.

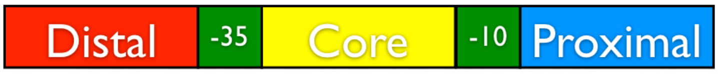


Figure 2.5: Promoters' architecture

Each unit was designed from a sequence known to be sensitive to one of the four transcription factors with variations in consensus sequence strength, transcription factor binding site strength, spacing and orientation [22]. The promoters have been assembled from a combination of a distal, a core and a proximal unit, with the position of each block defined by the 5' overhang, then they've been tested and sequenced in order to relate the functionality of each promoter to its architecture.

From these data the authors defined a set of heuristic rules that should guide the design of engineered regulated promoters:

- **Limits of Regulation.** Gene expression can be regulated over five order of magnitude. Regulated promoter activity is independent of unregulated activity. As a result, effective repression tends to increase with unregulated activity, whereas activation tends to decrease. Activation is limited by an absolute level of expression, at around 2.5% the level of the strongest unregulated promoter activities.
- **Repressor operator locator.** The effectiveness of repression depends on the operator location with core $\geq$ proximal $\geq$ distal.
- **One is enough.** Full repression is possible with a single operator between -60 and +20 at high repressor concentration. Activators function only upstream of -35 (distal), and have little positive or negative effect downstream, at core or proximal.
- **Repression dominates activation** producing an asymmetric logic.
- **Operator proximity.** Independent regulators generate SLOPE-like logic. Operator proximity increases competitive interactions, making the logic more AND-like.

The genetic circuit used to obtain these results is very simple and it's been designed so that the induction of each factor is expected to increase a target promoter's activity. Due to the construction method, many of the promoters of the library don't respect the previous rules and so their behaviour is not ideal, in particular the two Lux/Lac promoters with the activity closest to the desired one are the A81 and the B72 and their expressions are reported in Figure 2.6 and Figure 2.7.

The production of the reporter protein doesn't change adding VAI, the synthetic AHL, probably because the Lux site is in proximal.
Thus four new promoters were designed, following the rules stated in [22] and using operator sites very sensitive to the inducers, according to the available data.

The promoters' architecture is described in Table 2.1, they all have one region occupied by a Lac site and at least another one contains a Lux site, since they need to be sensitive to these transcription factors. Depending on how the third slot is filled, they can be divided into two groups. In the first one the promoters contain another Lux site while, in the elements of the second group, the sequences are completed with a constitutive block that shouldn't interfere with the gene expression. Regardless of the category, the

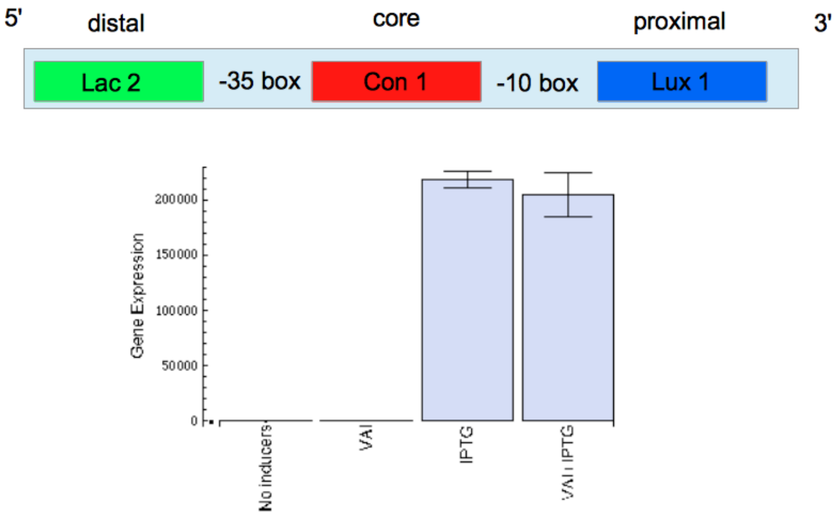


Figure 2.6: A81 promoter

	Distal	-35	Core	-10	Proximal
Prototype	Lux 1	TTTACT	Lac 2	GATACT	Lux 2
1	Lux 1	TTTACT	Lac 2	GATAAT	Lux 2
2	Lux 1	TTTACT	Lux 2	TATTTT	Lac 2
3	Lux 1	TTTACT	Lac 2	GATAAT	Con 2
4	Lux 1	TTTACT	Con 2	GATAAT	Lac 2

Table 2.1: Promoters' architecture

distal site is always filled with Lux 1, this sequence should be the most sensitive to this transcription factor and should grant the highest gain upon induction. Placing it in the most effective region for the activators should increase the chances of obtaining the correct behaviour.

The other Lux sensitive sequence that has been used is Lux 2, it is placed in proximal and core for promoters 1 and 2 respectively. The sequences of these promoters are completed with the Lac 2 site, the same contained in promoter A81.

This Lac sensitive sequence is used even in the promoters of the second group, in core for promoter 3 and proximal for promoter 4. The third slot for these constructs is filled with a constitutive region, Con 2.

Even the -35 and - 10 boxes are customizable, even if partly constrained by the upstream region. These sequences have been chosen, between the

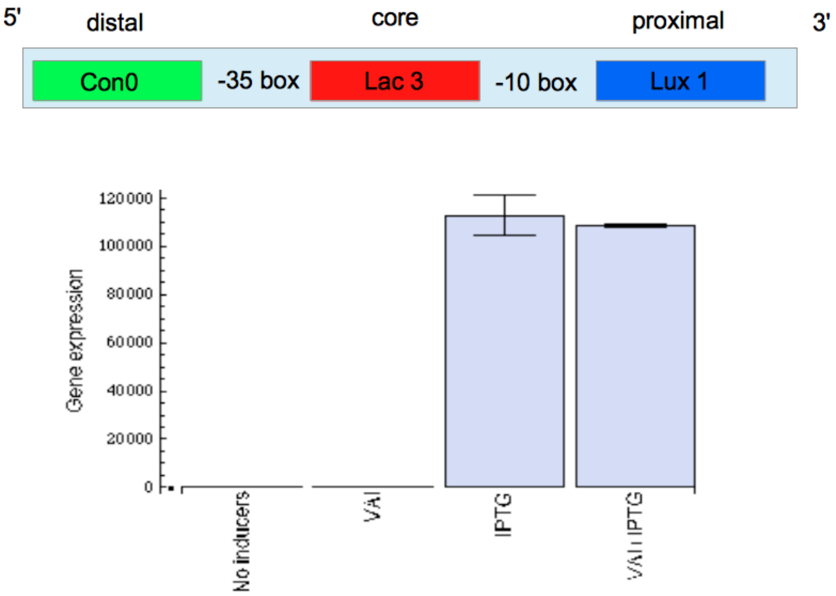


Figure 2.7: B72 promoter

available options, so to grant the lowest unregulated activity. Unregulated activity refers to the protein production in absence of transcription factors bounded to the DNA and, according to the specifics of the project, it needs to be as low as possible.
In Figure 2.8 are reported the unregulated activities of the promoters described in [22] divided according to the sequences of the -35 and -10 boxes. From these data should be possible to infer the strength of these regions without the interference of the transcription factors. But this representation has some limitations, for example it doesn't highlight the frequency of occurrence of each RNA-polymerase binding sequence in the library. The histograms of a very rare sequence might not be statistically significant, thus leading to a misleading interpretation of the graph. Furthermore, the data relative to promoters regulated by different transcription factors are considered together. This shouldn't have any effect on these results, but, since the interaction between these molecules and the DNA is not completely understood, a transcription factor might influence these data not only through its direct action but even through its frequency of appearance in the library.

Nevertheless the authors tried to infer the unregulated activity of each promoter from the median activity of the -35 and -10 box distribution. According to their results there's only a very weak correlation between the

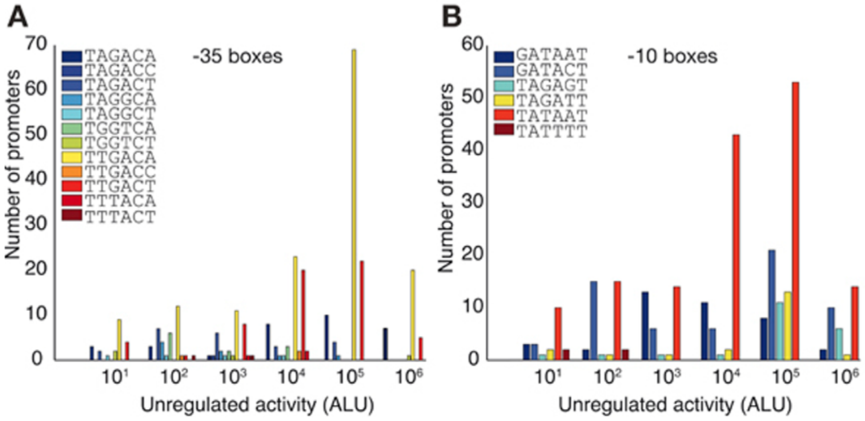


Figure 2.8: Unregulated activity

unregulated activity and the polymerase boxes, so the presence of strong polymerase boxes doesn't grant an high promoter activity, even if a strong promoter usually contains strong boxes.

Even with all the shortcomings, these data might provide a useful heuristic for choosing these regions of the promoters so, for the -35 box, were considered two options, compatible with the constrains imposed by Lux 1:

- TTTACA
- TTTACT

The second one was preferred, due to the lower number of promoters with unregulated activity between 10^3 and 10^6 and was used for all the promoters of the library.

As for the - 10 box the choices for promoter 1, 3 and 4 were:

- GATAAT
- GATACT

So, according to the B panel of Figure 2.8 the most suitable for the Leader Election project is the first. The Lux site, in the core region of the second promoter, constrains the first three bases of the -10 box to be *TAT*, so the choice obviously fell on *TATTTT*.

In the following analysis the activity of these promoters have been compared to the one of the prototype promoter. This promoter was the one originally designed for this module and its architecture is very similar to the first promoter, the only difference is in the sequence of the - 10 box.

2.2.2 The Shine Dalgarno (RBS) sequence

Another typical way of tuning gene expression is modifying the ribosome binding site. This kind of regulation acts at the translational level, modifying the interaction between the ribosome and the RNA. Due to the stability of this connection over the different constructs, it has been studied more closely and is better understood.

In fact there is a standard computational method to design synthetic RBS regions. In [12] is described a statistical thermodynamic model that quantifies the strength of the molecular interaction between the RNA and the ribosome complex, to predict the resulting translation initiation rate. Since the initiation phase normally is considered the slowest step of the translation process, the initiation rate can be used as an approximation of the rate of translation.

This model has been used to create an online application [17] for the automatic design of synthetic RBSs. Given the sequences upstream (about 5-20 nt) and downstream (at least 50 nt) of the Shine Dalgarno region, this system can design a ribosome binding sequence with a given translation initiation rate. It is even possible to impose some constrains on the sequence like the length or the bases in certain positions. In the same way it can solve the reverse engineering problem, retrieving the strength of the RBS from its sequence.

A substitution of the Shine Dalgarno sequence is certainly a good method to set the translation initiation rate in the desired range, but to tune it precisely around the working point it is better to use the method described in [21]. This technique uses hypermutable simple sequences repeats embedded in the spacer region of the ribosome binding site (Figure 2.9).

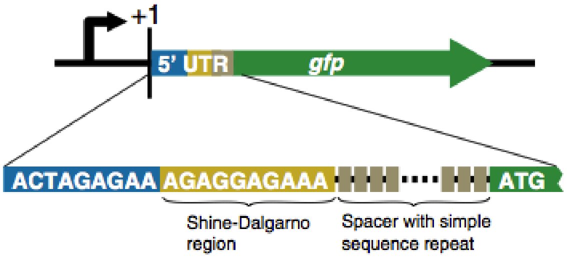


Figure 2.9: Structure of the RBS region

By changing the length of the spacer it is possible to tune the output, represented by the level of expression of a fluorescent protein, almost linearly in the semi-logarithmic scale, as reported in Figure 2.10. An increase in the spacer length gradually reduces the gene expression even if the trend of the decline is sensitive to the nucleotide composition of the repeats.

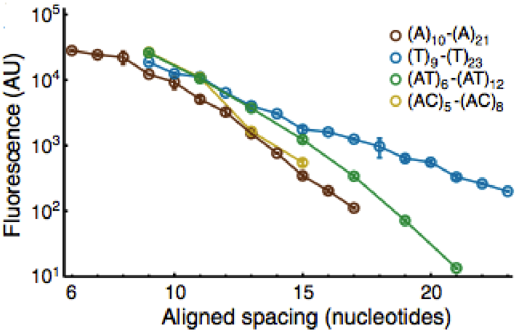


Figure 2.10: Relation between the length of the spacer and the fluorescence level

In the Follower module a poly-A repeat has been used, a sequence composed of only adenine triphosphate, of the length of ten nucleotides. This choice places the Shine-Dalgarno region and the first codon to be translated, at the ideal relative distance, providing the highest translation initiation rate and the opportunity of testing the system at the top of its productive capabilities. Another advantage of this system is that the protein production can be scaled down easily and predictably, if needed.

