

# UNIVERSITA' DEGLI STUDI DI PAVIA

FACOLTA' DI INGEGNERIA  
DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

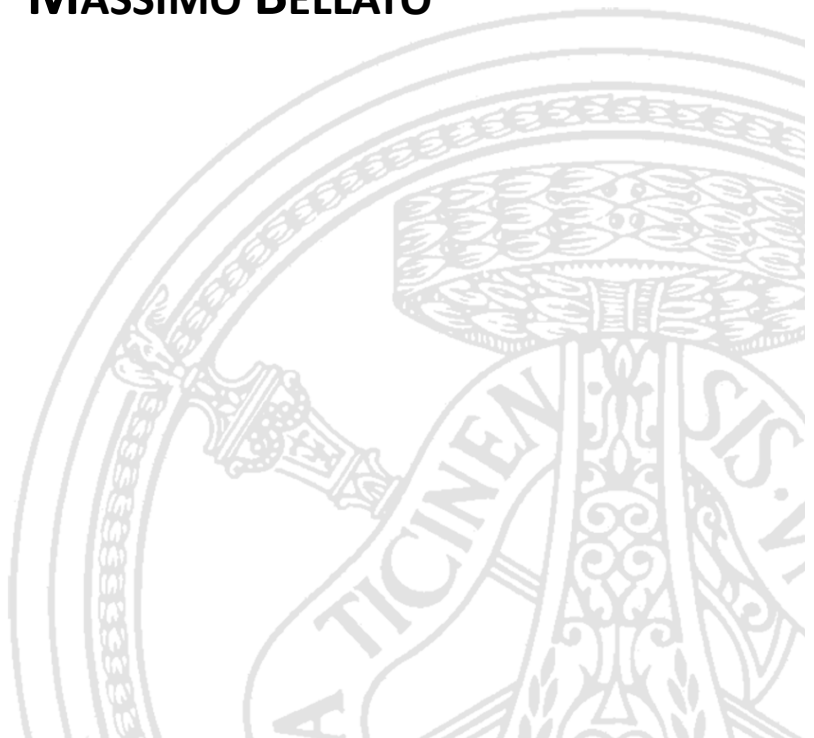
DOTTORATO DI RICERCA IN BIOINGEGNERIA E BIOINFORMATICA  
XXXI CICLO - 2018

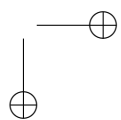
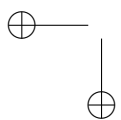
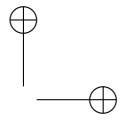
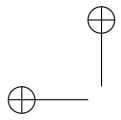
## OVERCOMING METABOLIC BURDEN IN SYNTHETIC BIOLOGY: A CRISPR INTERFERENCE APPROACH

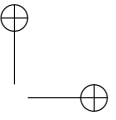
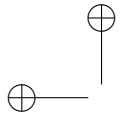
PhD Thesis by  
**MASSIMO BELLATO**

**Advisors:**  
Prof. Paolo Magni  
Prof. Lorenzo Pasotti

**PhD Program Chair:**  
Prof. Riccardo Bellazzi

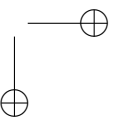
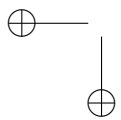


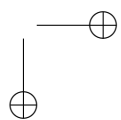
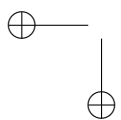
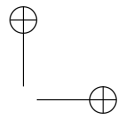
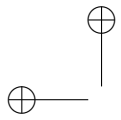


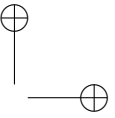
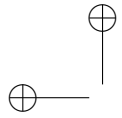


*Louis Pasteur's theory of germs  
is ridiculous fiction.*

Pierre Pachet,  
- Professor of Physiology at Toulouse, 1872



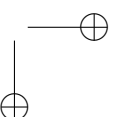
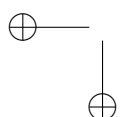


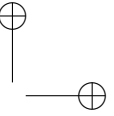
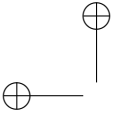


## Abstract (English)

In the last decade, Synthetic Biology is gaining an increasingly important role in the scientific community and dedicated research centers are rising all over the world. This discipline, which lies between life science and engineering, introduced the principles of abstraction, modularity and standardization in the biology world; nowadays, the application of these engineering principles is allowing the design of complex biological systems to program living cells, realizing all sorts of desired function in many fields, from agriculture to health. These systems consist of DNA sequences, rationally combined to program the genetic instructions for cell behavior customization. This is possible assuming that each part behaves as a biological “brick” for the design of complex genetic programs through functional building blocks; each DNA module undergoes an extensive characterization in order to provide documentation on its functioning enabling the rational design of complex circuits on the basis of the information available for each individual sub-part composing the whole system.

Mathematical modeling accompanies all the design procedure as a tool to describe the behavior of each single genetic module, in a bottom-up fashion that should allow the prediction of more complex systems





---

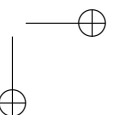
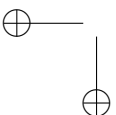
obtained by the interconnection of pre-characterized parts. However, many unpredictability sources hamper the ideally rational design of those synthetic genetic devices, mainly due to the tangled context-dependency behavior of those parts once placed into an intrinsically complex biological living system.

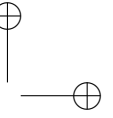
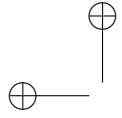
Among others, the finite amount of translational resources in prokaryotic cells leads to an effect called *metabolic burden*, as a result of which hidden interactions between protein synthesis rates arise, leading to unexpected counterintuitive behaviors.

To face this issue in rational design of synthetic genetic circuits hosted by bacterial cells, two actions have been proposed in this study: firstly, a recently proposed mathematical modeling solution that included a description of the metabolic load exerted by the expression of recombinant genes have been applied on a case study, highlighting its worth of use and working boundaries; second, a CRISPR interference-based architecture have been developed to be used as an alternative to high resource usage transcriptional protein regulators, studying the underlying mechanism in several circuitual configurations and optimizing each forming part in order to achieve the desired specifications.

In Chapter 1, an introduction on synthetic biology is presented; rational design paradigm and hurdle given by metabolic burden are described and the role of mathematical modeling discussed. In the second part, a brief overview on CRISPR technology and the overall aim of the study are reported.

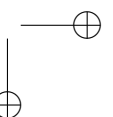
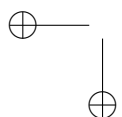
In Chapter 2, a case study evaluating the use of mathematical modeling to properly include metabolic burden in rational design of a set of transcriptional regulator cascades is reported. Firstly, the circuits and expected behavior are introduced, along with the discussion about experimental data, dissenting from what initially predicted. Secondly, the comparison between the use of a classical Hill equation-based model and an improved version that explicitly consider the translational load exerted by the expression of recombinant genes is reported.

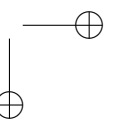
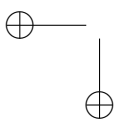
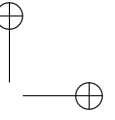
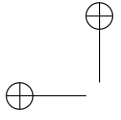




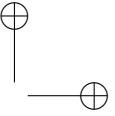
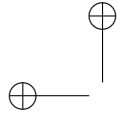
---

In Chapter 3, the design and deep characterization of a BioBrick<sup>TM</sup>-compatible CRISPR interference-based repression set of modules is shown; expression optimization of the molecular players is reported and its usability as a low-burden alternative is demonstrated with experimental data and mathematical modeling. Working boundaries, peculiar aspects and rooms for improvements are then highlighted. In Chapter 4, preliminary studies aimed to improve the CRISPR interference system are reported and some of its context-dependencies are highlighted. Effects on repression efficiency due to alteration in the sequence of the RNA molecules addressing the CRISPR machinery to the desired target are discussed; evaluation of problems and opportunities related to the expression of more of this RNA guides are then highlighted. Lastly, an example of behavior of the system in presence of a competitor transcriptional regulator is reported. In Chapter 5 the overall conclusions of this thesis work are drawn.





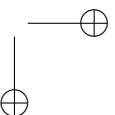
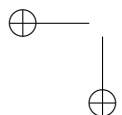


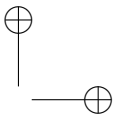
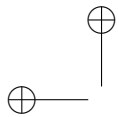


## Abstract (Italian)

Nell'ultimo decennio, la Biologia Sintetica sta assumendo un ruolo sempre più importante nella comunità scientifica e numerosi centri dedicati stanno sorgendo in tutto il mondo. Questa disciplina, a cavallo tra ingegneria e scienze della vita, ha introdotto i principi d'astrazione, modularità e standardizzazione nel mondo della biologia; oggigiorno, l'applicazione di questi principi ingegneristici sta consentendo la progettazione di sistemi biologici standardizzati per programmare cellule viventi, implementando le più varie funzionalità in molti campi, dall'agricoltura alla sanità. Questi sistemi consistono in sequenze di DNA, combinate razionalmente componendo istruzioni genetiche per personalizzare il comportamento della cellula. Questo è possibile solo assumendo che ogni parte si comporti come un "mattoncino" biologico per la progettazione di programmi genetici complessi; ogni parte è sottoposta ad un esteso processo di caratterizzazione, al fine di fornire una documentazione che permetta la progettazione razionale di circuiti complessi a partire dalle informazioni base disponibili su ogni sua sotto-parte costituente l'intero sistema.

La modellizzazione matematica accompagna le procedure di progettazione in quanto strumento utile a descrivere il comportamento di ogni





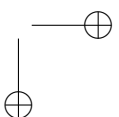
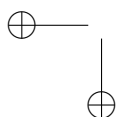
---

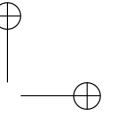
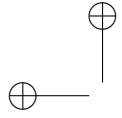
singolo modulo genetico, secondo un approccio bottom-up, consentendo successivamente la predizione di sistemi più complessi ottenuti interconnettendo parti precedentemente caratterizzate. Tuttavia, molte fonti di imprevedibilità ostacolano la progettazione razionale ideale di questi sistemi, principalmente a causa degli intricati comportamenti contesto-dipendenti che tali parti dimostrano una volta poste all'interno di un sistema vivente intrinsecamente complesso.

Tra queste, il limitato ammontare di risorse trascrizionali nella cellula procariote porta ad un effetto chiamato “fardello metabolico”, a causa del quale emergono interazioni nascoste tra i livelli di sintesi delle proteine espresse, causando comportamenti inattesi o controintuitivi. Per affrontare questo problema nella progettazione razionale di circuiti genetici sintetici ospitati da cellula batteriche, in questo lavoro di tesi sono state proposte due azioni:

1. Su un caso di studio, è stata adottata una soluzione recentemente pubblicata basata sull'impiego di modelli matematici che includano una descrizione del carico metabolico esercitato dall'espressione di geni ricombinanti, evidenziandone l'utilità ed i limiti;
2. è stata sviluppata un'architettura genetica basata sulla *CRISPR interference*, pensata come alternativa all'impiego di regolatori trascrizionali ad alto carico metabolico. Sono stati studiati meccanismo sottostante con diverse configurazioni circuitali ed ottimizzazione di ogni parte costituente il sistema, in modo da sottostare alle specifiche richieste.