The RBS used in the Follower module is the Bio-Brick number BBa_B0030, its sequence is present in the Registry of Standard Biological parts [2] and it a very common component that has been widely used in many different applications and tested widely. It is a very strong Shine Dalgarno region, according to the characterization reported in [2], and its efficiency is 0.6 with respect to the activity of the Bio-Brick number BBa_B0034, that is considered as a point of reference.

Chapter 3

Experimental Characterization

3.1 Assembly reaction

Each promoter's sequence is about one hundred nucleotides long and they've been designed to be inserted into the follower plasmid (Figure 3.1). This plasmid is a circular DNA segment built to test the Follower module in isolation, it contains some standard elements like the antibiotic resistance gene and the replication origin (ORI) together with the hybrid Lux-Lac promoter and the TetR protein.

The backbone that has been used in this construct, is the standard BioBrick $BBa\_K590010$ (pGA1A3), it is an high copy number plasmid, defined Gibson assembly friendly in [2], that confers ampicillin resistance and has a total length of about 3.6 kbp. The hybrid promoter is identified with D49delta in Figure 3.1, and is placed upstream of the RBS (BBa_0030) and of the sequence that codes for a fusion protein (TetR-mRFP).
To substitute the promoter region of the follower plasmid has been used a three pieces Gibson assembly reaction [8]. As shown in Figure 3.2 this reaction is realized adding T5 exonucleases, polymerases and Taq ligases to a mixture of the DNA sequences, in the appropriate concentrations, and then keeping the solution at 50^oC till the completion of the ligation (see the Material and Methods chapter for a detailed description).

In this case the three DNA sequences used in the reaction are the promoter region, from the bglBrkPrefix to the untraslated region (UTR) upstream of the RBS; the insert, that contains the sequence coding for the proteins of the Follower module and the backbone that goes from the bglBrk Suffix to the bglBrkPrefix. This solution was chosen over a “classical” two pieces reaction to increase the efficiency. The combination of DNA regions with very different lengths has usually a very low yield because most of the longer sequences

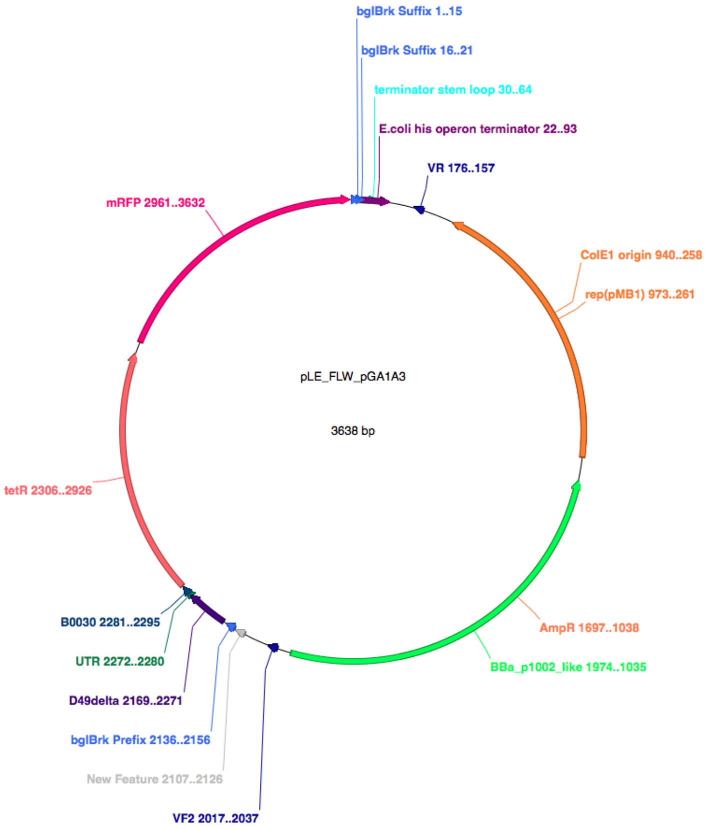

bglBrk Suffix 1..15
bglBrk Suffix 16..21
terminator stem loop 30..64
E.coli his operon terminator 22..93
VR 176..157
mRFP 2961..3632
ColE1 origin 940..258
rep(pMB1) 973..261
pLE_FLW_pGA1A3
3638 bp
tetR 2306..2926
B0030 2281..2295
AmpR 1697..1038
UTR 2272..2280
BBa_p1002_like 1974..1035
D49delta 2169..2271
bglBrk Prefix 2136..2156
New Feature 2107..2126
VF2 2017..2037

Figure 3.1: Follower plasmid

will close on themselves.

The DNA sequences used in this procedure were amplified with a PCR reaction, see Material and Methods for a description of this technique. Custom designed primers were used to isolate and amplify the desired regions from already existent constructs. Only the promoters were assembled differently, short DNA regions, named oligonucleotides, were synthetized and then combined, again with a PCR reaction, to build the complete sequence.

For each promoter were used four oligonucleotides, they were designed so that the ends are complementary to the neighbouring sequences with an overlapping region of about 25-30 nucleotides. The exact length of these regions have been calibrated so to maintain an almost constant melting temperature.

In Figure 3.3 is highlighted the structure of the promoters, as combination of oligonucleotides. Since the backbone doesn't change and the distal site is always filled with the Lux 1 site, the first oligo, identified with the letter

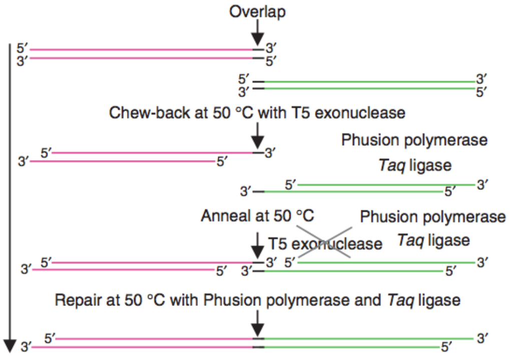

Overlap
5′
3′
3′
5′
5′
3′
3′
5′
Chew-back at 50 °C with T5 exonuclease
Phusion polymerase
Taq ligase
Anneal at 50 °C
Phusion polymerase
T5 exonuclease
Taq ligase
Repair at 50 °C with Phusion polymerase and Taq ligase

Figure 3.2: Gibson assembly reaction

A, is shared between all the promoters. The B sequence is present in both promoter 1 and 3, because even their core regions are the same.

3.2 Single module Characterization

The Characterization of the promoters has been divided in two phases: the analysis of the behaviour of the single module, that is needed to define the intrinsic activity of the regulatory sequences and the study of the connectivity with the other modules. In particular it has been evaluated the interaction with the Sender, to test potential unwanted interactions and the change in the activity induced by the contemporary presence of the two elements.

In the characterization experiments have been used reduced versions of the circuit in Figure 2.1, designed to highlight particular aspects of the behaviour of the system, like the reaction to the introduction of particular molecules, inducers, into the environment. In this case the intensity of particular fluorescent proteins has been used as output. This choice is incredibly convenient because it immediately provides information on the functioning of the circuit, since the production of RFP and GFP is strictly correlated with the activation of the genes characteristic of the two states.

The fluorescence is a type of luminescence that is found in some substances that, after absorbing light at a particular wavelength, are able to emit light at another wavelength, usually longer. This principle is used very often in biological assays, since it's easy to use and many proteins exhibit this propriety.

For this characterization, in particular, the fluorescence has been recorded with a flow cytometer. It measures the fluorescence of the single cells instead

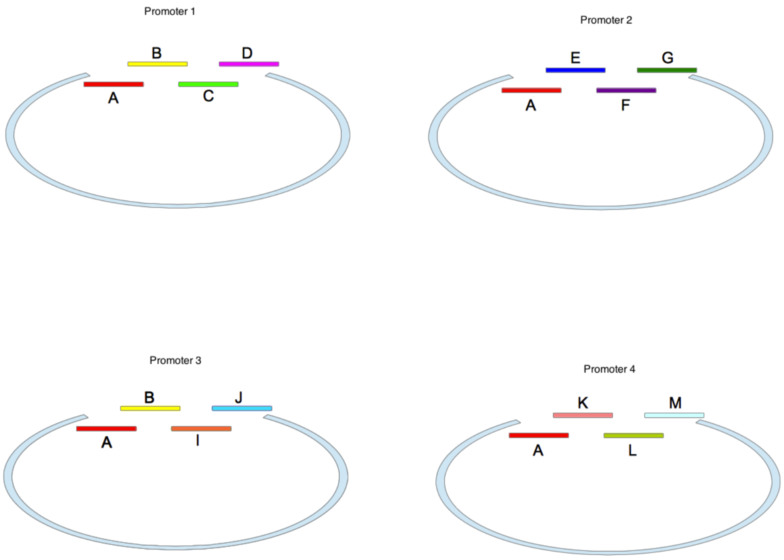


Figure 3.3: Oligo assembly structure

of the mean of the population and can record many fluorescent proteins at the same time.
In Figure 3.4 there's a scheme of the functioning of the instrument. The cells are suspended in liquid media and flow in a capillary where they are hit by the light emitted by a laser, at the excitation frequency. Sensors are then able to detect the light emitted from the cells, using special filters to isolate it. Adding more lasers and sensors makes it possible to measure, on the same cells, the intensity of different fluorescent proteins.

3.2.1 Desired Behaviour

The circuit in Figure 3.5 was used to characterize the functionality of the promoters in isolation. It contains both the Follower and the Receiver module due to the close relation between these two elements, the former needs the latter in order to sense the signal.

This system is transformed in MG1655 Z1 E. Coli cells, this strain has the characteristic of over-express LacI from the genome so, in absence of inducers, both the promoters are repressed. This block can be removed adding IPTG, this is a molecule that mimics the action of the allolactose in the natural

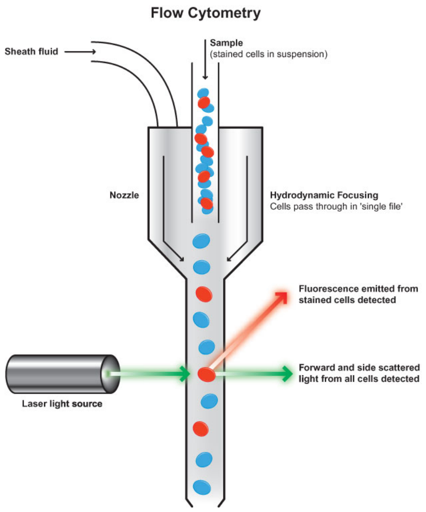

Flow Cytometry
Sheath fluid
Sample
(stained cells in suspension)
Nozzle
Hydrodynamic Focusing
Cells pass through in 'single file'
Fluorescence emitted from
stained cells detected
Forward and side scattered
light from all cells detected
Laser light source

Figure 3.4: Flow cytometer

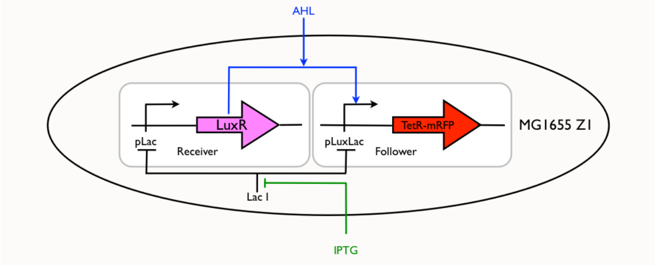

AHL
LuxR
pLac
Receiver
TetR-mRFP
pLuxLac
Follower
MG1655 Z1
Lac I
IPTG

Figure 3.5: Reduced circuit used to characterize the Follower module

Lactose digesting system.
The other input of this system is the synthetic AHL (VAI) that, bound with LuxR, can activate the expression of the Follower module. This circuit, can be compared to a boolean AND (Figure 3.6), the AND gate accepts two binary inputs and computes an exit according to the relation reported in Table 3.1. So we expect to have a considerably high expression of the output protein only in presence of both inducers.

But this comparison is not entirely correct because in a genetic circuit both the inputs and the output are continuous quantities and so, depending on the arbitrary threshold chosen to distinguish between the low and the high level, the classification of the behaviour might change. So it was necessary to find a better way to characterize the activity of the promoters.
In [22] are defined three indexes (3.1, 3.3, 3.4) specifically designed to solve

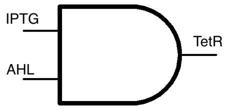


Figure 3.6: AND gate

IPTG	AHL	TetR-mRFP
0	0	0
0	1	0
1	0	0
1	1	1

Table 3.1: Truth table of a logic AND

this problem, they can be computed using the levels of fluorescence measured in four different conditions:

- b_1 no inducers
- b_2 and b_3: only one inducer
- b_4: both inducers.

The first index r, Equation (3.1), measures the gain of the system as the ratio between the fully induced and uninduced condition. The expression (3.2) is used to calculate the error on this index.

$$r = \frac{b_4}{b_1} \tag{3.1}$$

$$\Delta r = r\sqrt{(\frac{\Delta b_1}{b_1})^2 + (\frac{\Delta b_4}{b_4})^2} \tag{3.2}$$

In Equation (3.3) is defined the second index, a, that measures the asymmetry of the gate with respect to the two inputs, ranging from perfectly symmetric (a=0) to the completely asymmetric (a=1). A completely asymmetric promoter is sensitive only to one of the inducers.

$$a = \frac{log b_3 - log b_2}{log r} \tag{3.3}$$

The characterization of the behaviour is completed by the quantity defined in Equation (3.4), named l, that quantifies the logic type. This index is equal to 0 for a boolean OR and 1 for a boolean AND.

$$l = \frac{2log b_4 - log(b_2 b_3)}{2log r} \tag{3.4}$$

In order for these numbers to be meaningful it is necessary that the fluorescence levels used in the computation satisfy the following expression:

$$b_1 \leq b_2 \leq b_3 \leq b_4 \tag{3.5}$$

The information carried by the two last indexes, Equation (3.3) and Equation (3.4), can be synthetized in a single image. In Figure 3.7 the asymmetry is graphed with respect to the logic. Every possible gate has a behaviour that falls within the dotted triangle and so it is possible to define, with greater precision, the similarities between the activity of the genetic circuits of interest and the logic gates.

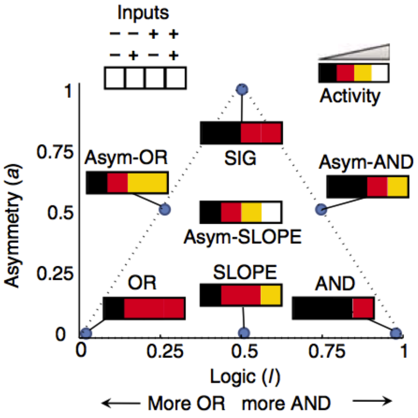


Figure 3.7: representation of the two indexes l and a for logic gates

3.2.2 Saturating concentrations

In the first stage of the characterization of the Follower module were used saturating concentrations of both inducers. This phase allowed to compare the basic behaviour of the promoters to the expected one. From these data it was even possible to determine the values of the indexes described in the above section. Due to the lack of time and some problems in the cloning procedure, the behaviour has been fully characterized only for promoter 1 and 3.

A concentration is saturating when in the environment there are enough inducers' molecules to fill all the regulatory sites, in any given moment. But, due to the variability in gene expression, it is not easy to define it uniquely. The most straightforward approach consists in empirically find a concentration of inducer, that is enough to elicit the peak of the response, without interfere with the basic cell functioning.
For these experiments was chosen a concentration of $500\mu M$ for the IPTG and of $100\mu M$ for the VAI. The details of the experimental protocol, used for these experiments, are described in the Material and Methods chapter but, the main steps are:

- Bring the overnight cultures back to the exponential growth phase.
- Add the inducers to the growth media, in the appropriate concentrations.
- Incubate at 37^oC for a defined period of time (usually a few hours).
- Measure the output, the expression of a fluorescent protein in this case.

Using the cytometer were recorded the levels of RFP from 50000 cells. These values were used to compute some statistical parameters like the mean, the median and the coefficient of variation. In Figure 3.8 are represented the mean values of the red fluorescent protein, recorded from the analysis of the reduced circuit in Figure 3.5. Each set of columns refers to a device with a different promoter and each color reports the data obtained with a different configuration of inducers.

On the left hand side there are the data relative to the prototype promoter. The blue column refers to the uninduced condition, the cyan and yellow ones to the presence of only one transcription factor, IPTG and VAI respectively, and the dark red column is relative to the presence of both IPTG and VAI in the media. The central set of data, identified with SSJ140, was recorded from the construct containing promoter number 1 while the last group of columns is part of the characterization of the third promoter. The errorbars in Figure 3.8 refers to the standard deviation computed over triplicates.
The overall behaviour of the three constructs is AND-like, with the expression of the fluorescent protein considerably higher only in presence of both inducers, but it is possible to point out many differences in the pattern of expression. For example promoters 1 and 3 have a gain three folds higher than the prototype, in presence of both inducers. Furthermore promoter 3 has a considerably higher variability and its sensibility is biased toward the

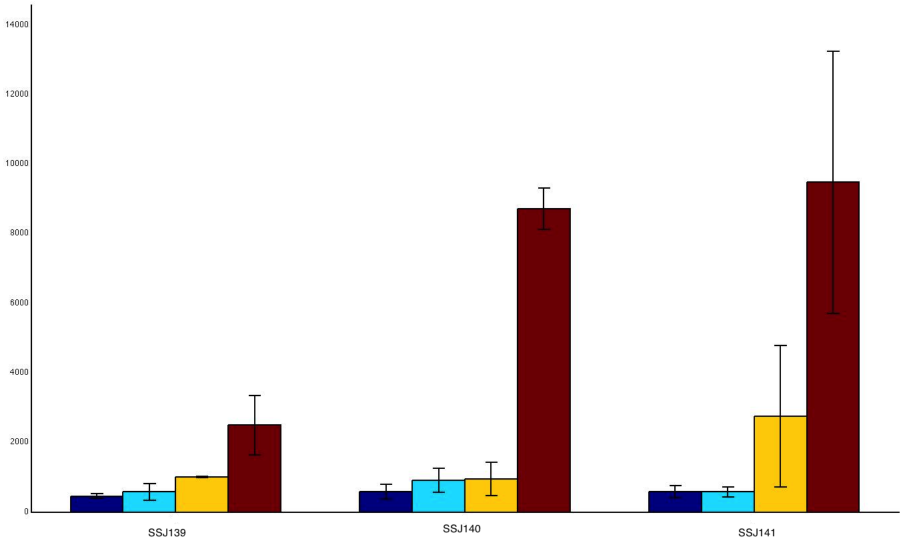


Figure 3.8: RFP expression of the Follower module when induced with saturating concentrations of inducers

	r $\pm\Delta r$	l	a
prototype	6 ± 2.02	0.53	0.17
1	14.5 ± 5.7	0.79	0.09
3	18.4 ± 5.5	0.72	0.56

Table 3.2: Indexes

VAI. To verify the statistical significance of these statements a T-test was performed confirming that the distributions are different with a 5% significance level.
The values obtained from this experiment were used to compute the three indexes defined in the previous section, Equation (3.1), Equation (3.3) and Equation (3.4). The results are reported in Table 3.2.

Again it is clear that the prototype promoter has a lower dynamic range, since r is smaller and it is even less AND-like, because l is further from 1. Additionally the third promoter is the least symmetric, since a is higher. From the position of the three promoters in the triangle of Figure 3.7, it is clear that the promoter number 1 has the behaviour most similar to a logic AND, since it is closer to the lower right corner, the prototype promoter has a response closer to a SLOPE gate, while the third promoter is definitively an Asymmetric-AND (Figure 3.9).

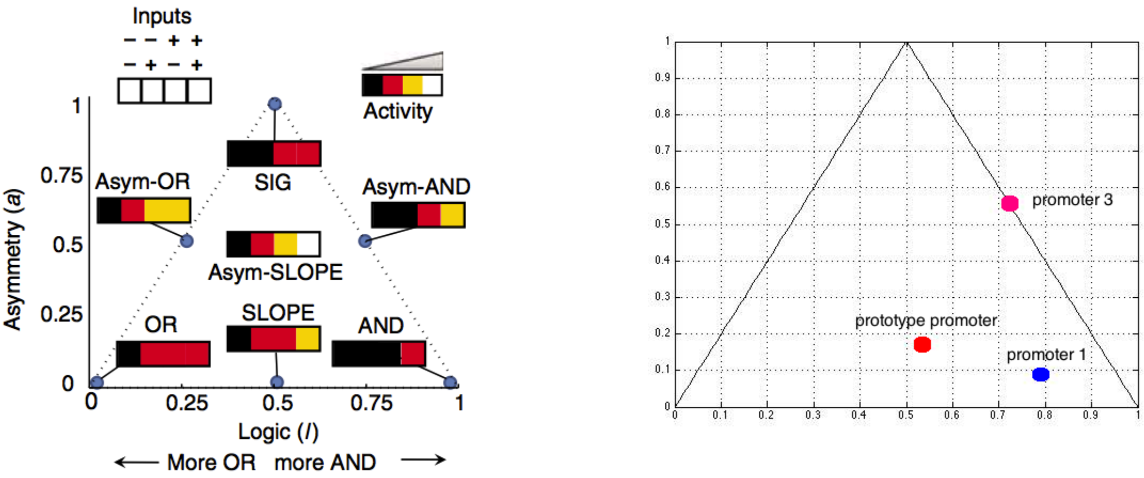


Figure 3.9: representation of the two indexes l and a for the reduced circuits

3.2.3 Dose Response

Since the Follower module needs to be integrated in a bigger circuit, in which the signals that are now used as input are produced by other cells, the next step was to verify that the correct behaviour was maintained even at lower concentrations of inducers, this was tested by building dose response curves. A dose response curve aims to characterize the variation of the behaviour of the system induced by the variation of its inputs. It is realized stimulating the circuit with different values of the inputs and measuring the variation of the output. The bi-dimensional continuous space has been sampled in a few interesting points performing the same experiment described in the previous section, just with different concentrations of the inducers.
Considering the saturating values as upper bound, three lower values were chosen for IPTG (0 μM, 100 μM, 250 μM) and four for VAI (0 μM, 500 nM, 10 μM, 50 μM). These results were combined to the previous ones to obtain the graphs in Figure 3.10, Figure 3.11 and Figure 3.12.

The behaviour of the system maintains the same characteristics, even for lower concentrations of the inputs. This robustness is fundamental for the functioning of the complete circuit, since is not going to be possible to directly control the concentration of the inducers. For promoters 1 and the

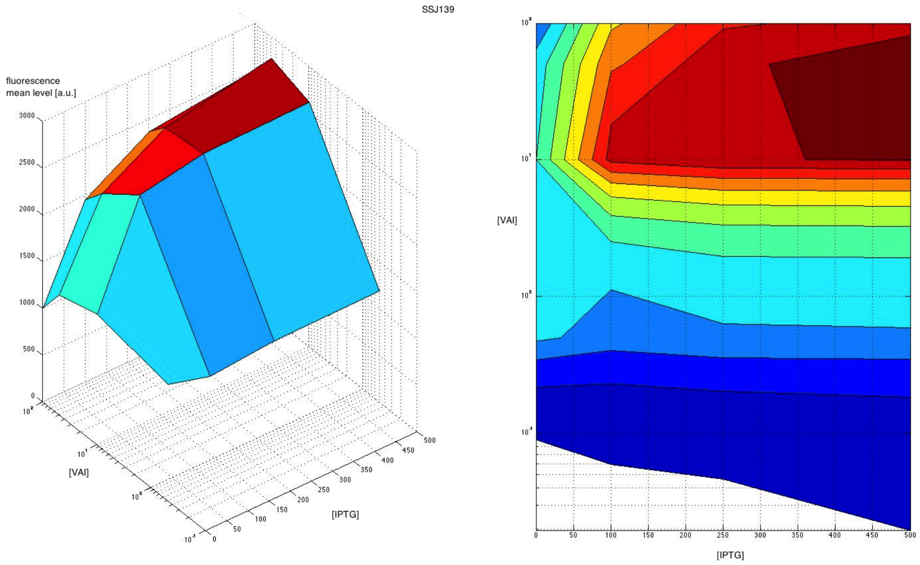


Figure 3.10: Dose response curve for the prototype promoter

prototype the saturating concentration of VAI seems to be already too high, since the protein expression is decreasing. The shape of the three surfaces is similar, but there are some differences: again it is possible to see that the third promoter is asymmetric, with a higher sensitivity to VAI, and that the dynamic range of promoters 1 and 3 is higher.