Nel Capitolo 1, viene presentata un'introduzione sulla biologia sintetica; vengono descritti progettazione razionale ed il problema del carico metabolico, assieme alla discussione del ruolo della modellizzazione matematica. Nella seconda parte, sono riportate una breve panoramica sulla tecnologia CRISPR e lo scopo generale dello studio.





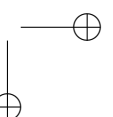
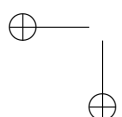
---

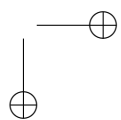
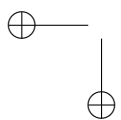
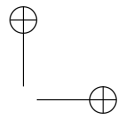
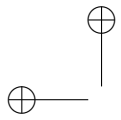
Nel Capitolo 2, è riportato un caso di studio relativo alla valutazione dell'impiego di modelli matematici per includere il carico metabolico nella progettazione razionale di un insieme di cascate di regolatori trascrizionali. Inizialmente, vengono introdotti il circuito e la risposta attesa, assieme alla discussione relativa ai dati sperimentali. Successivamente viene riportato un confronto tra un modello classico basato su equazioni di Hill ed una sua versione arricchita dalla descrizione matematica del carico esercitato dall'espressione dei geni ricombinanti.

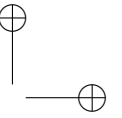
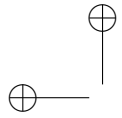
Nel Capitolo 3, sono mostrate la progettazione e la caratterizzazione di un insieme di moduli di repressione BioBrick<sup>TM</sup> compatibili basati sulla *CRISPR interference*; ne viene riportata l'ottimizzazione dell'espressione degli attori molecolari e ne viene dimostrata l'efficienza come alternativa a basso carico metabolico attraverso dati sperimentali e modellizzazione matematica.

Nel Capitolo 4, sono riportati degli studi preliminari atti al miglioramento di sistemi basati su CRISPR interference e ne vengono evidenziate le sue dipendenze contestuali. Vengono studiati gli effetti sull'efficienza di repressione causati dall'alterazione della sequenza delle molecole di RNA pilotanti il macchinario molecolare al sito desiderato; vengono inoltre valutati problemi ed opportunità relativi all'espressione di più d'una di queste guide ribonucleotidiche. Infine, viene riportato un esempio del comportamento del sistema in presenza di un regolatore trascrizionale competitivo.

Nel Capitolo 5 sono riportate le conclusioni di questo lavoro.

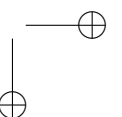
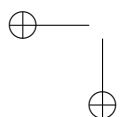






# Contents

<b>Table of contents</b>	<b>viii</b>
<b>1 Background</b>	<b>1</b>
1.1 Synthetic Biology . . . . .	1
1.1.1 Rational design . . . . .	2
1.1.2 Unpredictability sources . . . . .	6
1.2 Metabolic Burden in Rational Design . . . . .	7
1.3 CRISPR/Cas9 system . . . . .	9
1.3.1 Mechanism and first usages . . . . .	9
1.3.2 CRISPR interference and derivatives . . . . .	13
1.4 Project Idea and bigger picture . . . . .	15
<b>2 Metabolic Burden Modeling using transcriptional cas-</b>	<b>17</b>
<b>    cades as case study</b>	
2.1 In-vivo characterization . . . . .	18
2.1.1 Circuits description . . . . .	18
2.1.2 Circuits characterization . . . . .	20
2.1.3 Data processing . . . . .	21
2.2 Models description . . . . .	23





## CONTENTS

---

2.2.1	No-burden model (NBM) . . . . .	23
2.2.2	Burden model (BM) . . . . .	25
2.3	Application of the models . . . . .	28
2.3.1	Data overview . . . . .	28
2.3.2	Model fitting and analysis . . . . .	30
2.3.3	Circuit predictability with no-burden model . . . . .	32
2.3.4	Circuit predictability with the Burden Model . . . . .	39
2.3.5	Fitting using all the available experimental data . . . . .	45
2.3.6	Fixing non-functional cascades via rational design . . . . .	46
<b>3</b>	<b>Engineering low-burden synthetic circuits via CRISPRi</b> . . . . .	<b>51</b>
3.1	dCas9 expression cassette . . . . .	52
3.1.1	dCas9 toxicity . . . . .	52
3.1.2	Repression efficiency . . . . .	54
3.2	sgRNA expression cassette design . . . . .	63
3.2.1	Flow cytometry analysis of IPTG-induced system . . . . .	67
3.2.2	TSS Optimization . . . . .	70
3.3	Characterization of inducible sgRNA systems . . . . .	73
3.3.1	Constitutive dCas9 expression: MC targets . . . . .	75
3.3.2	Constitutive dCas9 expression: HC targets . . . . .	77
3.3.3	IPTG-driven sgRNA: $P_{LtetO1}$ target analysis . . . . .	79
3.3.4	CRISPRi Mathematical model . . . . .	82
3.3.5	Overall observations . . . . .	89
3.4	Dependence on promoter strength . . . . .	90
3.4.1	$P_{luxRep}$ library with MC target . . . . .	92
3.4.2	$P_{luxRep}$ library with HC target . . . . .	92
<b>4</b>	<b>Further investigations</b> . . . . .	<b>97</b>
4.1	CRISPRi-Transcription Factor interference . . . . .	97
4.2	Approaches to improve tunability and repression range . . . . .	101
4.2.1	Altered sgRNAs . . . . .	101
4.2.2	CRISPRi competition . . . . .	107



---

## CONTENTS

<b>5 Discussion and conclusions</b>	<b>117</b>
<b>Appendix</b>	<b>128</b>
<b>A Successful rational design example: Synthetic close-loop controller</b>	<b>129</b>
A.1 Introduction . . . . .	129
A.2 Close-loop circuit design . . . . .	132
A.3 Materials and Methods . . . . .	141
A.4 Results . . . . .	147
A.5 Conclusion . . . . .	155
A.6 Supplementary Notes . . . . .	158
<b>B Supplementary information for Chapter 2</b>	<b>183</b>
B.1 Strains, reagents and cloning . . . . .	183
B.2 Data overview . . . . .	184
B.3 Additional fittings and sensitivity analysis . . . . .	194
B.4 Evolutionary stability analysis . . . . .	205
B.5 Degradation tag sensitivity analysis . . . . .	208
B.6 GFP autofluorescence analysis . . . . .	209
B.7 Tables of parts and parameters data . . . . .	210
B.8 Evidence of metabolic burden using a different monitor cassette . . . . .	212
<b>C Supplementary information for Chapter 3</b>	<b>215</b>
C.1 List of constructs and estimated parameters . . . . .	215
C.2 Materials and reagents . . . . .	220
C.3 Cloning . . . . .	221
C.4 Microscopic Images . . . . .	226
C.5 Flow Cytometry . . . . .	227
C.6 Blanking of the auto-fluorescence in the GFP signal . . . . .	228
C.7 Fitting and prediction from Section 3.3.4 . . . . .	230

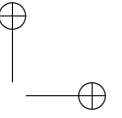
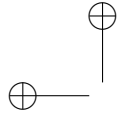


## CONTENTS

---

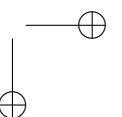
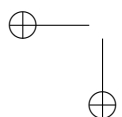
<b>Bibliography</b>	<b>249</b>
<b>List of publications</b>	<b>267</b>





# List of Figures

1.1	Engineering cycle. . . . .	4
1.2	Bacterial Type II CRISPR/Cas Immune Response . . .	10
1.3	Design of the sgRNA with Tetraloop . . . . .	11
1.4	Transcription initiation block . . . . .	14
2.1	Collection of circuits analyzed in this study. . . . .	19
2.2	Overall prediction performance by the two models analyzed in this work. . . . .	33
2.3	Measured output of the circuits at all the HSL concentrations tested in this work plotted against the predicted values. . . . .	34
2.4	NBM prediction of the measured HSL-dependent output in all the test set circuits without Monitor cassette. . . . .	36
2.5	BM prediction of the measured HSL-dependent RFP output in all the test set circuits with Monitor cassette. . . . .	43
2.6	BM prediction of the measured HSL-dependent GFP output in all the test set circuits with Monitor cassette. . . . .	44
2.7	Re-designed circuit variants schema . . . . .	48



## LIST OF FIGURES

---

2.8	Analysis via NBM of the measured HSL-dependent output in the re-designed circuit variants, without Monitor cassette. . . . .	49
3.1	Toxicity of dCas9 . . . . .	53
3.2	Morphological Analysis of dCas9 Expressing Cells . . . . .	55
3.3	CRISPRi with inducible dCas9 and constitutive sgRNA configuration . . . . .	56
3.4	Inducible dCas9 Targeting Medium Copy $P_{LtetO1}$ . . . . .	57
3.5	Inducible dCas9 Targeting Medium Copy $P_{LlacO1}$ . . . . .	59
3.6	Inducible dCas9 Targeting High Copy $P_{LtetO1}$ . . . . .	61
3.7	Inducible dCas9 Targeting High Copy $P_{LlacO1}$ . . . . .	62
3.8	Non-optimized Luciferase-inspired Inducible Knob Characterization . . . . .	65
3.9	Non-optimized Lac Operon-inspired Inducible Knob Characterization . . . . .	66
3.10	Single Cell Analysis of IPTG-Inducible Cassette . . . . .	69
3.11	Optimized HSL-Inducible Promoter Characterization . . . . .	71
3.12	Optimized IPTG-inducible Promoter Characterization . . . . .	72
3.13	Schema of the model system used in this study. . . . .	74
3.14	HSL Induction Cassette Targeting $P_{LtetO1}$ in Medium Copy . . . . .	75
3.15	HSL Induction Cassette Targeting $P_{LlacO1}$ in Medium Copy . . . . .	76
3.16	HSL Induction Cassette Targeting $P_{LtetO1}$ in High Copy . . . . .	77
3.17	HSL Induction Cassette Targeting $P_{LlacO1}$ in High Copy . . . . .	78
3.18	IPTG Inducible Cassette Targeting $P_{LtetO1}$ in Medium Copy . . . . .	80
3.19	IPTG Induction Cassette Targeting $P_{LtetO1}$ in High Copy . . . . .	81
3.20	Effects of co-transformed plasmids and dCas9 expression on GFP monitor level. . . . .	85
3.21	Schema of $P_{luxRep}$ library repression system . . . . .	91

LIST OF FIGURES

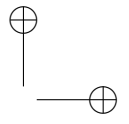
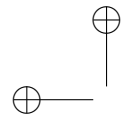
3.22	Medium Copy $P_{luxRep}$ Promoter Library Repression . . .	92
3.23	High Copy $P_{luxRep}$ Promoter Library Repression . . . . .	93
3.24	Normalized High Copy $P_{luxRep}$ Promoter Library Re- pression . . . . .	94
4.1	LuxR CRISPRi Competition . . . . .	99
4.2	Design of less effective sgRNA . . . . .	102
4.3	Behaviour of imperfect sgRNAs . . . . .	104
4.4	Repression efficiency comparison between several mod- ification approaches . . . . .	105
4.5	CRISPRi competition with constitutive competitor ex- pression . . . . .	109
4.6	CRISPRi competition with inducible competitor expres- sion . . . . .	110
4.7	CRISPRi competitor with double inducible system . .	113
4.8	Effects of competitor expression on RFP expression . .	114
A.1	Close loop genetic controller schema. . . . .	134
A.2	HSL close loop controller preliminary simulations. . . .	140
A.3	List of engineered strains tested in this study. . . . .	142
A.4	Fitting of steady-state transfer functions of the individ- ual devices of close-loop circuit. . . . .	149
A.5	Analysis of interconnected subcircuits. . . . .	151
A.6	Prediction of open-loop and close-loop circuits behavior.	153
A.7	Model prediction of disturbance rejection on HSL output.	156
A.8	Example of data from HSL production and degradation assays. . . . .	162
A.9	Characterization of the constitutive promoters used to drive aiiA. . . . .	164
A.10	Fitting of SensRegRFP-MC and OL1-SensRegRFP in presence of IPTG to over-express LuxR. . . . .	166
A.11	Time course experiments for SensRegAct-MC. . . . .	168



## LIST OF FIGURES

---

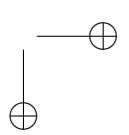
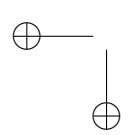
A.12	Time course experiments for CL1 without ATc. . . . .	168
A.13	Measured vs predicted HSL concentration in time course experiments of SensRegAct-MC and CL1 without ATc. . . . .	169
A.14	Time course experiment of OL1-SensRegRFP in response to ATc addition ( $100ng/ml$ at $t = 0$ ). . . . .	172
A.15	Predicted time course of HSL in turbidostat at $X = 0.2$ for OL1 and CL1. . . . .	174
A.16	Sensitivity indexes for the parameters involved in close-loop control for $X = 0.2$ . . . . .	176
A.17	Simulated steady-state level of intracellular AiiA ( $A^{SS}$ ) for different values of $X$ . . . . .	177
A.18	Cell density time courses in turbidostat experiments of OL1. . . . .	178
A.19	HSL time courses in turbidostat experiments of OL1. . . . .	179
A.20	Cell density time courses in turbidostat experiments of CL1. . . . .	180
A.21	HSL time courses in turbidostat experiments of CL1. . . . .	181
A.22	Summary of cell density values distribution in all the turbidostat experiments of OL1 and CL1. . . . .	182
B.1	RFP output data for the cascade circuits tested. . . . .	186
B.2	Growth rate data for the cascade circuits tested. . . . .	187
B.3	GFP output data for the cascade circuits with Monitor cassette. . . . .	188
B.4	Comparison plots between strains without and with Monitor cassette. . . . .	189
B.5	Correlation between GFP and growth rate in all the strains with the Monitor cassette. . . . .	190
B.6	Correlation between GFP and growth rate for each strain with the Monitor cassette. . . . .	191
B.7	Correlation between GFP and RFP for each strain with the Monitor cassette. . . . .	192



## LIST OF FIGURES

---

B.8	OD <sub>600</sub> , raw GFP and raw RFP values measured in culture, supernatant and pellet of three strains. . . . .	193
B.9	NBM fitting of the measured HSL-dependent output in all the training set circuits without Monitor cassette. .	194
B.10	Comparison of the two NOT gate characteristics as a function of the predicted per cell concentration of the TetR or LacI repressor in presence of Monitor module.	195
B.11	Univariate sensitivity analysis of the NBM by applying a variation on the $\delta$ parameter of the Hill functions. . .	196
B.12	Univariate sensitivity analysis of the NBM by applying a variation on the $\alpha$ parameter of the Hill functions. . .	197
B.13	Univariate sensitivity analysis of the NBM by applying a variation on the $K$ parameter of the Hill functions. .	198
B.14	Univariate sensitivity analysis of the NBM by applying a variation on the $\eta$ parameter of the Hill functions. . .	199
B.15	Fitting and prediction results for the NBM learned and simulated against RFP data of the circuits with the Monitor cassette. . . . .	200
B.16	BM fitting of the measured HSL-dependent RFP output in all the training set circuits with Monitor cassette. . .	201
B.17	BM fitting of the measured HSL-dependent GFP output in all the training set circuits with Monitor cassette.	202
B.18	Results of fitting using all the available data (training and test set) using NBM and BM: RFP data . . . . .	203
B.19	Results of fitting using all the available data (training and test set) using BM: GFP data . . . . .	204
B.20	Evolutionary stability of the X <sub>1</sub> T <sub>L</sub> r and X <sub>rep</sub> T <sub>L</sub> r circuits.	205
B.21	Simulation of X <sub>1</sub> T <sub>L</sub> r and X <sub>rep</sub> T <sub>L</sub> r with the NBM for different values of $\lambda_{tet}$ and $\lambda_{lac}$ parameters. . . . .	208
B.22	Raw GFP autofluorescence dependence on OD <sub>600</sub> . . .	209
B.23	Evidence of metabolic burden with a different monitor system. . . . .	213

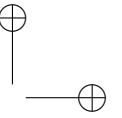
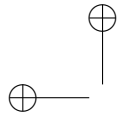




## LIST OF FIGURES

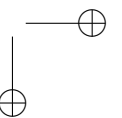
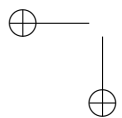
---

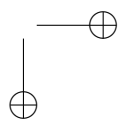
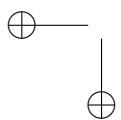
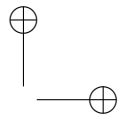
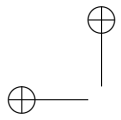
C.1	Linear regression of $q(\mu)$ and $m(\mu)$ . . . . .	229
C.2	Burden model: Global fitting of sgRNA expression cas- settes . . . . .	230
C.3	Burden model: Global fitting of MC targets . . . . .	231
C.4	Burden model: Global fitting of HC targets . . . . .	232
C.5	No Burden model: Global fitting of sgRNA expression cassettes . . . . .	233
C.6	No Burden model: Global fitting of MC targets . . . . .	234
C.7	No Burden model: Global fitting of HC targets . . . . .	235
C.8	Burden model: fitting of sgRNA expression cassettes .	236
C.9	Burden model: fitting w/o predictions of MC targets .	237
C.10	Burden model: fitting w/o predictions of HC targets .	238
C.11	No Burden model: fitting of sgRNA expression cassettes	239
C.12	No Burden model: fitting w/o predictions of MC targets	240
C.13	No Burden model: fitting w/o predictions of HC targets	241
C.14	Burden model: fitting w/ predictions of MC targets . .	242
C.15	Burden model: fitting w/ predictions of HC targets . .	243
C.16	Burden model: predictions of HSL-driven gPtet . . . .	244
C.17	No Burden model: fitting w/ predictions of MC targets	245
C.18	No Burden model: fitting w/ predictions of HC targets	246
C.19	No Burden model: predictions of HSL-driven gPtet . .	247



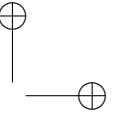
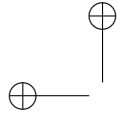
# List of Tables

2.1	Parameters description and estimated values . . . . .	35
4.1	List of altered gRNAs with relative sequences features.	103
A.1	Model parameters. . . . .	137
A.2	RBS selection for luxI and aiiA. . . . .	163
B.1	Parts used in this study. . . . .	210
B.2	Statistics on circuits bearing the burden monitor . . . .	211
C.1	Table of constructs. . . . .	215
C.2	CRISPRi model parameters . . . . .	219
C.3	Table of primers. . . . .	222
C.4	List of gRNAs and relative targets. . . . .	225







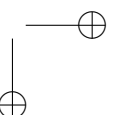
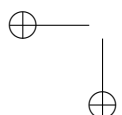


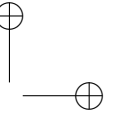
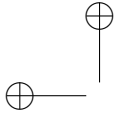
# Chapter 1

## Background

### 1.1 Synthetic Biology

The knowledge around DNA structure and manipulation achieved in the last century have resulted in an exponential growth of advanced biotechnologies and genetic engineering [1, 2], giving way to the divergence of many fields. Despite some notable exceptions [3, 4], how DNA - a molecule composed by nucleotides encoding all the information necessary to the development and functioning of living cells - is decoded and used in all the known living cells is described by the central dogma of molecular biology [5]: DNA coding sequences, called genes, are transcribed in RNA molecules, that are in turn translated in proteins by ribosomes. DNA molecules can also undergo replication processes that allow the propagation of the information to the progeny. Synthetic biology is a recent framework based on engineering and molecular biology. Although several branches and various applications as well as laboratories all around the world arose from the works of the pioneers in synthetic biology, the leitmotiv driving the studies in synthetic biology could be briefly described as did in [6]:





## 1. Background

---

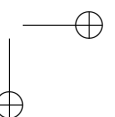
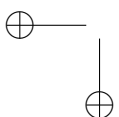
Synthetic biology aims to design and engineer biologically based parts, novel devices and systems as well as redesigning existing, natural biological systems.

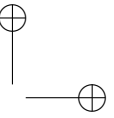
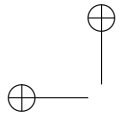
Through this simple sentence, all the fundamental aims of synthetic biology are explained: from implementing new biological systems by using synthetic molecules, to rearrange pre-existing biological entities in new structures implementing functionalities in living systems; the former chases to deeply understand and model mechanisms standing at the very basement of life, the latter to find smart and innovative solutions to solve specific problems by using an approach based on living microscopic tools.

### 1.1.1 Rational design

The interdisciplinary field of synthetic biology has allowed scientists to tackle biological problems with an engineering mindset. Differently from biotechnology and genetic engineering, in synthetic biology the rational design - a fashion that rises from the engineering related concepts of abstractions, standardization, modeling and engineering cycle - is indeed the common and shared approach adopted to implement novel living systems [6, 7, 8].

Working with living systems such as microorganisms is not trivial, since they are composed of biological elements (molecules) interacting and interfering with each other, possibly altering their behavior. Cells are also able to grow and evolve, features that could further affect the output of a system; in particular, the output of a system can be dependent on variations in growth rate of the cell population, leading to the impossibility of complying the design specifications [9, 10]. Moreover, evolution could lead further problems in systems stability when the predicted behavior of a biological system is unpreserved in time





## 1.1. Synthetic Biology

---

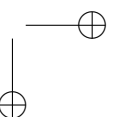
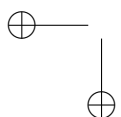
due to unexpected mutations of one or more components. To work in a biological context means also to interact with organisms that naturally react to environmental changes such as temperature and pH, being therefore active elements rather than passive chassis.

It comes obvious that the adoption of a rational approach to assemble and test novel systems is a key challenge, but is becoming of high importance with the increasing complexity in the design of synthetic biological systems.

Guidelines and sets of key rules to design biological systems with robust and predictable behavior, defining proper working conditions and reliability boundaries, need therefore to be drawn up. Under this lens, despite the differences in the context of application, this subject can be considered analogous to the others engineering fields such as electric and computer engineering [11]. The design process of a new genetic device in synthetic biology follows the engineering cycle (Figure 1.1):

1. definition of system specifications;
2. design of the system in accordance with the specification and development of a mathematical model describing the system behavior to predict its output;
3. physical implementation of the system (using suitable standardized and pre-characterized parts);
4. test and validation to verify the compliance of the system with the specifications;
5. in case of not, restart.