3.2.4 Fitting

The dose response curves were fitted to the product of two Hill's functions, Equation (3.6), to test how the differences in gene expression influence the parameters of the mathematical representation.
G_{max} is the gain of the system while AHL and IPTG are the concentrations of the inducers. K_{IPTG} and K_{AHL} are the half activation constants and represent the concentrations of inputs necessary to obtain half of the maximum expression. The Hill's coefficients, n and m, define the shape of the curve, as showed in Figure 3.13. An higher value for these parameters cause an increase of the steepness of the function, making it more similar to a step function. From the curves in Figure 3.13 it is even possible to notice that the most drastic change takes place between the values 1 and 2, this range is probably even the one in which m and n will be, according to literature

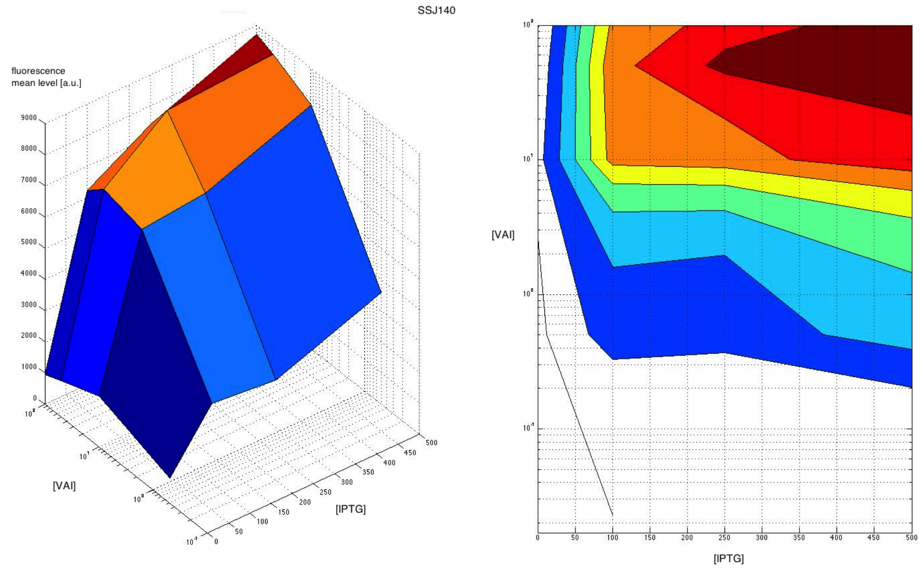


Figure 3.11: Dose response curve for the first promoter

data. The last parameter of the function, b, is the offset.

$$f = G_{max} \frac{AHL^n IPTG^m}{(K_{AHL}^n + AHL^n)(K_{IPTG}^m + IPTG^m)} + b \tag{3.6}$$

The values of these parameters were found with the least square method, starting from the initial condition reported in Table 3.3. A Matlab script was used to minimize a cost function, defined as the absolute value of the sum of the elements of a vector. This vector was computed as the difference between the value of Equation (3.6), in the current point of the parameter space, and the experimental data.

The optimal values of the parameters for the three promoters are reported in Table 3.4.

The surfaces, obtained using these values, are plotted in the right panels of Figure 3.14, Figure 3.15 and Figure 3.16. They correctly reproduce the experimental data, represented in the left panels of the same images, but are unable to capture all the details and the differences of the three systems.

Even if the approximation is not perfect these values give us some information about the possible causes of the variations between the different constructs.

One thing to notice is that G_{max} is sensibly lower for the prototype promoter

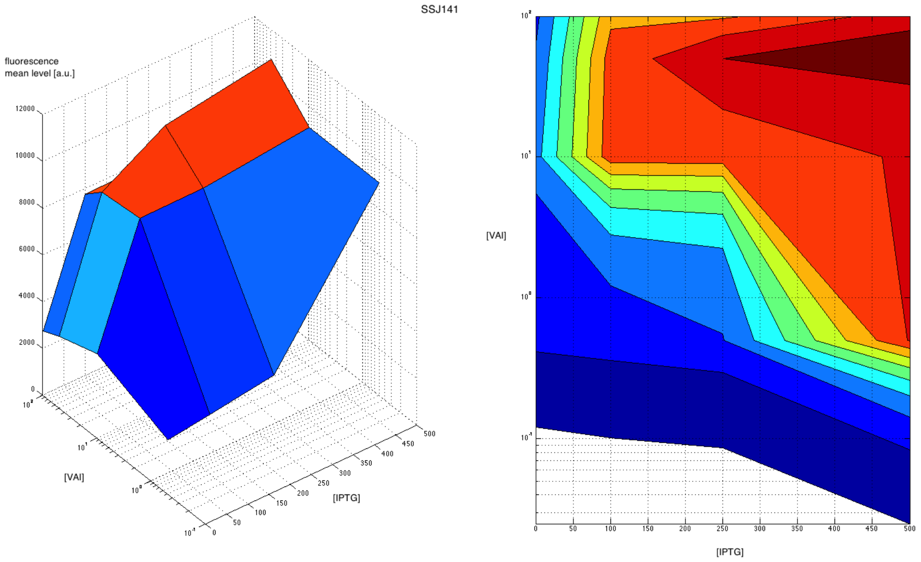


Figure 3.12: Dose response curve for the third promoter

while it is of the same order of magnitude for promoter 1 and 3. The same goes for K_{AHL} but, in this case, the prototype promoter's value is higher, meaning that this construct has a lower sensitivity to VAI and that an higher concentration of inducer is necessary to reach the half activation.
Promoter 3 has a considerably lower value of K_{IPTG}, this can partly account for the asymmetry of this system since, given a fixed concentration of VAI, the amount of IPTG necessary to reach the half activation is lower.
The Hill coefficient for the VAI, n, decreases going from the construct with promoter 3 to the one with the first promoter, and it is even further reduced

Parameter	Value
G_{max}	8800
K_{AHL}	10
K_{IPTG}	10
n	1
m	1
b	500

Table 3.3: Initial values

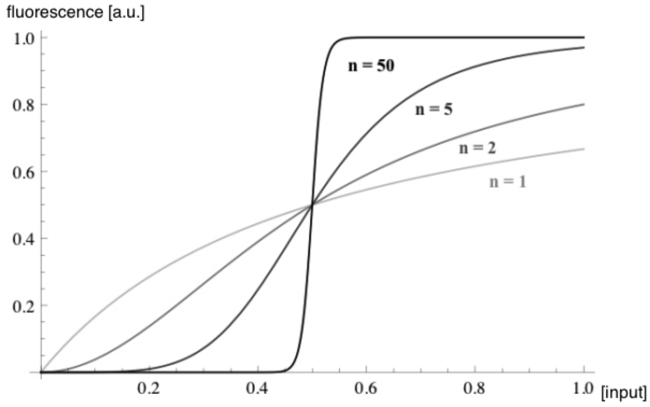


Figure 3.13: Effect of the Hill coefficient on the shape of the curve

	Promoter 1	Promoter 3	Prototype promoter
G_{max}	8411.9	8394.7	3317.3
K_{AHL}	0.99	0.84	15.45
K_{IPTG}	18.96	6.56	13.69
n	1.38	2.05	0.38
m	0.48	0.91	0.63
b	924.73	889.77	576.13

Table 3.4: Optimal parameters

for the prototype promoter. This means that the steepness of the function is diminishing, as shown in Figure 3.13. This information is important to determine the local sensitivity of the system to the inducer.
The IPTG's Hill coefficient, m, stays more or less constant over the three constructs, but is always less then one, this makes the curve concave, see Figure 3.14 were both n and m are in this range as an example. The offset is numerically lower for the prototype promoter but, due to the lower gain of this version of the system, its dynamic range is still quite small.

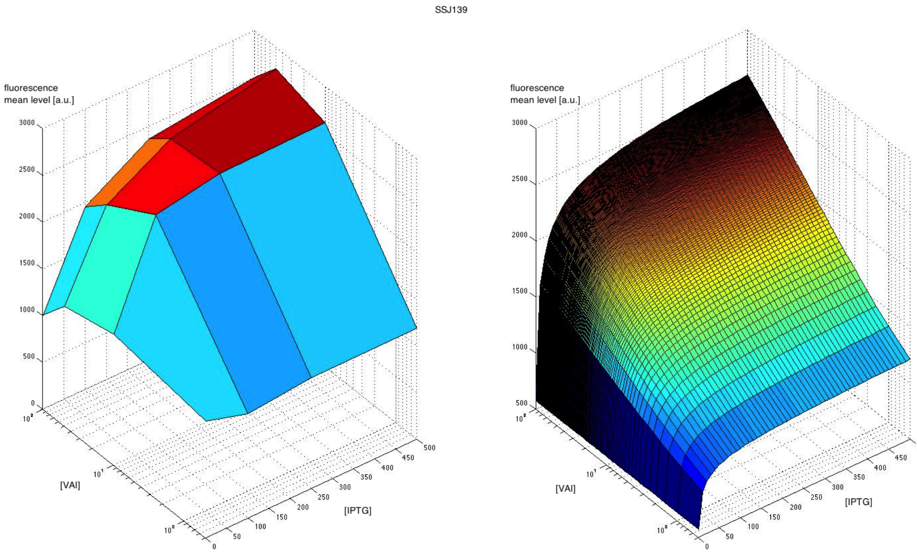


Figure 3.14: Fitting of the data of the system with the prototype promoter

3.3 Interaction with the other components

Another fundamental step in the characterization, is the evaluation of the changes introduced in behaviour through the interaction with the other modules. For the follower module, in particular, the most important interaction is with the Sender. These two modules realize the mean of communication between the different cells, that is essential for the realization of the complete behaviour. At the same time the introduction of the Sender element fundamentally change the experimental environment [27], and this may even induce changes in the behaviour of the Follower module. So this condition was completely analysed with another series of induction assays.

The first set of experiments, the Supernatant assay, aims to test the ability of the Follower module to receive the signal produced by the Sender while, in the Coculture assay, Sender and Follower cells are introduced in the same environment to test the communication *in vivo*.

3.3.1 Supernatant assay

For this first assay it has been used the same construct employed in the characterization of the single module (Figure 3.5). The difference, in this

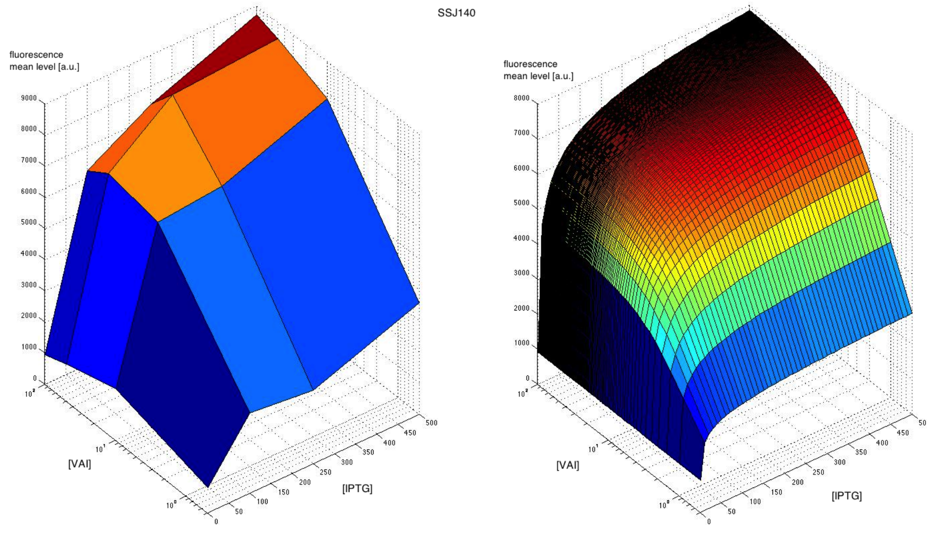


Figure 3.15: Fitting of the data of the system with the first promoter

case, is that instead of adding synthetic VAI it was used the AHL produced by the Sender cells.
The Sender cells are schematically represented in Figure 3.17, they contain both the Sender module and the RGE 271 system. The former contains, as previously described, the LuxI gene driven by a promoter sensitive to AraC bound with L-Arabinose, while the latter constitutively produce GFP, the green fluorescent protein.

These cells were cultured overnight in presence of L-Arabinose (0.5%) to induce the production of the signal, the supernatant of this culture was then used as inducer for the Follower module showed in Figure 3.5. This kind of preliminary assay is possible because the AHL, produced by the Sender cells, can freely diffuse through the cell membrane, so its concentration in the supernatant and inside the cells is the same.
To replenish the nutrients in the supernatant of the overnight culture, it was mixed with fresh media in ratios of 1:1 and 1:3. This step is fundamental to allow the growth of the Follower cells but, at the same time, reduces the AHL's concentration making the induction less effective. This assay is still very important because it allows to easily test a fundamental aspect of the behaviour of the complete circuit. After about four hours of induction, with the different combinations of IPTG (500μM) and induced supernatant, the

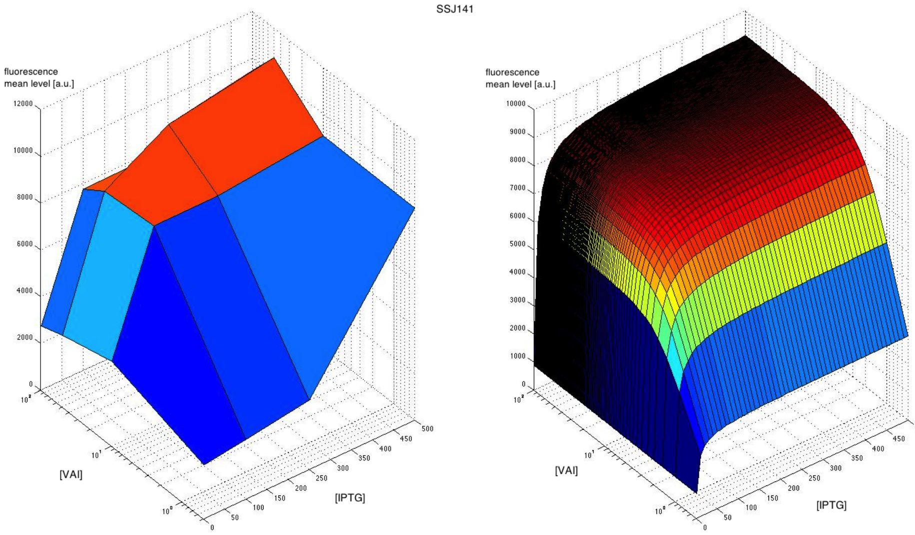


Figure 3.16: Fitting of the data of the system with the third promoter

levels of RFP were recorded with the Flow Cytometer and the results are reported in Figure 3.18. The blue column refers to the absence of IPTG and induced supernatant, the red column represent the data collected in presence of both the inducers while the cyan and yellow ones were recorded when only one transcriptional regulator was added in the environment, IPTG and AHL respectively.

It is clear that the production of fluorescent protein is considerably lower than in the previous experiments, but the signal's concentration is drastically reduced too, both due to the dilution with fresh media and its production through the Lux I induction in the Sender cells. The AND-like behaviour is maintained nevertheless and it is still possible to recognize some of the defining characteristics of the three constructs, like the higher asymmetry of the third promoter, due to the bias in its sensitivity toward AHL.
This is very important because it means that the behaviour, even if less emphasized, is not random but is determined by the activity of the hybrid promoters. Another interesting aspect is the reduction of the variability, this increase in reliability is important for the realization of a complex behaviour, like the Leader Election, where every component needs to work as expected. By comparing these experimental conditions to the previous ones it seems that this improved response is probably related to the autoinducer, in par-

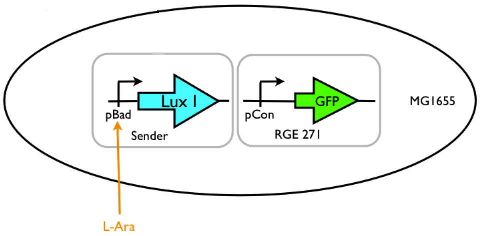


Figure 3.17: Sender Cells

ticular to its concentration or to the use of AHL produced by E. coli cells, instead of synthetic VAI. In the previous experiment one of these factors, or a combination of both, has probably caused an unwanted interaction with another cellular component causing an increase in the standard deviation of the samples. The absence or the reduction of the undesired reaction in the Supernatant assay causes an equal decrease in the side effect.

From the data in Figure 3.18 the ratio 1:1 between supernatant and fresh media is clearly the best, the fully induced level of RFP is almost two fold higher and, since the fluorescence level in the uninduced condition is quite uniform, the AND-like behaviour is more defined. This is a direct consequence of the higher concentration of AHL, due to the lower dilution of the supernatant. This even means that the concentration of the nutrients in the culture is lower but, according to the available data of bacterial density, it doesn't seem to have a significant effect on the population's growth.

3.3.2 Coculture assay: Saturating concentrations

Proven that the communication between Sender and Follower cells is possible, a coculture assay was devised. In this last part of the characterization of the Follower module the two cells strains were mixed in the same culture and then induced with IPTG and L-Arabinose to realize the intercellular communication.

The two populations were grown separately overnight and then mixed in two defined ratios, 1:3 and 1:10. These values were empirically chosen after many tries and compensate the difference in growth rate between Sender and Follower cells. The bacterial cells transformed with the Follower-Receiver module (Figure 3.5) grow slower, due to the over-expression of LacI, so it is necessary to bias the initial condition to prevent the Sender cells (Figure 3.17) from taking over the entire culture. The cells were then subdivided in four batches and introduced in environments containing different combinations of

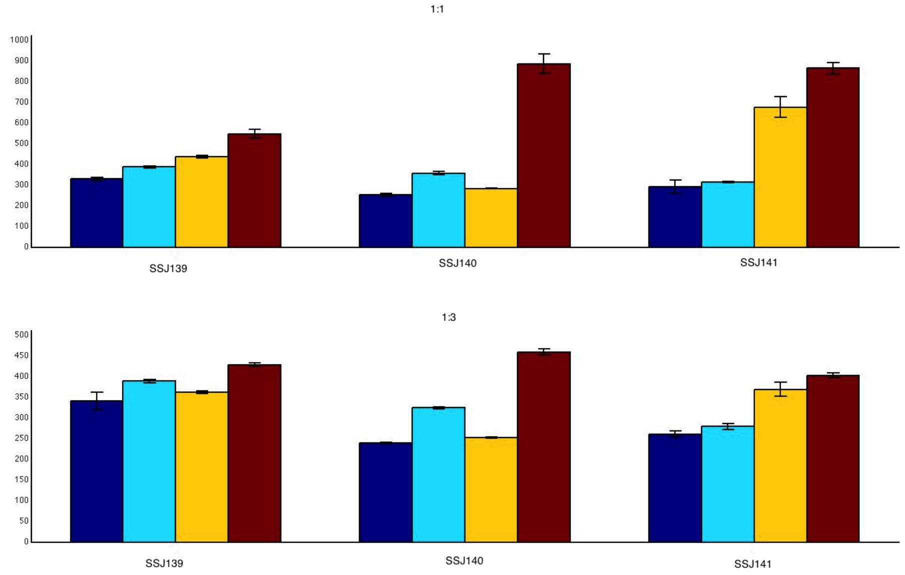


Figure 3.18: Results of the supernatant assay

the two inducers. After about 7 hours both the level of RFP and GFP were recorded with the Flow Cytometer. These data are very useful because allow to differentiate the two cell kinds and determine many interesting quantities like the final ratio between the populations.
The cells transformed with the Sender module will have a high production of green fluorescence protein, due to the presence of the RGE271 plasmid. Since, in the Follower population, the level of this protein shouldn't rise over the background level, the fluorescence distribution should be bimodal, like the one in Figure 3.19.

This representation makes it really easy to separate the Follower cells from the Senders, since the two populations are clearly distinguishable. Once the cells, classified as Followers, have been isolated it is possible to consider only the RFP levels from this group. This method has been applied to the construction of the histograms in Figure 3.20. Again the three set of columns refer to a different construct, in particular the one on the left hand side contains the prototype promoter, the one in the center the first promoter and the one on the right hand side the third promoter. The first bar reports the data obtained in absence of both IPTG and Arabinose, the level of fluorescence in cyan was recorded when in the environment was added IPTG, in the concentration of 500 μM. The yellow column represent the data measured

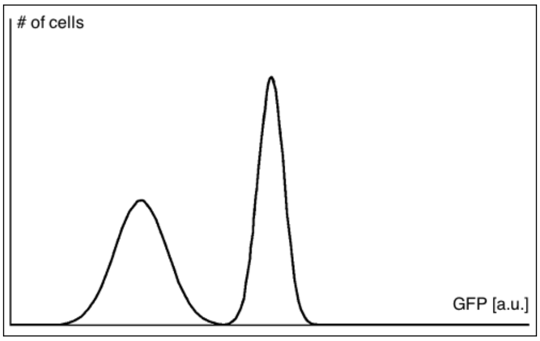


Figure 3.19: Distribution of the GFP levels for the coculture assay

only in presence of Arabinose 0.5%, the red column, as usual, was obtained adding both IPTG and Arabinose in the previous concentrations.

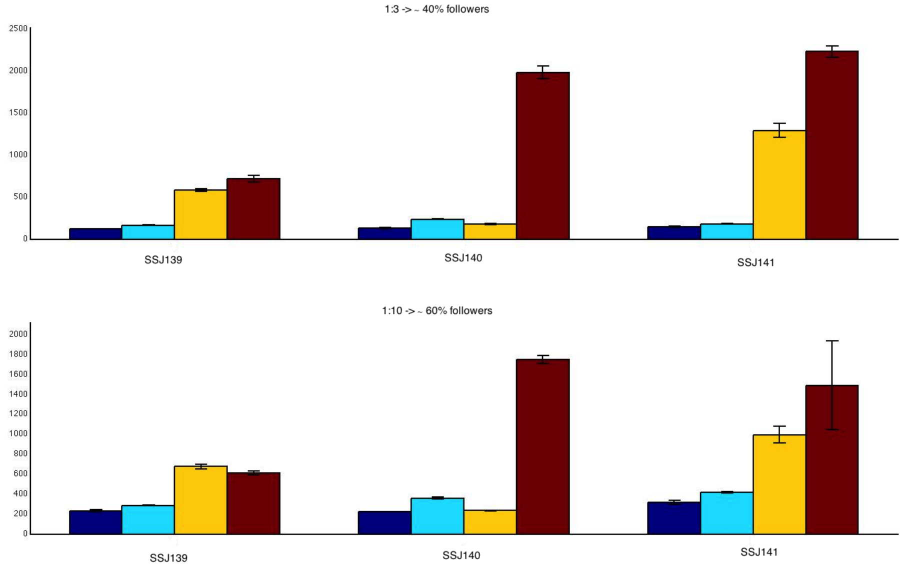


Figure 3.20: Results of the Coculture assay

As mentioned above, these data refers only to the fraction of the entire population classified as Follower, according to the level of GFP, and this fraction is 40% for an initial ratio of 1:3 and 60% for a ratio of 1:10.
In this experiment, the fully induced level of fluorescence is almost twice the one obtained in the Supernatant assay, while for the other conditions the

results are comparable, supporting the idea that the concentration of inducers is most influential only in presence of both the transcriptional factors. For this assay it is safe to say that the signal's concentration is higher, since it is continuously produced during the whole experiment, while with the previous setup the initial concentration was fixed and it decreased over time due to degradation.

Even in these histograms it is possible to identify the characteristics of the activities of the three promoters, the gain is higher for promoters 1 and 3 and the third promoter is asymmetric due to its biased sensitivity. These features are present in both the sets of data, that represent the two initial conditions, but are better defined in the upper plot, obtained with the 1:3 ratio. This is probably caused by the higher percentage of Sender cells, with respect to the other initial ratio. In this case even the signal concentration is higher, since there are more cells that can actively produce it.

Since it was possible to obtain the communication between the two kind of cells and it was found a ratio between the two populations that, for saturating concentrations of the transcription factors, ensures the correct behaviour, the assay was repeated, with different values of the inputs, to build dose response curves.

3.3.3 Coculture assay: Dose response

This assay was designed to complete the evaluation of the behaviour of the Follower module and of the signal transmission system. Building dose response curves allows to test the robustness of the behaviour to the change of the inducers concentration, in particular it was important to evaluate the reactions of the constructs to smaller values of the inputs. The stability of the response, in this case, is slightly less important since the Arabinose is an external input even for the complete circuit, and so it is possible to set its concentration arbitrarily. While the IPTG will not be present in the final design, it is just temporary used to compensate the activity of the genomic LacI.

The concentrations of IPTG that were considered for this assay, are the same used for the previous dose response: 0 μM, 100 μM, 250 μM and 500 μM, while for L-Arabinose were tested the following four values: 0%, 0.01%, 0.1% and 0.5%.

Like the previous assay the two cell's types were grown separately and mixed, in ratio 1:3, only upon induction. The levels of fluorescent proteins were measured after about 7 hours. For the coculture assay it was necessary to increase the induction time because, in this case, instead of adding the signalling molecule, it is stimulated the synthesis of the enzyme that, in turn, will catalyze the production of the autoinducer. This additional step increases the time necessary to the full induction to occur. Again, from the GFP level, was possible to isolate the Follower cells and consider only this sub-population level of RFP. The results of this operation are reported in Figure 3.21, Figure 3.22 and Figure 3.23.

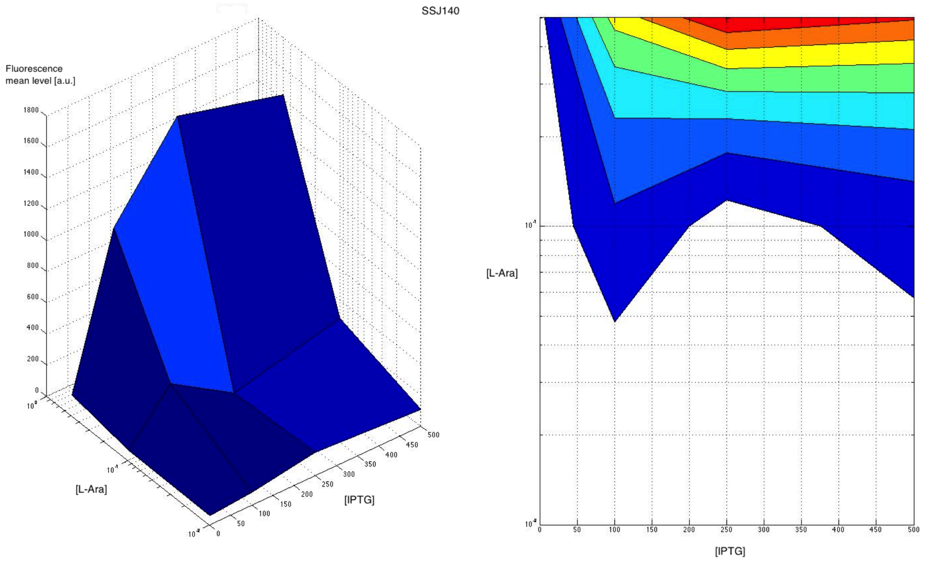


Figure 3.21: Coculture assay: dose response promoter 1

Unlike the previous dose response curves it is clear that, in order to obtain a well defined AND-gate, it is necessary to add 0.5% of L-Arabinose to the culture. As for the IPTG, even lower concentrations are able to elicit the correct behaviour, if combined with a saturating concentration of the other inducer, but the peak of the expression is obtained with the highest values tested for both the inputs. A notable exception is the first promoter, where the RFP level recorded in presence of $250\mu M$ of IPTG is equivalent to the one obtained with $500\mu M$.