The engineering concepts of abstraction and modularity in synthetic biology allow the definition of a hierarchy of parts and devices, based on intrinsic complexity and interface possibilities with each other [12, 7], enabling the design of complex systems with predictable behavior



## 1. Background

---

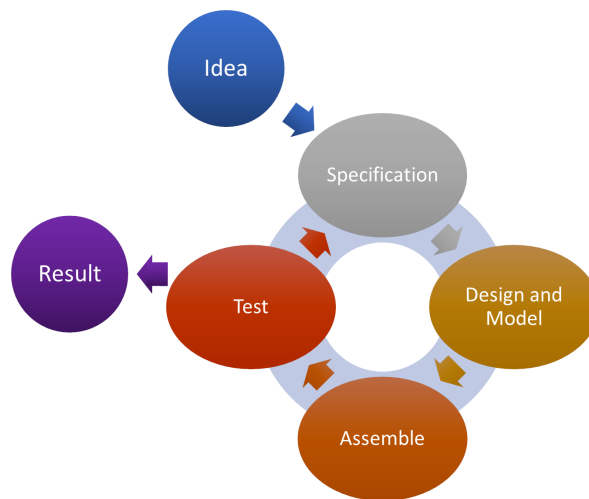
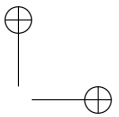
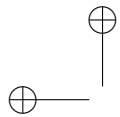


Figure 1.1: **Engineering cycle.** Design process starts with the system specifications definition, followed by the development of a mathematical model of the system to describe its behavior; subsequently, the choice of parts and connections are taken. Once assembled, the system undergoes to test and validation steps to understand whether it is compliant with the initial design specifications, or not. If not, modifications on the basis of the obtained results are applied, restarting the cycle [6].



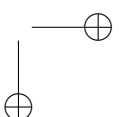
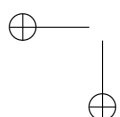
## 1.1. Synthetic Biology

---

by interconnection of basic parts, knowing the quantitative characteristics of each single sub-module [13, 14]. Under the hypothesis of modularity, the integrity of the properties of a module should be ensured [15], allowing the prediction of the behavior in a complex system just by knowing the transfer function of each single, simplified, part composing it.

To manage the complexity of these systems and increase their predictability, mathematical modeling is probably the best instrument, allowing the description of arbitrarily complex systems by using a limited amount of parameters. However, even if it is possible to define sophisticated models to be solvable exploiting the computational power of a moderns computer, it is worth noting that the intrinsic complexity of biological systems is still a major issue in the modeling step [16]. Not only mathematical models could be defined before physically assembling of a system to predict its output and evaluate its compliance with the design specification; models can indeed be used during the engineering cycle to predict how a system could react to disturbances or variations on the behavior of some of its composing parts, avoiding trial-and-error experiments [17]. In particular, differences between model predictions and experimental data could highlight inaccuracy in previous hypothesis on biological systems and, through that, explain possible unexpected behaviors of synthetic devices [6].

A successful example of rational design in synthetic biology is reported in Appendix A, where the implementation of a synthetic close-loop controller circuit for the regulation of an extracellular molecule based on quorum sensing is proposed. Nonetheless, a number of unpredictability sources may affect the behavior of a system, as discussed in the following section. Potentially, deviation between expected and observed behavior can increase for higher complexity circuits and current models are still far from being able to include all the possible effects arisen from biological complexity. Despite several studies aimed to elicit this variability and hypothesize or demonstrate its sources, some





## 1. Background

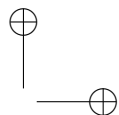
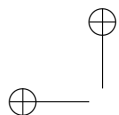
---

modeling [18, 19] or design [20, 21, 22] solutions have also been proposed; as always, further efforts are still needed to be spent in this direction.

### 1.1.2 Unpredictability sources

The predictability of the designed circuits is a central issue in synthetic biology, since only in a predictable framework biological systems can be constructed from the bottom-up. Mathematical models can support the design process, enabling the rational engineering of complex systems and avoiding trial and error approaches [23, 24, 25]. Although standardized approaches for the characterization of parts have been recently proposed, the intrinsic complexity of biological components currently limits the predictability of parts function when they are re-used in different contexts [26, 27].

The major unpredictability sources for biological components are context dependent and cell-to-cell variability, cross-talk, evolutionary stability, retro-activity and cell burden [28, 26]. Efforts towards the reproducible characterization of parts function include standardized measurement approaches for transcriptional activity [29] and biophysical models for the quantification of ribosome binding sites (RBSs) strength [18, 30] or transcriptional terminators efficiency [31, 32]. The study of simplified model systems can help to elucidate the feasibility boundaries of the bottom-up design approach in biological engineering. Within this framework, different studies used mathematical models to study the superposition of the effects of multiple independent gene expression cassettes [33], context-dependent variability of individual or interconnected devices [34, 13, 35] retroactivity effects due to the interconnection of biological modules which share common resources [15] and prediction of quantitative behaviour of logic functions [36, 37, 38] or feed-forward circuits [39]. High-throughput studies have also been carried out to evaluate the variation of parts activity in a large num-



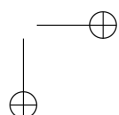
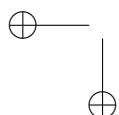
---

## 1.2. Metabolic Burden in Rational Design

ber of diverse expression systems, showing the variations expected for promoters, RBSs and genes with different codon composition [40, 41]. Efforts towards the improvement of biological components modularity have recently been carried out by proposing insulated promoters [42, 20], a bicistronic design for gene expression cassettes that makes RBS efficiency more predictable [43], a device for timescale separation to mitigate retroactivity [44] and ribozyme-based insulators at 5'-UTR [21]. Recent works have also proposed methods to guide biological engineers in parts selection, via statistical analyses to evaluate promoter and RBS collections [45], and a computer-aided design tool for the choice of logic devices to construct reliable functions [46]. In the latter study, devices are selected via model-based approach from the knowledge of their transfer function, also considering the minimization of cell burden and failure rate caused by the multiple use of the same part in a circuit.

## 1.2 Metabolic Burden in Rational Design

One of the main factors leading to unpredictable behavior of synthetic circuits is cell burden [47]. The unnatural load caused by heterologously expressed genes can lead to transcriptional and translational resources depletion, exerting important global effects on the functioning of the designed circuit [47, 48]. Synthetic circuit designs aimed to reduce the metabolic load for the cell have been reported [49, 50, 51, 52], in which systems with superior protein yield or functions were obtained. Experimental and in-silico methods have also been recently proposed to analyze cell burden in synthetic circuits [51, 53, 19, 54]. The use of a constitutive expression cassette for a reporter gene, integrated in the bacterial chromosome, has been adopted as a real-time and in vivo burden monitor, to indirectly quantify the cellular resources limitation via microplate assay [51]. This method-





## 1. Background

---

ology was demonstrated to be more sensitive than growth rate measurement for burden quantification. Other works have used the same approach, with the constitutive cassette assembled in plasmid, to study cell burden via modelling frameworks based on electronic engineering [54] and microeconomics [19]. Different mathematical models have been proposed for the analysis of protein expression in a limited resources context [53, 55, 56]. Such recently proposed models have been useful to identify the expression systems behaviors occurring when resources are limiting and cannot be trivially explained via simple Hill function-based activation/repression models [57]. However, such burden models still have shortcomings, e.g., they are unable to explain the possible separation of cellular resource pools among chromosome and plasmids (as suggested by Gyorgy et al. [19]) and the relationship between cell growth rate and resource pools is still lacking in such models, although it has been included in one recent study on dynamics of protein expression [58]. While most of the literature studies analyzed cell burden in non-interacting gene expression systems, a recent in-silico study indicated that burden can largely affect the quantitative function of interconnected circuits, in which non-trivial activation and repression functions may emerge [53]. Such previously unexpected behavior was confirmed by recent in vivo experiments involving a simple gene regulatory network, tested with two diverse regulatory gene RBSs and circuit copy numbers [59]. In the same work, an interaction graph-based theoretical framework was proposed to describe the effective interactions occurring among network modules, and eventually guide the design of circuits with different topologies [59]. The works mentioned above demonstrate the need of further steps towards the testing of a rigorous bottom-up approach in the design of interconnected synthetic circuits and they also highlight that cell burden is an important feature to be modeled in order to describe otherwise unpredictable outputs.



## 1.3 CRISPR/Cas9 system

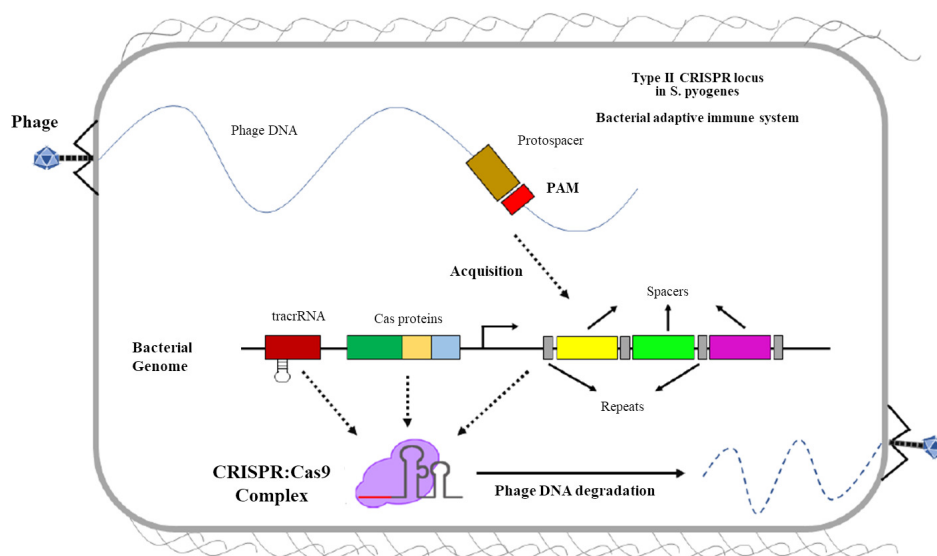
### 1.3.1 Mechanism and first usages

Up until recently, the function of *Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)* and their *CRISPR-associated proteins (Cas)* remained unknown. The former were firstly mapped in bacterial genomes because of their particular structures: short sequences (*protospacers*) separated by identical palindromic DNA spacers, altogether known as a *CRISPR array*. More recent studies revealed that those particular DNA structures are the hallmark of bacterial adaptive immunity against viral infections caused by bacteriophages [60, 61, 62]; several orthologous versions of this biological machinery have been found in almost 40% of Bacteria and 90% of Archea [63]. In *Streptococcus pyogenes*, the first and most studied Bacteria naturally expressing the CRISPR system, when a bacteriophage injects its genome into a host bacterial cell, the immune system produces a pool of Cas proteins that form a complex able to recognize a *protospacer adjacent motif (PAM)* sequence in the exogenous genome; from the phage DNA, a *protospacer* sequence of 20 nucleotides, adjacent to the PAM, is then excised and finally integrated into the CRISPR array. From this point on, the bacterial cell is said to be immune to the specific type of phage which protospacer sequence have been acquired: if the cell progeny is re-infected by the same kind of phage, the CRISPR array - containing, amongst others, the spacer relative to the infecting phage - is transcribed (pre CRISPR RNA or pre-crRNA) and spliced into several *CRISPR RNAs (crRNAs)* (one for each acquired spacer) along with the synthesis of another Cas protein called *Cas9*. This enzyme has nuclease activity and associates with the a linker RNA called *trans-activating-CRISPR-RNA (tracr-RNA)* that binds its 5' side to the non-complementary part of the crRNA and the 3' side to Cas9, thus forming a RNA structure called guide RNA

## 1. Background

(gRNA) and linking it to the enzyme. Cas9 nuclease can be therefore guided to the foreign phage DNA thanks to the gRNA with which is complexed, as described in Figure 1.2. Thus, the phage genome is degraded, thus halting replication and infection.

It is worth to notice that again, the gRNA:Cas9 complex is able

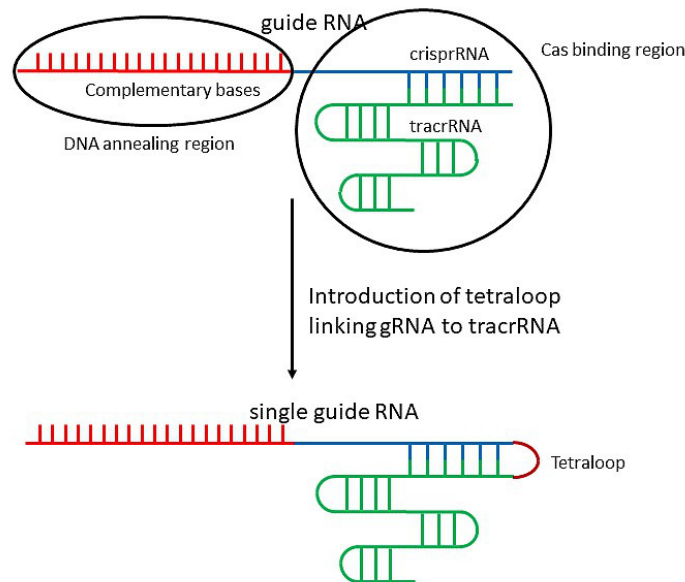


**Figure 1.2: Bacterial Immunization.** When a bacteriophage infects a bacterial cell, CRISPR-associated proteins locate a PAM sequence in the phage genome, excise a fragment of DNA (protospacer) and integrate that fragment into the CRISPR array in the bacterial genome. Subsequent infection with the same bacteriophage activates an immune response that ends with phage genome degradation.

to bind only DNA sequences that are adjacent to a PAM site, which recognition by one of the Cas9 domains is the first mandatory step necessary for the formation of a stable DNA:gRNA:Cas9 complex [64]. Inspired by the naturally occurring system, groups have isolated the *Streptococcus pyogenes* Cas9 protein and have created a chimeric version of the gRNA-tracrRNA pair of RNAs by linking the two via a loop

### 1.3. CRISPR/Cas9 system

of 4 nucleotides [65]. In this way it has become possible to achieve precise DNA strand breaks by simply expressing a single protein along with single RNA molecule for desired target, virtually addressing every possible 20 nucleotide sequence on a DNA molecule adjacent to a PAM sequence (see Figure 1.3).



**Figure 1.3: sgRNA design.** Qi et al. [65] have created a single guide RNA by linking the guide RNA and tracr RNA via a tetraloop. This has allowed for easy and cheap design and synthesis of CRISPR/Cas systems applied for a variety of purposes. The creation of a sgRNA leads to reduced degrees of freedom when designing the complementary RNA strand for Cas targeting, resulting in a simpler, more basic method with less room for error and variation.

Before the CRISPR/Cas system was used as a means of genetic modification, other different sequence specific nucleases (SSNs) were developed and patented by different groups. The first to be discovered were meganucleases, that have the ability to recognize specific 18-23

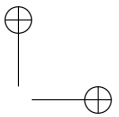
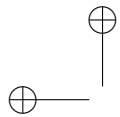


## 1. Background

---

nucleotide sequences and produce double strand breaks at specific loci [66]. However they proved too large and complex for efficient manipulation of the guiding sequence. Next, zinc-finger domains were discovered to bind DNA through conserved protein motifs, where every zinc finger has the ability to recognize 3 sequential nucleotides [67]. These zinc-finger domains were exploited for the DNA recognition ability, and fused to a FokI domain with nuclease activity, resulting in a multi-domain protein with the ability to find and cleave specific sequences, known as zinc-finger nucleases (ZFNs)[66]. A more recent targeting mechanism was discovered in plants, who possess transcription activator-like effector (TALE) proteins. Similarly to zinc fingers, they are composed of highly conserved domains that are able to recognize only one nucleotide. These TALE domains are subsequently fused to a FokI domain with nuclease ability resulting in a sequence specific nuclease - TALEN -, whose domains are able to recognize one nucleotide at a time [68, 69]. TALENs and ZFNs both rely on protein engineering and protein-DNA interactions for recognition; nevertheless, the repetitive nature of the conserved domains makes it more difficult to assemble a functional SSN, since the cleavage domain of FokI effects a double-strand break only upon dimerization of two subunits located on the two complementary strands of the double helix; therefore two proteins are necessarily expressed in the desired cell to induce the DNA modification, with consequent problems in both delivery and metabolic load due to the size of the encoding genes.

The CRISPR revolution was pioneered by Jennifer Doudna among others, who thought to exploit the targeting mechanism of the CRISPR/Cas system for the purpose of gene editing, and developed a more facile method for generating a functional gRNA (see Figure 1.3) [65]. This has allowed for a boom in genetic modification experiments, since CRISPR-based SSN became widely available and easy to assemble. The main advantage of CRISPR/Cas based targeting is the bypass of the complex protein engineering required to manufacture ZFNs or



---

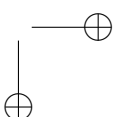
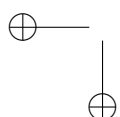
### 1.3. CRISPR/Cas9 system

TALENs, both of which are protein based, because Cas proteins are guided by none other than a guide RNA that associates with the Cas protein and directs it to the specific sequence; the mechanism of pairing between the CRISPR:Cas complex and the target DNA relies solely on DNA-RNA complementarity, and consequently, designing guides has proven to be simpler, faster, and cheaper.

However, the knowledge on DNA repair mechanisms of cells allowed us to exploit the cell ability to drive homologous recombination in response to DNA damage. As such, techniques emerged for targeted mutagenesis of host cells by providing template DNA fragments in combination with SSN that upon DNA cleavage activate DNA repair enzymes and augment the probability of correct homology-directed repair of the gene locus. Through this technology, scientists are able to insert or exchange segments of coding DNA with unprecedentedly high precision, as well as produce functional knock-outs via inducing non homologous end joining (in cells where this kind of repair is present), where small indels alter the reading frame in the coding sequences of a target gene. These targeted mutagenesis can affect protein expression and in turn cell fate [70, 71, 72].

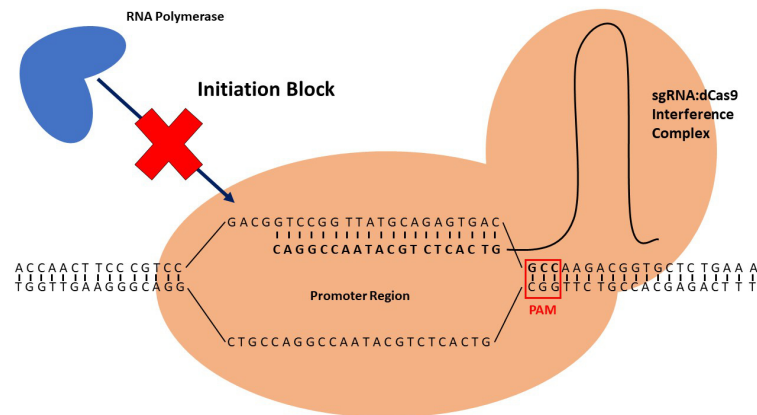
#### 1.3.2 Catalytically inactive Cas9: CRISPR interference and derivatives

Modifications of the original CRISPR/Cas system emerged to serve as transcriptional regulators of genes. The dead Cas9 (dCas9) engineered protein for the targeted repression of genes, lacking nuclease catalytic activity due to single nucleotide deletions in its two cleaving domains, has been developed in [65]. This system, known as *CRISPR interference* - *CRISPRi*, exploits targeting via gRNAs to achieve gene silencing by steric occupancy to elongation or even binding of RNA-



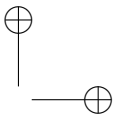
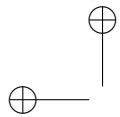
## 1. Background

polymerases (see Figure 1.4). Further, CRISPR/dCas9 were also fused



**Figure 1.4: Transcription initiation block by CRISPR/dCas9 complex.** As a means of silencing gene expression, sgRNA have to be designed complementary to promoter regions of reporter genes. The binding of the interference complex to the promoter region prevents transcription initiation by sterically hindering the recruitment of polymerase to the promoter.

to transcription activators, that ultimately lead to an increase in gene expression [73] despite this system so far have been developed mainly in eukaryotic cells. Other studies fused dCas9 to domains recruiting chromatin remodeling proteins such as histone acetylases/deacetylases, histone methylases, DNA methyl transferases, all of which either repress or enhance gene expression without altering the sequence of DNA [74]. Moreover, sequence-specificity of these systems implies potent selection methods of specific gene sequences in a large pool of cells with different genomes, to be applied ultimately to complex bacterial ecosystems such as the human microbiome [75, 76]. A group of scientists have been able to select and kill only pathogenic bacteria present in complex bacterial ecosystems by employing a phage vector carrying a CRISPR/Cas system directed at specific sequences in pathogenic bacteria known as virulence genes. Studies are still ongoing, but it



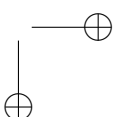
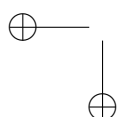
---

## 1.4. Project Idea and bigger picture

demonstrates the powerful ability of SSNs to be used as a selection tool for the targeted elimination of specific bacterial species. Another application of this system, in which studies were performed *in vivo*, involves the use of a viral vector containing a combination of guide RNAs targeting retroviral LTRs and structural genes to achieve the efficient excision of pro-viral DNA from infected host cells [77], providing a viable method for the cure of diseases that until this day only have no remedy, such as AIDS and herpes. These systems rely on the multiplexing of gRNA to target several sequences while maintaining expression of the same single Cas protein; desired repression of multiple target genes with transcription factors requires complex systems and many cellular resources rendering it impractical. A recently developed CRISPR imaging system fuses GFP proteins to dCas9 for the visualization of several loci in the genome; this system, called CRISPR-Tag, allows an in-depth look at the effect of spatial-temporal organization of genes [78].

## 1.4 Project Idea and bigger picture

In this study, firstly, synthetic transcriptional cascades in *Escherichia coli* obtained upon interconnection of different inducible and repressible devices were analyzed. To elucidate the reliability of currently available mathematical models applied to this circuit collection, the predictive power of a widely used Hill function model and of one of the recently proposed models that considers cell burden due to resource limitation were tested. Specifically, the latter considers Hill functions to describe specific interactions among circuit elements together with cell burden that modulates protein expression. The feasibility boundaries of a bottom-up approach are elucidated, together with the importance of taking into account cell burden in quantitative predictions.