In Figure 3.23, the prototype promoter presents a very unusual reaction to the addition of $250\mu M$ of IPTG. Since there is no record of a similar behaviour in the previous characterization, and the values reported in this graph are the average of two experiments, it is likely that an outlier biased this result.

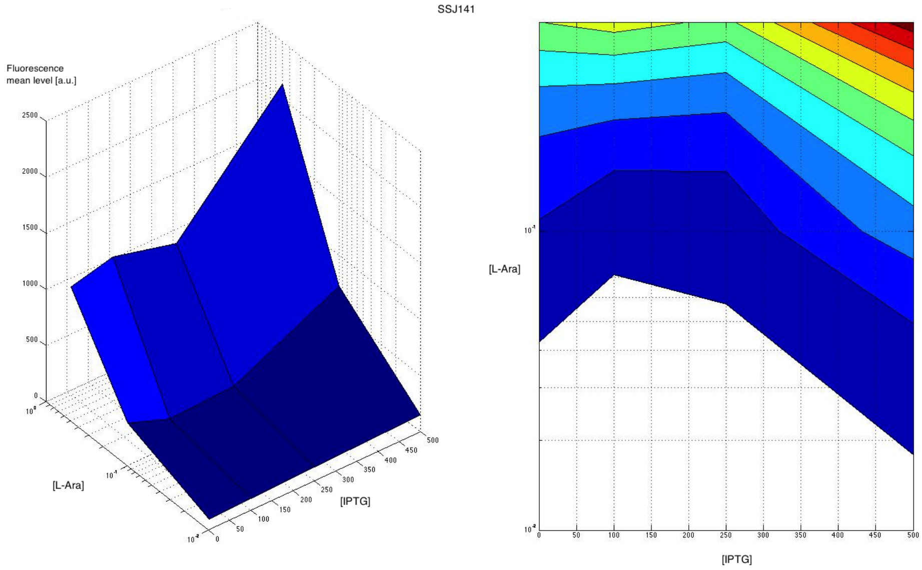


Figure 3.22: Coculture assay: dose response promoter 3

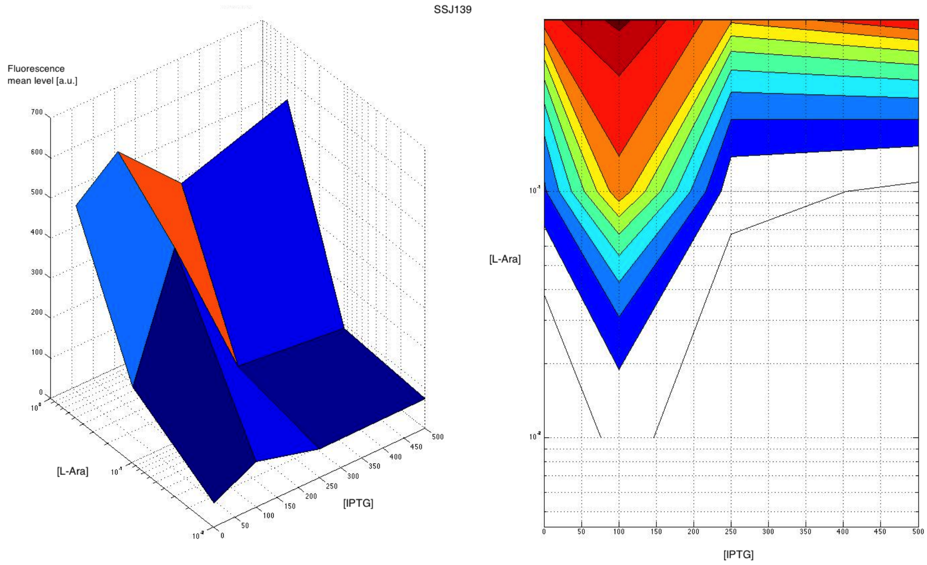


Figure 3.23: Coculture assay: dose response prototype promoter

Chapter 4

Materials and Methods

In this chapter are described the molecular biology techniques and the protocols of the assays applied for the realization of this project. In the first three sections are reported the *in vitro* technologies that were used to build the plasmids to analyse. They exploit changes of temperature, enzymes and other molecular machineries to produce the desired changes in the DNA sequence.
The last three sections report the protocols employed to test the behaviour of the constructs and to characterize them. These are obviously *in vivo* techniques and rely on the activity of fluorescent proteins and on instruments that are able to detect their signal.
The fourth section describe the process that allows the transition from the *in vitro* domain to the *in vivo*. This is possible because this technique is one of the technologies that allow to introduce the exogenous DNA into the bacterial host.

4.1 PCR, Polymerase Chain Reaction

This is a fundamental instrument in Molecular Biology because it makes it possible to amplify a DNA fragment even one million times. It was developed by Kary Mullis in 1984 and for this reason he obtained the Nobel Prize for Chemistry in 1993.
Knowing the sequence of the region to be amplified it is possible to chemically synthesize short oligonucleotides, named primers, complementary to the extremes of the sequence of interest. These fragments, together with the DNA-polymerase, realize the *in vitro* synthesis of the DNA.
The PCR reaction is schematized in Figure 4.1, it begins heating up a solution of DNA, primers, a mixture of nucleotides, DNA-polymerase and other

molecular components necessary to the synthesis of the filament.

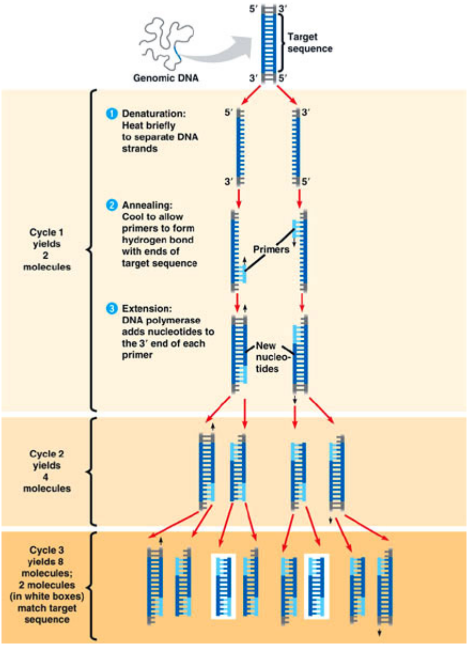

5′
3′
Target sequence
Genomic DNA
3′
5′
Cycle 1 yields 2 molecules
① Denaturation: Heat briefly to separate DNA strands
5′
3′
3′
5′
② Annealing: Cool to allow primers to form hydrogen bond with ends of target sequence
Primers
③ Extension: DNA polymerase adds nucleotides to the 3′ end of each primer
New nucleo-tides
Cycle 2 yields 4 molecules
Cycle 3 yields 8 molecules; 2 molecules (in white boxes) match target sequence

Figure 4.1: PCR reaction

A temperature increase causes the denaturation of the double strand allowing the primers to access their binding site on the DNA. In the annealing phase, a slight decrease in the temperature of the environment causes the primers to bind to the template. The successive incubations allows the DNA-polymerase to extend the primers, using the DNA single strand as a template. Repeating these steps about 20-30 times allows to amplify the DNA region of interest to a concentration detectable through electrophoresis on agarose gel.

The temperature of annealing of the primers and the extension time are parameters fundamental parameters for the success of the reaction. The former is determined by the oligonucleotide sequences and it is related to the strength of the bond between the primers and the DNA. It is important that these temperatures are in the same range for both primers, with a difference of at most 2^oC. The best way to compute this parameter is using one of the many available online applications, like the one on the Techno Scientific website, but, for an approximate calculation, it is possible to consider 2^oC for each A and T and 4^oC for every C and G of the annealing region. The extension time is set by the length of the sequence to be amplified and the speed of the polymerase. When using a Phusion DNA polymerase, it can be

roughly computed considering 1 minute for each kbp of the sequence to be synthesized plus 20-30 s to ensure the completion of the reaction.
This technique is extremely sensitive and specific and it can be realized very easily since a single instrument, named thermocycler (Figure 4.2), can be programmed to realize the different phases of this reaction (Figure 4.3).

Figure 4.2: Thermocycler

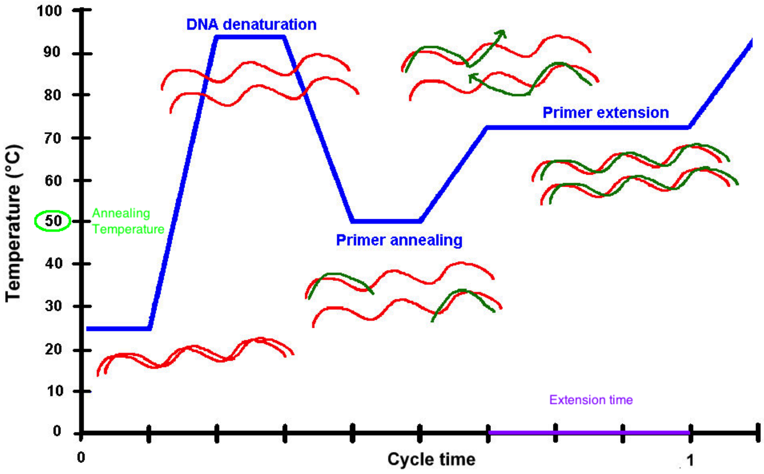

100
90
80
70
60
50
40
30
20
10
0
Temperature (°C)
DNA denaturation
Annealing Temperature
Primer annealing
Primer extension
Extension time
0
Cycle time
1

Figure 4.3: Scheme of the phases of the PCR reaction

4.2 DNA Electrophoresis

This technology allows to separate DNA fragments according to their length and, in order to do that, it takes advantage of the uniform negative charge of the DNA. The samples are loaded in an agarose gel matrix and an electric field is applied to the system, this pushes the DNA segments toward the positive pole with a strength proportional to their length. Longer sequences will experience a stronger attraction but, on the other hand, will migrate slowly because of the stronger resistance that they experience within the gel. After a given period, the fragments will be positioned in the matrix according to a gradient of lengths, making it possible to identify and isolate a desired element just by knowing its size (Figure 4.4).

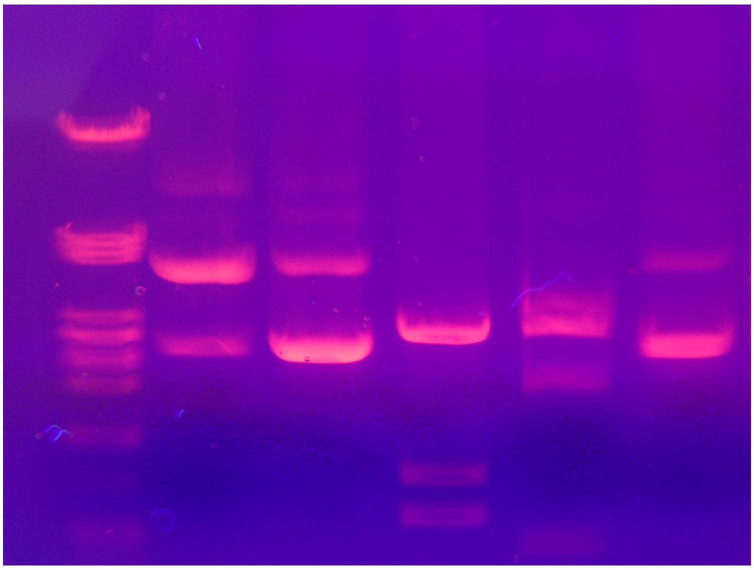


Figure 4.4: Example of Electrophoresis in Agarose gel

One of the wells, used to load the samples in the Agarose gel, contain the gel ladder. This is a mixture of DNA fragments of different lengths that is used as a reference for the other samples, see the first column of Figure 4.4 for an example. The length and the concentration of each fragment in the ladder is known so it can be used to determine, by comparison, the length of the fragments of interest.

4.3 Gibson Assembly

This *in vitro* recombination system allows to assemble and repair overlapping DNA molecules in a single isothermal step [8]. It is alternative to other ligation methods and, since it doesn't require restriction sites, it is very useful for cloning multiple inserts into a vector, rapidly constructing large DNA molecules.