## 1. Background

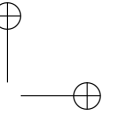
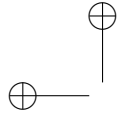
---

To expand the possibilities in rational design of genetic circuits bearing a notable amount of cellular burden, a CRISPRi-based model system has been implemented in BioBrick<sup>TM</sup> standard and deeply characterized, highlighting its working range and drawing up an usage blueprint. The flexibility of gRNA design allows to target almost any desired sequence, granted the presence of an adjacent PAM motif required for system function, whereas transcription factors have affinity towards a single or few sequence(s) of preference. Nonetheless, sgRNAs are macromolecules that do not undergo the translation process, which is known as the main source of metabolic load (thus the net load is given by dCas9 expression only). Once the circuit is designed, a mutagenesis is sufficient to introduce the desired sgRNA sequence, in order to customize the regulatory machinery by changing target and specifically tuning the transcription of the desired genes.

In the last part of this work, CRISPRi-based circuits are adopted in the design of synthetic circuits of increasing complexity. In particular, different aspects of high interest for the use of sgRNAs in circuit design are touched: use of protein repressors and sgRNAs to implement double regulation, use of sgRNAs as low-burden alternative of protein repressors, tuning the efficiency of sgRNA repression, and competition effects among different sgRNAs expressed in the same cell.

Taken together, the investigations reported in this thesis are expected to improve our design capability of synthetic circuits by providing means of taking into account and decrease cell burden. The main tool used is based on CRISPRi, for which a thorough characterization is provided in different contexts, highlighting advantages and limits. A mixture of *in silico* and *in vivo* studies is adopted to reach the described goals.





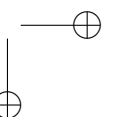
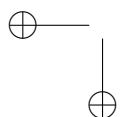
## Chapter 2

# Metabolic Burden Modeling using transcriptional cascades as case study<sup>1</sup>

In this study, synthetic transcriptional cascades in *Escherichia coli* obtained upon interconnection of different inducible and repressible devices, were analyzed. To elucidate the reliability of currently available mathematical models applied to this circuit collection, it was aimed to test the predictive power of a widely used Hill function model and of one of the recently proposed models that considers cell burden due to resource limitation. Specifically, the latter considers Hill functions to describe specific interactions among circuit elements together with cell burden that modulates protein expression. These models are described in 2.1 and will be referred to as *no-burden model* (NBM) and *burden model* (BM), respectively. The study presented in this work

---

<sup>1</sup>This chapter has been taken and re-adapted from the article “*Re-using biological devices: a model-aided analysis of interconnected transcriptional cascades designed from the bottom-up*”[79]





## 2. Metabolic Burden modelling

---

elucidates the feasibility boundaries of a bottom-up approach and the importance of taking into account cell burden in quantitative predictions.

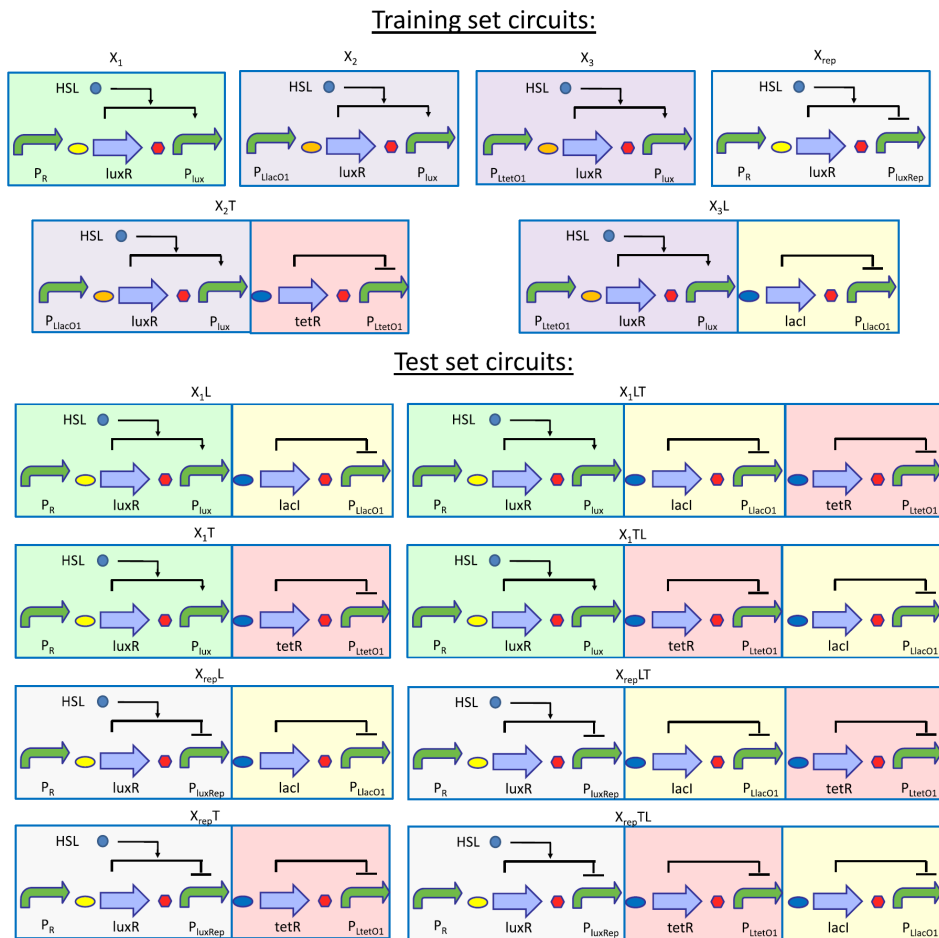
### 2.1 In-vivo characterization

#### 2.1.1 Circuits description

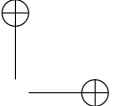
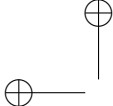
The circuits analyzed in this study are described in Figure 2.1. Their design is based on the widely used lux, tet and lac systems elements, and the RFP and GFP reporter genes (see Table B.1 for a description of all the basic parts used). The circuits topology implements transcriptional cascades [20, 80] composed by an HSL-inducible or -repressible input block upstream of NOT gate blocks (none, one or two) connected in series. Finally, an RFP expression device is assembled downstream of the cascade to serve as detectable circuit output. All of them have been assembled in the low-copy pSB4C5 vector [81]. Input blocks include a constitutively expressed luxR gene with a strong RBS (BBa\_B0030 or BBa\_B0034) under the control of the  $P_R$ ,  $P_{LtetO1}$  or  $P_{LlacO1}$  promoter [82], and the wild-type inducible  $P_{lux}$  promoter [83] or the strongest member of a synthetic repressible promoter library [84], herein called  $P_{luxRep}$ , downstream. The NOT gates include the tetR or lacI repressor gene, with a weak RBS (BBa\_B0031), and their cognate repressible promoter  $P_{LtetO1}$  or  $P_{LlacO1}$ , respectively, downstream [85]. The tetR and lacI genes both have an LVA fast-degradation tag for the translated protein [86].

One- to three-block cascades have been studied by using different combinations of these devices. A collection of the same circuits has also been constructed with a reporter expression cassette downstream, composed by a GFP expression system driven by a constitutive promoter (BBa\_J23100). This additional gene expression cassette will be

## 2.1. In-vivo characterization



**Figure 2.1: Collection of circuits analyzed in this study.** All of them are available with an RFP expression system downstream of the output promoter (indicated in the text with the *r* suffix), and also with a GFP expression cassette driven by a constitutive promoter downstream (indicated in the text with the *rg* suffix, meaning that both RFP and GFP can be measured to quantify circuit output and cell burden, respectively). Curved green arrows represent promoters; straight violet arrows indicate coding sequences; red hexagons represent transcriptional terminators; ovals represent RBSs (BBa\_B0030 yellow; BBa\_B0034 orange; BBa\_B0031 blue); circle represents HSL. Activation and repression are indicated as thin arrows. Block color is consistent among the circuits



## 2. Metabolic Burden modelling

---

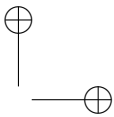
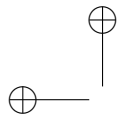
referred to as Monitor cassette, and it will be adopted to quantify cell burden, as previously carried out [51, 19, 54].

The described circuits were divided into training and test sets (see Figure 2.1). In particular, the  $X_1$ ,  $X_2$ ,  $X_3$  and  $X_{rep}$  configurations have been used as specific measurement system constructs for the characterization of input blocks, whereas  $X_2T$  and  $X_3L$  have been used to characterize individual NOT gates. The ones to characterize input blocks include the input device with RFP downstream. The ones to characterize NOT gates include a pre-characterized input device upstream, in order to tune the expression of tetR and lacI genes over a range of levels, and RFP downstream to measure the NOT gate block output. In the context of a bottom-up approach, the characterization of the circuits above was used to predict the behavior of the test set circuits, which are composed by different combinations of the characterized blocks.

The used input devices provide homogeneous transcriptional output with no bimodal distribution of gene expression [37, 84, 87]. For this reason, the network topology used for the circuits in this work provides unimodal outputs at all the cascade levels.

### 2.1.2 Circuits characterization

Fluorescence and absorbance of recombinant bacteria incubated in a microplate reader were measured over time as previously described [13, 87, 84]. Briefly, bacteria from a glycerol stock were streaked on a selective LB agar plate. After 16– to 20 –  $h$  incubation at 37°C, 1ml of selective M9 was inoculated with a single colony. For strains expressing a repressor in absence of HSL, the inducer was added at this step, at a proper concentration, to allow them to reach a steady-state of intracellular proteins, and to avoid long dynamics due to repressor proteins degradation and dilution during the microplate assay. After 21 –  $h$  incubation at 37°C, 220rpm, in an orbital shaker, cul-



---

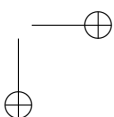
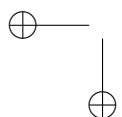
## 2.1. In-vivo characterization

tures were 100 fold diluted in a final volume of  $200\mu l$  in a 96-well microplate. HSL ( $2\mu l$ ) was added when required, to reach the desired final concentration. Cultures were not placed in the external wells of the plate to avoid intensive evaporation during incubation. The microplate was incubated with lid in the Infinite F200 microplate reader (Tecan) and it was assayed via kinetic cycle: 15s linear shaking (3mm amplitude), 5s wait, absorbance (600nm) measurement, fluorescence measurements, 5min sampling time. RFP and GFP fluorescence was measured with a gain of 80 with the 535/620nm and 485/540nm filter pairs, respectively. Control wells were always included, as described in the following Data processing Section 2.1.3, to measure the background of absorbance and fluorescence, and to provide internal control references for relative activity calculations. At least three biological replicates, starting from different colonies, were assayed for each strain.

### 2.1.3 Data processing

Data analysis and related graphs were carried out via Microsoft Excel and Matlab R2007b or R2017b (MathWorks, Natick, MA). Pairwise correlations and corresponding p-values, as well as correlation matrices, were computed via the Matlab *corr* function. Linear regressions were carried out via the Matlab *regress* function.