The reaction is schematized in Figure 3.2. Double stranded DNA sequences, overlapping at the ends, are mixed with a cocktail of enzymes and incubated at $50^o C$ for about one hour. During this period the T5 exonuclease enzymes chews back some of the bases at the 5' end of the DNA fragments, creating the sticky ends. At this point, Phusion polymerases and Taq ligases combine the complementary sequences.

A great advantage of this technique is that the enzymes involved don't compete with each other, so they can be active simultaneously [8], making the reaction fast and easily realizable.

4.4 Electroporation

A fundamental step in the construction of genetic circuits is the introduction of the exogenous DNA into the host cells. This process is named transformation and works by temporary destabilizing the cell membrane, allowing the vector to cross it.

One technique, that can be used to increase the permeability of the membrane is electroporation. A mixture of plasmids and electrocompetent cells, bacteria that underwent a particular procedure to make them able to acquire external DNA, is placed in a plastic cuvette with two aluminum electrodes on its sides Figure 4.5.

Figure 4.5: Cuvettes for Electroporation

The cuvette is then inserted into an electroporator and an intense electric field, about 1250 V, is applied to the sample. This procedure allows the plasmids to enter the host through the pores in the membrane opened by the electrical shock. Immediately after the electroporation the cells are rescued, with fresh media, and spread on agar plates. This technique has an efficiency lower than one and so it is necessary to devise a method to identify the transformed cells. This can be realized using antibiotics, a particular antimicrobial can be added to the agar, on which the bacteria will grow and the correspondent resistance gene, inserted into the vector, will allow only the transformed cells to grow.

4.5 Induction assay

This assay was used in the characterization of the Follower module in isolation, to assess the response of the hybrid promoters to the inducers. It is performed by measuring the expression of the red fluorescent protein a few hours after adding the transcription factors in the environment. The instrument employed in these experiments is the Flow Cytometer, because it allows to obtain detailed and precise information on the promoter's behaviour, by measuring the level of protein in each cell.

Flow Cytometry is a technology widely used in synthetic biology and biological research in general, because it supports the contemporary measure of many different chemical and physical properties on thousands of cells.

The functional scheme of the instrument is represented in Figure 3.4, it is composed of a very thin channel, in which the cells flow in single file, surrounded by lasers and detectors.

A fluorescent protein is a molecule able to emit a luminous signal at a specific wavelength, when hit with a light beam at another wavelength, usually shorter. These proteins are very commonly utilized due to their simplicity of use and their efficacy: by simply placing this element under the control of the promoter of interest it will be possible to have a complete and precise report of its activity.

To characterize the Follower module, the red fluorescent protein, mRFP, was fused to the TetR protein, that is massively produced only when this module is active.

Cells, transformed with the synthetic genetic circuit in Figure 3.5, were cultured overnight in LB media, in presence of the appropriate antibiotic (Ampicillin). The different cultures, one for each combination of promoters and concentrations of inducers, were diluted 1:200 in M9CA media in the morning. After 1 hour of incubation at 37^oC, the transcription factors were added to the environment. For these assays were used IPTG and VAI, the synthetic AHL, in different concentrations (Table 5.1).

After 4 hours of incubation at 37^oC the samples were tested with the flow

Inducer	Concentrations (μM)
IPTG	0, 100, 250, 500
VAI	0, 0.5, 10, 50, 100

Table 4.1: Concentrations of inducers used in this assay

cytometer, measuring the RFP intensity over 50000 events or 50 μl. From each measurement were obtained mean, median and coefficient of variation of the fluorescence, in arbitrary units. Even the number of events and the amount of culture collected (in μl), were recorded since they allow to quantify the cell's density.

Each combination of the various concentrations of inducers was tested three times for each promoter and, from the collected data, it was possible to characterize the response of the follower module to different concentrations of inducers, both saturating and not. These values were even used to compute some indexes, fundamental for the comparison between the behaviour of this system and a logic gate (Equation 3.1, Equation 3.2, Equation 3.3, Equation 3.4).

4.6 Supernatant assay

This test was necessary to verify the theoretical possibility of realizing the intercellular communication within the population. In this case the two kinds of cells, Follower (Figure 3.5) and Sender (Figure 3.17), aren't mixed and the construct measured in the flow cytometer is the same employed in the previous assay. The difference is that, instead of inducing with VAI, was used the natural AHL, synthesized by the LuxI enzyme. This can be done because the AHL can freely diffuse through the cell membrane and so it can be found in the media.

From the Sender cells, cultured overnight in presence of 0.5% of L-Arabinose, the supernatant was extracted, it was then sterilized and the nutrients were replenished by mixing the supernatant with fresh media in defined proportions, 1:1 and 1:3.

At this point the Follower cells were induced with different combinations of IPTG, in the concentration of 500 μM, and induced supernatant. The samples were incubated at 37^{o}C for 4 hours. The subsequent measure of the statistical values of the fluorescent signal and of the population density, confirmed the ability of the Follower cells to sense the amount of AHL produced by the Senders.

Inducer	Concentrations
IPTG (μM)	0, 100, 250, 500
L-Ara (%)	0, 0.01, 0.1, 0.5

Table 4.2: Concentrations of inducers used in this assay

4.7 Coculture assay

This last assay was devised to test the system of communication realized by the LuxI and LuxR genes, through the AHL molecule, in the specific context of the Leader Election project.
The two cell strain, Follower (Figure 3.5) and Sender (Figure 3.17), were cultured separately overnight in M9CA media. In the morning the two cultures were diluted and mixed in two defined ratios, 1 Sender cell to 3 or 10 Follower cells. These values were empirically chosen so to counterbalance the bias in the growth rate. The Sender cells grow much faster than the Followers because they don't overproduce the Lac repressor, so it was necessary to adjust the proportion between the two populations to prevent the senders to overtake the entire culture. During this phase the samples were induced with different concentrations of the two transcription factors (Table 5.2).

The cells were then incubated at 37^{o}C for 7 hours, this increase in the induction time is necessary to ensure the completion of the reaction. In this system, instead of adding the intercellular signal, the production of the enzyme that catalyses the synthesis of AHL, is induced. The addition of this step makes it necessary to extend the duration of the assay.
The results collected with this assay, two replicates of each construct with every possible combination of inducers, were used to build dose response curves that confirmed the functionality of the communication system.

Chapter 5

Conclusions

5.1 From sequence to gene expression

The regulation of gene expression through devices that modulate transcription or translation, like promoters and RBSs, is relatively flexible as nucleotide sequences directly determine the specificity and efficiency of the reaction [9].
After the experimental characterization of the three hybrid promoter, an analysis of their sequences has been performed trying to identify which regions are the most important to determine the strength of the promoter and its experimental behaviour.
In the generic bacterial promoter there are fundamentally three regions that determine the strength of the binding with the RNA polymerase.
Besides the -35 and - 10 boxes, a research group of the University of Wisconsin [25] identified another sequence, upstream of the -35 box, that is able to interact with a subunit of the RNA polymerase, facilitating its interaction with the DNA. This element is named UP and its standard sequence is:
NAAAWWTWTTTTNNNAAANNNNTTGACA
Where **N** is used to represent any nucleotide, while **W** can be either an adenine or a thymine. The underlined bases, on the right hand side, form the classical -35 box sequence and their presence is fundamental to identify correctly the UP element. The sequences of the UP element of the three characterized promoters are reported in Table 5.1.

Since the distal region has always been filled with the sequence identified with Lux1 in [22], the UP element is conserved over the three versions of the hybrid promoter. If compared with the standard sequence for the UP element, it is possible to identify 10 bases that are different. Since this element is only present in bacterial promoters and it's not fundamental for

	-57 -47	-44 -41
Promoter 1	AGTACCTGTAGGATCG	ATCGTACAGG
Promoter 3	AGTACCTGTAGGATCG	ATCGTACAGG
Prototype Promoter	AGTACCTGTAGGATCG	ATCGTACAGG

Table 5.1: Up element sequences

	-35	-10
Standard	TTGACA	TATAAT
Promoter 1	TTTACT	GATAAT
Promoter 3	TTTACT	GATAAT
Prototype Promoter	TTTACT	GATACT

Table 5.2: -35 and -10 boxes

the transcription to begin, it is not clear how these mutations affect gene expression but, since all the promoters have the same sequence, it is safe to assume that they will be influenced all in the same way.
In the next phase of the analysis the sequences of the -10 and -35 boxes were considered. The comparison between the standard elements and the ones employed in the characterized constructs is illustrated in Table 5.2.

The -35 box is shared between all the promoters and it differs from the standard one in two points. The effect of these mutations is probably a reduction in the affinity between the promoter and the RNA-polymerase, since the standard sequence should be the one that the transcriptional machinery will most likely recognize and bind.
The -10 box is the same for promoters 1 and 3, it was chosen because, between the feasible alternatives, it was the one with the lowest unregulated activity. It is even more similar to the standard sequence for this regulatory element, since only the first base is different. In the prototype promoter the sequence of the -10 box differs, from the one of the other constructs, for the value of the fifth base. This substitution is probably responsible, at least in part, for the significantly lower gain of the system containing the prototype promoter, since the only difference between this construct and the one comprising the first promoter is in this region. This consideration is even supported by evidences, reported in [25], that identify the -10 box as the most important of the three consensus sequences of the bacterial promoter.

5.2 Regulated Promoters Engineering

The use of regulated promoters is a classical technique employed to regulate gene expression. In particular it is the method of election whenever it is necessary to relate the production of a certain protein to the presence, or absence, of a certain molecule. This situation is very common and encompass every application that require the system to sense the environment and respond to external stimuli, like the realization of sensors or logic gates.
Almost every regulated promoter employed in biological research has been obtained from a natural system, even if the organism in which it is used might not be the one in which it was originally present. The possibility of importing, in a bacterial strain, a regulatory element that evolved in another system is very important because, beside widening the range of possible functions and behaviours, it should minimize the interactions between the exogenous DNA and the host. Having a genetic circuit functionally independent from the host reduces the uncertainties that may arise with multiple sources of production for the same protein like, for example, both the plasmid and the genome. At the same time it is possible to use the regulatory element independently from its original function, making it more versatile. On the downside, the introduction of a completely new system can elicit unpredictable reactions in the host, that may change or disrupt the regulatory abilities of the complex.
A regulated promoter contains one or more sequences that can be recognized by specific molecules, they are usually very short, about 5-6 nucleotides, and their position is almost as important as their sequence. This concept is clearly represented in Figure 2.6 and Figure 2.7 where a wrong positioning of the Lux site makes the promoters completely insensitive to VAI.
As a general rule, a repressor will be more effective if placed close to the transcription initiation, since it generally acts preventing the RNA-polymerase from binding the DNA, while an activator will be more powerful if placed upstream of the -35 box, where it can aid the binding of the molecular machinery to the promoter, without hampering it. Some repressors can be effective even if placed tens of bases upstream of the promoters, the repressor binds this site and another regulatory region, positioned inside the promoter, this operation cause the DNA to bend and makes the promoter inaccessible [15]. An example of this is reported in Figure 5.1 and it refers to the natural Lactose digesting system.

These considerations, supported by the experimental data presented before, confirm some of the rules reported in [22]. Comparing the behaviour of the construct with promoter 1 and the one with promoter 3 it is even possible to notice that the addition of the second Lux site has the effect of slightly reducing the protein production, obtained upon induction. The binding of

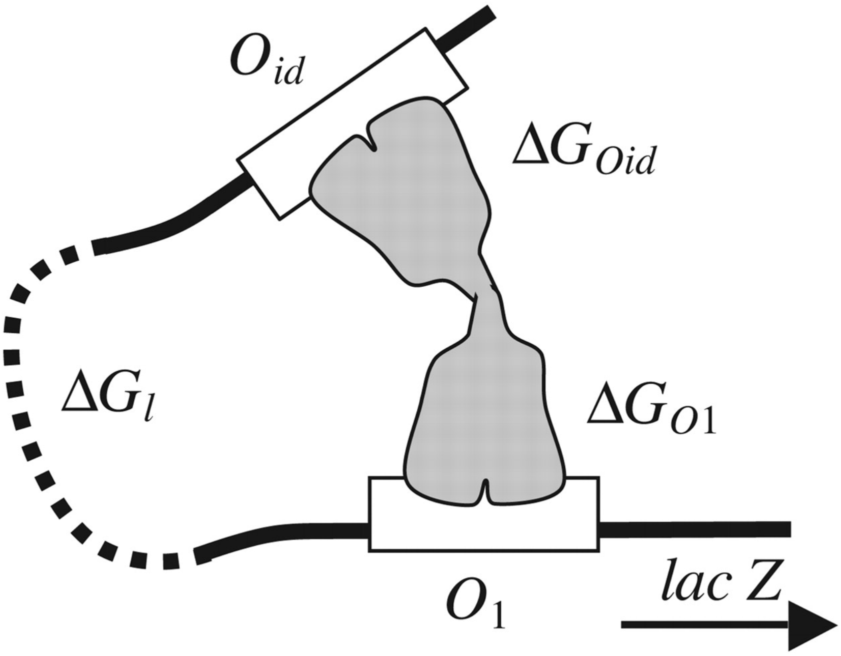


Figure 5.1: A transcription factor can bind two operators sites at the same time

the activator in core probably interferes with the polymerase binding, instead of facilitating it. The analysis performed on the behaviour of the first promoter confirms another rule, of the ones defined in [22], the proximity of the regulatory sites makes the interaction more competitive and so, more AND like (Figure 3.9).

Observing the data relative to promoter three it is clear that, in this case, the activation dominates the repression, in contrast with what reported in [22]. Since only this construct exhibit this behaviour, it is not clear if it's a peculiarity of this promoter or if the directive in the paper is imprecise. Certainly when an activator site and a repressor site coexist in the same promoter, one will dominate on the other. This will increase the asymmetry of the promoter but, probably, both the activator and the repressor can be the predominant, depending on the strength of the consensus sequences employed. Another factor worth considering is that, in the Follower cells, the LacI protein was produced by the genome, while the LuxR gene was on a low/medium copy number plasmid. Even if the strain chosen for these experiments over-expresses the Lac repressor, it isn't possible to exclude that LuxR may be over-produced, with respect to LacI, and that its higher concentration may bias the effect of the transcription factors. The increased symmetry of the first promoter is probably caused by the presence of the additional Lux sensitive sequence, their combined effect is probably able to balance the action of the two inducers.

The process of promoters engineering has greatly benefited from the increasing completeness and standardization of the characterization of the regulatory systems.

In particular this work demonstrates that, following simple design principles, it is possible to obtain promoters with the desired characteristics. But the complete and precise prediction of the behaviour of these elements is still very far since even very similar sequences, like the ones here presented, have significantly different responses to the same experimental conditions.

The construction and screening of big libraries of components can be really helpful in the determination of heuristics and general design criteria, but they don't fully explain the molecular mechanisms that lead to a certain behaviour. This lack of knowledge drastically limits the ability of engineers to rationally design promoters and genetic circuits, increasing the time necessary to tune their behaviours and the chance of obtaining unpredicted reactions.

Deeper analysis of natural and engineered systems coupled with computational modelling and simulation will be key elements in the realization of complex, reliable genetic circuits with real life applications.

5.3 Complex circuits realization

Synthetic Biology has had a great development in the last few years, many applications have been presented and significant advances has been made in many aspects of the field like design, modelling and experimental characterization.

But, as Luis Serrano ably says, "We are still like the Wright Brothers, putting pieces of wood and paper together" [14]. There are still many unsolved problems and many applications are still beyond the verge of realizability due to their complexity and the lack of knowledge of the basic components.

The biological world is incredibly complicated and heterogeneous, it slowly evolved and molded itself to solve problems and increase its fitness for survival. The mankind began to shape this world very early in its history, domesticating plants and animals and selectively breeding them to enhance useful features.

But understanding how a long series of consecutive changes and modifications lead to the organisms that now inhabit the earth is really complex and until recently, the nature of most of the information about the functioning of biological systems and processes has been qualitative or semi-qualitative. The introduction of engineering principles and techniques in this field is of prime importance in order to increase the knowledge of the basic phenomena and really becoming able to modify them purposely.

As previously described, the pivotal points of classical engineering, standardization, decoupling and abstraction are fundamental, their application allows the development of simple systems, completely characterized and with defined functions and features that can be used and combined as the electrical components of the natural world.

But there are aspects that these principles are unable to encompass, due to their rigidity and strict categorization. The organic matter is prone to changes, might react unexpectedly to external stimuli and for every stated rule, there are tens of exceptions. This calls for a redefinition of the concepts of standardization, abstraction and decoupling for the biological world.

Developing standard modules and techniques for Synthetic Biology applications will surely involve the ideation of new experiments and technologies to obtain quantitative data, independent of the particular instrument and attentively measured with very clear and detailed protocols. Up until now many efforts have been focused on developing standard techniques for assembling genetic circuits, but none of the proposed approaches has yet imposed itself. Simultaneously, the DNA synthesis is becoming really cheap.

Soon many laboratories will find convenient to synthesize most of their constructs, drastically reducing the importance of the standardization of the assembly techniques. The standardization of the results, on the other hand, will be very useful because it will make the data obtained by different laboratories readily comparable and it will be easier for the scientific community to develop well characterized, plug and play, systems and concentrate the efforts toward common goals.

Decoupling is the principle that has probably been exploited the most in the work here presented. The complex Leader Election circuit has been subdivided in smaller, more manageable modules that have been built independently and tested both in isolation and in combination. In this particular area the development of accurate computational models will be crucial since it will allow the identification of the critical points and major issues of the particular circuit. This knowledge will guide in the choice of the assays necessary to fully characterize the behaviour of each genetic element, while providing suggestions for obtaining the desired response. Combining this strategy with a standardized and quantitative experimental characterization, will help to identify the most suitable modules to use in the final system while reducing possible unwanted interactions. All these methodologies will combine leading

to a reduction of the time necessary for the realization of complex circuits.
In the previous description two kinds of abstraction have been introduced for their importance in the biological engineering context. One of them is strongly correlated to the decoupling principle and states that each module should be simple to model and easily used in combination. This typology of abstraction incorporates even the Occam principle, since it suggests to prefer the simplest solution, that can realize the desired behaviour.
In the other are described abstraction hierarchies, these structures should allow the researchers to focus only on one determined level of complexity without regard for the others. In order to fully realize this concept is fundamental to limit the informational exchange between the different hierarchies. This entails that inputs and outputs of each level should be defined very clearly. Currently this form of abstraction isn't achievable due to the insufficient knowledge of the genetic circuits, the hosts and their interactions. Abandoning the organisms that are most currently used and focus on more artificial life forms, like the minimal cell, might be a step toward the realization of this goal. This organism doesn't have a clear definition yet but, removing every element and function that is not fundamental for the survival of the system should minimize the interactions between the host and the exogenous DNA. Furthermore, the development of this organism would probably lead to a deeper knowledge of bacterial cells' physiology and functioning, thus clarifying the informational exchanges necessary between each level of complexity.

Ringraziamenti

La realizzazione di questa tesi non é stata solamente la culminazione del mio periodo di studi universitario, ma é stata soprattutto un'esperienza di vita che mi ha portato lontano da casa, a conoscere nuove persone, una nuova cultura e a crescere sia dal punto di vista delle competenze tecniche e teoriche che come persona.
A questo proposito vorrei cogliere l'occasione di ringraziare la mia famiglia che mi ha appoggiata in questa scelta e supportata, anche economicamente, durante tutto il mio stage all'Universitá di Washington.
Inoltre un enorme ringraziamento va a tutto l' S.O.S. laboratory e in particolare al Professor Eric Klavins e a Shelly Jang che mi hanno seguita durante la realizzazione del mio progetto e mi hanno insegnato moltissimo, anche ad approcciare e risolvere indipendentemente i problemi che frequentemente emergono nell'attivitá di ricerca.
Grazie anche al mio relatore, Professor Stefano Severi al mio correlatore, Dott. Emanuele Giordano e alla Dott.ssa Francesca Ceroni per il loro aiuto e la loro disponibilitá.
Un ringraziamento particolare va al mio ragazzo, Andrea, che ha condiviso con me questa esperienza e il percorso universitario in generale. Grazie per avermi supportata e sopportata in tutti gli alti e i bassi di questi anni.
Infine vorrei rimarcare come questo lavoro sia il frutto del lavoro di molte persone che mi hanno fatto conoscere la Biologia Sintetica, mi hanno ispirata e hanno alimentato la mia passione per la scienza, l'ingegneria e la ricerca.

Acknowledgements

The preparation of this thesis has been not only culmination of my University career, but even a life experience that brought me far from home and made me met new people, know a new culture and improve myself both on the professional and personal level.
For this reason I would like to thank my family that supported me, even economically, during my stage at the University of Washington.
Furthermore a big thanks goes to the members of the S.O.S laboratory, and in particular to Professor Eric Klavins and Shelly Jang that helped me during the realization of this project and from whom I've learned a lot, even to independently approach and solve the problems that frequently arise in scientific research.
Thanks even to my advisers Professor Stefano Severi and Dr. Emanuele Giordano and to PhD Francesca Ceroni for their help and willingness.
Additionally thanks to Andrea, that shared with me this experience and the university career in general. Thank you for support, and sometime bear, me over the ups and downs of these years.
In the end I would like to stress how this work is the result of the work of many people that introduced me to Synthetic Biology, inspired me and nurtured my passion for science, engineering and research.

Bibliography

[1] Meaning and structure in biology and physics: Some outstanding questions.

[2] Registry of standard biological parts. `http://partsregistry.org/Main_Page`.

[3] *Chimica, biochimica e biologia applicata*. 2009.

[4] D. A. Fletcher A. P. Arkin. Cheap, fast and somewhat in control. *Genome Biol.*, 2006.

[5] C.A. Voigt A. Tamsir, J.J. Tabor. Robust multicellular computing using genetically encoded nor gates and chemical "wires". *Nature*, 2012.

[6] W.T. Astbury. Molecular biology or ultrastructural biology? *Nature*, 1961.

[7] B.L. Bassler. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current Opinion in Microbiology*, 1999.

[8] R.Y. Chuang J.C. Venter C.A. Hutchinson H.O. Smith D.G. Gibson, L. Young. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature Chemical Biology*, 2009.

[9] D.K. Karig R. Weiss E. Andrianoantoandro, S. Basu. Synthetic biology: new engineering rules for an emerging discipline. *Molecular System Biology*, 2006.

[10] D. Endy. Foundations for engineering biology. *Nature*, 2005.

[11] C. Sanchez J. Monod S. Eldestein F. Jacob, D. Perrin. The operon: a group of genes with expression coordinated by an operator. *C.R. Acad Sci. Paris*, 1960.

[12] C.A. Voigt H.M. Salis, E.A. Mirsky. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 2009.

[13] S. Jang and E. Klavins. Manuscript in preparation. 2013.

[14] R. Kwok. Five hard truths for synthetic biology. *Nature*, 2010.

[15] H.G. Garcia U. Gerland T. Hwa J. Kondev T. Kuhlman R. Phillips L. Bintu, N.E. Buchler. Trascriptional regulation by the numbers: applications. *Current Opinion in Genetics and Development*, 2005.

[16] R. Weiss F.H. Arnold L. You, R.S. Cox. Programmed population control by cell-cell communication and regulated killing. *Nature*, 2004.

[17] Salis Lab. Rbs calculator. `https://salis.psu.edu/software/forward`.

[18] B.L. Bassler M.B. Miller. Quorum sensing in bacteria. *Annual Reviews in Microbiology*, 2001.

[19] R. Weiss P.E.M. Purnick. The second wave of synthetic biology: from modules to systems. *Nature reviews Molecular Cell biology*, 2009.

[20] B. Bowerman R. Li. Symmetry breaking in biology. *Cold Spring Harbor Perspective in Biology*, 2010.

[21] E. Klavins R.G Egbert. Fine-tuning gene networks using simple sequence repeats. *PNAS*, 2012.

[22] M.B. Elowitz R.S. Cox, M.G. Surette. Programming gene expression with combinatorial promoters. *Molecular System biology*, 2007.

[23] C.H. Collins F.H. Arnold R. Weiss S. Basu, Y. Gerchman. A synthetic multicellular system for programmed pattern formation. *Nature*, 2005.

[24] Bismita Srichandan. Leader election. `http://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDYQFjAB&url=http%3A%2F%2Fwww.cs.gsu.edu%2F~cscyqz%2Fcourses%2Faos%2Fslides08%2Fch4.6-Fall08.ppt&ei=Tv2xUNbiC8mi4gSezIHgCA&usg=AFQjCNHwy8Or45B3ultAWb4PFfxvF9u4aQ`, Fall 2008.

[25] W. Ross R.L. Gourse S.T.Estrem, T. Gaal. Identification of an up element consensus sequence for bacterial promoters. *Proceedings of the National Accademy of Science*, 1998.

[26] L. Tsimring J. Hasty T. Danino, O. Mondragon-Palomino. A synchronized quorum genetic clock. *Nature*, 2010.

[27] J.J. Collins T.K. Lu, A.S. Khalil. Next generation synthetic gene networks. *Nature Biotechnology*, 2009.

[28] Herk van der Belt. Playing god in frankenstein's footsteps: Synthetic biology and the meaning of life. *Nanoethics*, 2009.

[29] E.P. Greenberg W.C.Fuqua, S.C.Winans. Quorum sensing in bacteria: the luxr-luxi family of cell density-responsive transcriptional regulators. *Journal of Bacteriology*, 1994.

[30] Greenberg W.C.Fuqua, E.P. Signalling: Listening in on bacteria: acyl-homoserine lactone signalling. *Nat Rev Mol Cell Biol*, 2002.