Raw absorbance and red fluorescence time series were blanked by background subtraction as previously reported [13, 88] to obtain  $OD_{600}$  and RFP time series. Sterile medium and a non-fluorescent TOP10 culture were used as absorbance and red fluorescence background, respectively. Since a significant cell density-dependent autofluorescence was previously reported for GFP measurements with the adopted experimental setup [89], green fluorescence was blanked via a different procedure: for each GFP-expressing strain, a control strain with identical circuit and HSL concentration, but without GFP expression cas-





## 2. Metabolic Burden modelling

---

sette, was considered. The raw green (auto)fluorescence ( $GFP_{auto}$ ) vs  $OD_{600}$  characteristic (at least two biological replicates) was fitted via an exponential regression:

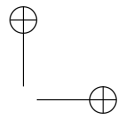
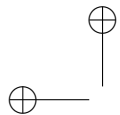
$$GFP_{auto}(t) = e^{q+m \cdot OD_{600}(t)} \quad (2.1)$$

This curve was used to estimate the green fluorescence background of a target culture, given its  $OD_{600}$  at each time point. The  $GFP_{auto}$  value was subtracted from the raw fluorescence of the target culture to obtain a signal proportional to the GFP level in the whole culture. Fluorescent protein synthesis rate per cell ( $S_{cell}$ ) was computed for each culture and fluorescent protein as:

$$S_{cell} = \frac{dF}{dt} \cdot \frac{1}{OD_{600}} \quad (2.2)$$

where  $F$  is the RFP or GFP fluorescence level in the whole culture; it was then averaged over the exponential growth phase ( $0.05 < OD_{600} < 0.18$ ) [84]. The obtained values were divided by the average  $S_{cell}$  of a reference culture, constitutively producing RFP or GFP with the same expression system under the control of the BBa\_J23101 promoter, yielding  $S_{cell,norm}$ . Reference cultures for RFP and GFP have the BBa\_J107029 and BBa\_K173001 expression cassettes, respectively. When the growth rates of target strain and reference cultures are similar,  $S_{cell,norm}$  is equivalent to the Relative Promoter Unit (RPU) value. A strain only including the Monitor cassette was also considered (herein called *Monitor culture*) to estimate the GFP level without the cell load caused by the circuits. In this strain, a BBa\_B0015 transcriptional terminator was assembled upstream of the Monitor cassette to enable GFP measurements with the BBa\_J23100 promoter in the same surrounding context of all the circuits, in which this terminator is always present upstream of the Monitor cassette.

Growth rate was computed via linear regression of  $\ln(OD_{600})$  vs time



## 2.2. Models description

---

characteristic in the  $0.05 < OD_{600} < 0.18$  window [84, 89].

The inclusion of specific control strains without Monitor cassette in different conditions for target strain autofluorescence estimation was necessary because such background value was found to be not only  $OD_{600}$ -dependent, but also growth rate-dependent (see Figures B.8, B.22), and strains bearing circuits with or without Monitor have similar growth rates (see Figure B.4). The background of the GFP reference culture and the Monitor culture was estimated using the RFP reference culture as control.

## 2.2 Models description

### 2.2.1 Mathematical description: No-burden model (NBM)

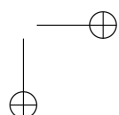
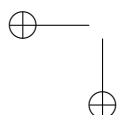
Models including Hill functions were used to describe activation and repression of proteins expression in the analyzed circuits. Intracellular protein levels were modeled via dynamic equations as previously performed in many works [38, 21, 46], assuming the steady-state of all the intracellular species in exponentially growing cells, and assuming no metabolic burden affecting the cells.

The level of a repressor protein ( $P_j$ ) in the NOT gate blocks is computed as:

$$P_j = \frac{1}{\mu + \gamma} \cdot S_j \quad (2.3)$$

where  $\gamma$  is the protein degradation rate due to the LVA tag,  $\mu$  is the cell growth rate, which depends on recombinant strain and HSL concentration, and  $S_j$  is the  $P_j$  synthesis rate per cell, defined as:

$$S_j = \delta_j + \frac{\alpha_j}{1 + \frac{K_j^{\pm\eta_j}}{I_j}} \quad (2.4)$$





## 2. Metabolic Burden modelling

---

where  $\delta$ ,  $\alpha$ ,  $K$  and  $\eta$  are the Hill equation parameters that characterize the upstream regulated promoter; in particular,  $\delta$  is the basic expression rate in the off-state,  $\delta + \alpha$  is the maximum expression rate,  $K$  is the input ( $I$ ) level corresponding to 50% of the expression rate range, and  $\eta$  is the Hill coefficient (positive if the upstream promoter is inducible, negative if repressible); finally,  $I$  is the function input, which can be a per-cell protein level (if the NOT gate has another NOT gate block upstream) or HSL concentration (if the upstream block is an input block). Growth rate is assumed to affect protein dilution rate due to cell doubling, but not all the other processes (e.g., transcription and translation).

The immature (i.e., non-fluorescent) reporter protein per-cell level ( $R$ ) is computed as:

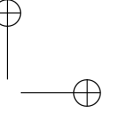
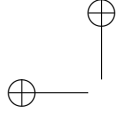
$$R = \frac{1}{\mu + a} \cdot S_j \quad (2.5)$$

where  $a$  is the fluorescent protein maturation rate and the other symbols have the same meaning as above. Finally, the mature reporter protein synthesis rate per cell ( $S_{cell,norm}$ ), which is the measured output of the circuits, is computed as [29]:

$$S_{cell,norm,RFP} = a \cdot R \quad (2.6)$$

$S_{cell,norm,RFP}$  is expressed as arbitrary units of RFP ( $AU_R$ , if considering circuit output) per cell per time ( $AU_R cell^{-1} min^{-1}$ ). RFP per cell concentration is assumed to be proportional to its respective arbitrary units. To support the predictable interconnection of biological devices, inputs and outputs of all the circuit blocks need to be expressed with the same units [38]. To this aim, the regulated promoter of all the blocks is characterized in  $AU_R cell^{-1} min^{-1}$  units and the intracellular levels of all the proteins of the network can be expressed as  $AU_R cell^{-1}$ . The basic underlying assumptions are that a promoter is able to drive any downstream-connected gene to the same





---

## 2.2. Models description

activity-dependent expression level [21], and the resulting protein level is assumed to be proportional to the gene expression level [29]. Such assumptions enable to model and re-use different biological devices by expressing their activities in comparable units [38], in absence of context-dependent variation of parts function [90].

### 2.2.2 Modeling in a limited resource context: Burden model (BM)

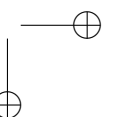
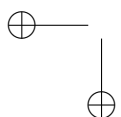
In a limited resource context, RNA polymerase and ribosome intracellular levels have also to be taken into account. In order to include this aspect in the mathematical description of the systems, the model proposed in the work by Qian et al. [53] was adopted and integrated in the previously described *No-burden model* - *NBM*. While the full model derivation procedure and assumptions are described in the work by Qian et al. [53], from a structural point of view the only difference between the BM and the standard model based on Hill equation (NBM) is the presence of a denominator ( $D$ ) that affects all the protein synthesis rates. Referring to Eq. 2.4, the protein synthesis rate term becomes:

$$S_j = \frac{S_{max,j}}{D} \quad (2.7)$$

$$S_{max,j} = \delta_j + \frac{\alpha_j}{1 + \left(\frac{K_j}{I_j}\right)^{\pm\eta_j}} \quad (2.8)$$

$$D = 1 + \sum_{k=1}^M J_k \cdot S_{max,k} \quad (2.9)$$

where  $M$  is the number of expressed proteins in the cell,  $S_{max,k}$  is the maximum achievable synthesis rate of the  $k$ -th expressed protein, and  $J_k$  is the related resource usage parameter, representing a measure of





## 2. Metabolic Burden modelling

---

the burden caused by the  $k$ -th protein. This denominator includes not only the effect of the genes in the synthetic network, but also the ones of the organism. The sum of the organism gene contributions to the burden ( $z$ ) can be assumed to be a constant circuit- and induction-independent term:

$$z = \sum_{k=1}^C J_k \cdot S_{max,k} \quad (2.10)$$

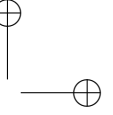
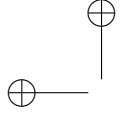
where  $C$  is the number of expressed organism genes. The denominator  $D$  can be re-written as:

$$D = 1 + z + \sum_{k=1}^Y J_k \cdot S_{max,k} \quad (2.11)$$

where  $Y = M - C$  is the number of proteins expressed in the synthetic circuit. Being  $(1 + z)$  a constant term,  $D$  can be rescaled by dividing each term by  $(1 + z)$  to obtain  $\hat{D}$ :

$$\hat{D} = 1 + \sum_{k=1}^Y J_k \cdot \hat{S}_{max,k} \quad (2.12)$$

where  $\hat{S}_{max,k}$  are the maximum achievable synthesis rates rescaled by  $(1 + z)$ . In this case, all the synthesis rates and intracellular protein levels in the model are rescaled by this term. For this reason, the  $P_j$ ,  $S_{max,j}$  and  $S_j$  in the BM can be interpreted as the protein level, maximum synthesis rate and actual synthesis rate relative to the endogenous resource usage term,  $1 + z$ , without affecting their units or the functionality of the model. In the BM, the Hill equation represents the maximum achievable synthesis rate (Eq. 2.7-2.8) and has a different interpretation compared to the NBM, in which it represents the actual synthesis rate (Eq. 2.4). Since all the circuits analyzed with the BM contain the Monitor cassette, the constant contribution



## 2.2. Models description

---

of GFP expression was included among the organism genes (although GFP expression was found to have a negligible contribution to cell burden, as described in Section 2.3) without affecting the meaning of the described quantities.

The resource usage terms,  $J$ , are expressed in ( $AU_R^{-1} cell min$ ) units. However, the contribution to  $\hat{D}$  of non-regulated proteins in the circuits (i.e., LuxR in the input blocks) is herein expressed by the dimensionless constant term  $\Sigma = J \cdot \hat{S}$ .

Differently from the NBM, GFP expression is also modeled to enable the quantification of cell burden. The intracellular level ( $G$ ) of immature GFP is computed as:

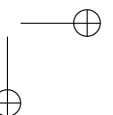
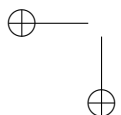
$$G = \frac{1}{\mu + a_G} \cdot \frac{S_m}{\hat{D}} \quad (2.13)$$

where a  $G$  is GFP maturation rate,  $S_m$  is the synthesis rate in the Monitor cassette (expressed as arbitrary units of GFP -  $AU_G cell^{-1} min^{-1}$ ), and the other symbols have the same meaning as above.

Analogously to the RFP output, the Monitor cassette output is described as:

$$S_{cell, norm, GFP} = a_G \cdot G \quad (2.14)$$

In addition to the NBM assumptions, RNA polymerase and ribosome levels were further assumed to being not affected by cell growth rate, and that cell burden and growth rate do not considerably affect the transfer function of input devices due to LuxR protein level variation [54]. The inclusion of these two phenomena would require a model relating growth rate and RNA polymerase/ribosome levels, and the explicit modelling of LuxR production and binding with HSL via a mechanistic model, such as the one proposed by Carbonell-Ballester et al. [91]. Both interventions are beyond the scope of this work and can be topic of additional modeling studies.





## 2. Metabolic Burden modelling

---

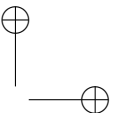
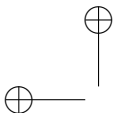
### 2.3 Application of the models

#### 2.3.1 Data overview

Cascade output level at steady-state was measured as a function of HSL for all the circuits via RFP analysis (see Figure B.1). Considering circuits without Monitor, their RFP output span a wide range of values ( $> 640$  fold), with all the circuits showing a relevant output variation as a function of HSL concentration, from 3 fold ( $X_{rep}TLr$ ) to 141 fold ( $X_1r$ ). The growth rate of these recombinant strains span a 2.5 fold range (see Figure B.2). The quantitative behavior of individual devices was consistent with previously published characterization data (see Appendix B) [13, 84, 87, 91, 92, 93].

According to the inducible or repressible behavior of the constructed circuits that can be inferred from the individual blocks, all the 1- and 2-block circuits showed the expected logic behavior. However, only two of the four 3-block circuits ( $X_1LT_r$  and  $X_{rep}LT_r$ ) showed the expected logic output trend: the  $X_1TL_r$  and  $X_{rep}TL_r$  circuits, in which the output should be an increasing and decreasing function of HSL, respectively, showed a clear decreasing output ( $X_1TL_r$ ), and an increasing and then decreasing ( $X_{rep}TL_r$ ) output. Cell resource limitation may give rise to such qualitatively unpredictable behavior, in which the logic activation and repression rules were affected by hidden interactions caused by cell burden [53]. For this reason, analysis of circuits in presence of a burden measurement system was carried out.

Considering circuits with Monitor cassette, GFP was analyzed (see Figure B.3), in addition to RFP and growth rate (see Figures B.1-B.2), and used to indirectly measure cell burden. The RFP output and growth rate are highly similar to the ones of the circuits without Monitor, suggesting that the Monitor itself does not provide relevant burden to the cell (correlation value of 0.99 and 0.83, respectively, see



### 2.3. Application of the models

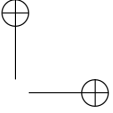
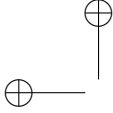
---

Figure B.4).

In presence of a monitor cassette, a strong correlation between growth rate and GFP was previously observed, caused by growth rate decrease in presence of cell burden [51]. By contrast, here a statistically significant but very low correlation (0.27) was observed (see Figure B.5). Considering individual circuits, only three of them ( $X_2Trg$ ,  $X_1Trg$  and  $X_1TLrg$ ) showed a statistically significant growth rate-GFP correlation (see Table B.2 and Figure B.6), with the 0.41, 0.72 and 0.84 values, respectively. These constructs have in common a highly expressed tetR repressor, while in the other circuits its transcription is driven by weaker promoters. Consistently, these three circuits are also characterized by the lowest GFP levels among the tested circuits (see Figure B.3). Such data suggest that a correlation between growth rate and GFP can be detected only in the circuits causing the highest cell burden, while in other recombinant strains no significant growth rate-GFP correlation could be seen, even though GFP and growth rate showed large variations and GFP exhibits a clear HSL-dependent trend.

In previous studies of circuits including single non-interconnected expression cassettes, RFP and GFP also showed strong correlation because Monitor levels decrease when the expression of a second protein is triggered, due to resource allocation [19, 54]. The same negative strong correlation can be seen here for the four input devices (see Table B.2 and Figure B.7). This trend cannot be observed for the other circuits, which include the regulated expression of different proteins, whose expression, together with the one of RFP, may provide a burden for the cell and give rise to complex RFP-GFP relationships.

The illustrated inter-relationships among RFP, GFP and growth rate confirm the usefulness of a burden monitor to measure cellular capacity instead of typically used growth rate measures. In fact, the output of a monitor cassette can not only provide an early cellular burden signal that precedes a growth rate decrease in dynamic experiments, as



## 2. Metabolic Burden modelling

---

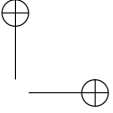
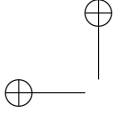
previously described [51], but also a more sensitive measure of cellular capacity, demonstrated by the clear RFP-GFP negative trend for the input devices (see Figure B.7), not reflected by growth rate changes (see Figures B.1-B.3).

### 2.3.2 Model fitting and analysis

Matlab R2007b was adopted for model fitting and analysis. Fitting was performed using the weighted least squares method via the *lsqnonlin* function. For each data point at a given HSL concentration, the weight of the  $i$ -th squared residual was set to  $w_i = \frac{1}{\sigma_i^2}$ , where  $\sigma_i$  is the standard deviation of all the biological replicates at the given HSL concentration. Biological replicates showed a relatively low variability in terms of growth rate (average CV of 12%, with a range of 1 – 36%); for this reason, the growth rate of recombinant strains was set to the average growth rate value at a given HSL concentration.

Unless differently stated, the NBM and BM were fitted sequentially: in the NBM, the Hill parameters of the four input blocks were first learned individually; then, the Hill parameters of the two NOT gates were learned individually, by setting the Hill parameters of their input devices to the values estimated in the first learning step. In the BM, the four input blocks were first simultaneously fitted to estimate the respective Hill parameters and the burden-related parameters, i.e.,  $J_{RFP}$ ,  $\Sigma_{X\lambda}$ ,  $\Sigma_{Xlac}$  and  $\Sigma_{Xtet}$ ; then, the Hill parameters of the two NOT gates, as well as their burden-related parameters,  $J_{tet}$  and  $J_{lac}$ , were learned individually as before, by setting the other parameters to the previously estimated values.

Implicit equations, commonly occurring in the BM due to the presence of protein levels on the left and right hand side (see Eqs. 2.4, 2.7-2.9), were solved using a custom Matlab script implementing the



### 2.3. Application of the models

---

fixed point method.

The NBM and BM were also fitted by using the data of all the circuits, and their parameters were all simultaneously estimated. In this case, the two models were compared via the Likelihood Ratio (*LR*) test, in which the log-likelihood value was computed as:

$$LL = -(\sqrt{2\pi} \sum_i^N \sigma_i + \frac{1}{2} \sum_i^N r_i^2) \quad (2.15)$$

considering RFP data, where  $N$  is the number of data points and  $r_i^2$  is the  $i$ -th weighted squared residual, assuming that experimental data are affected by uncorrelated Gaussian error with standard deviation  $\sigma$ .

Unless differently indicated, fixed values were used for the following parameters in all the fitting and simulation procedures:  $\lambda_{tet} = 0.0173 \text{min}^{-1}$  [86],  $\lambda_{lac} = 0.0533 \text{min}^{-1}$  [94],  $a = 0.0167 \text{min}^{-1}$  [45] and a  $G = 0.0462 \text{min}^{-1}$  [45].

A Monte Carlo approach was adopted to estimate parameter uncertainty and to propagate it throughout the model fitting procedure. For each model fitting step, 10,000 synthetic datasets were created by adding Gaussian noise (with zero mean and variance  $\sigma_i^2$ ) to the model prediction computed with the estimated parameters [84]. Negative data were set to zero. The fitting procedure was carried out for each dataset and a distribution of estimated parameters was obtained. In the stepwise procedure, parameter sets were randomly extracted from the previously obtained distribution instead of fixing them during the NOT gates model identification step, to properly propagate the uncertainty of parameter estimation to the downstream learning steps.

Univariate sensitivity analysis (i.e., performed on a single Hill parameter -  $\delta$ ,  $\alpha$ ,  $K$  or  $\eta$  - for all the blocks of a circuit) was carried out by following the Monte Carlo method illustrated above, but replacing the target parameter distribution with a Gaussian distribution with



## 2. Metabolic Burden modelling

---

the same mean and  $CV = 25\%$ . This variability was set to impose that the 95% confidence intervals of parameters ( $p$ ) are  $0.5 \times p$  and  $1.5 \times p$ , which are reasonable context-dependent variability values seen in other studies (although larger variability can be observed in distribution tails [40]). Multivariate sensitivity analysis (i.e., performed on all the four Hill parameters in all the blocks) was carried out analogously, except that the Hill parameters were extracted from a multivariate Gaussian distribution, taking into account the correlation between parameter estimates.

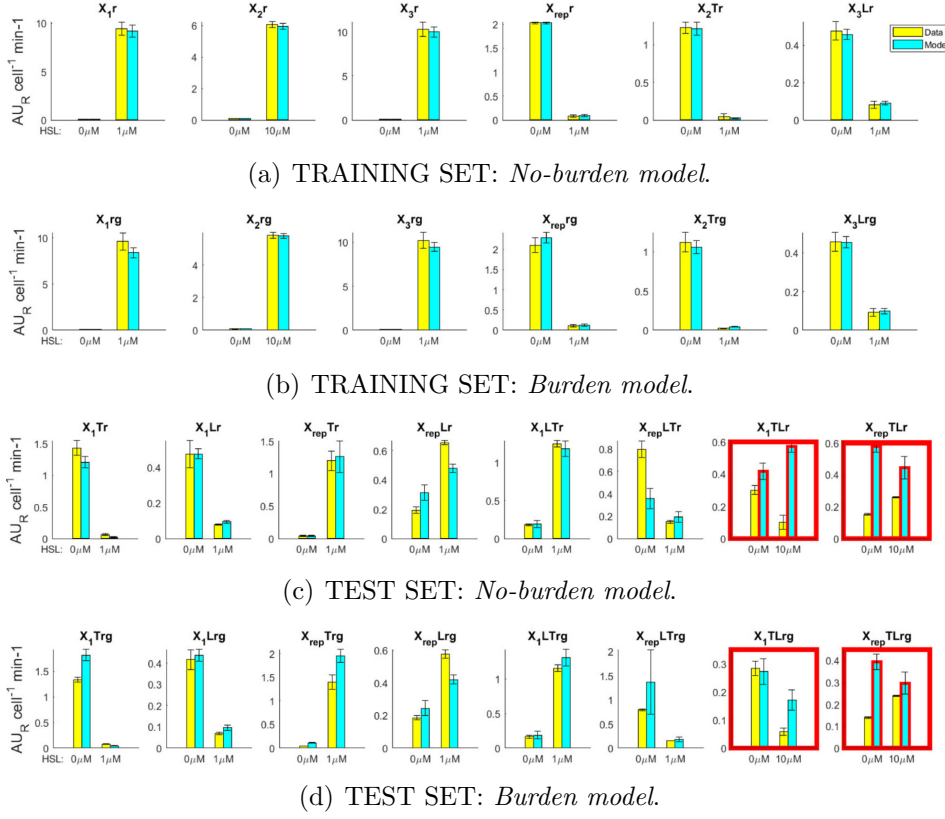
Monte Carlo model simulations, aimed to predict the test set circuits output, were carried out by extracting parameter sets from the estimated distributions. Predictions were performed by fixing the growth rates to the experimentally measured values.

### 2.3.3 Circuit predictability with no-burden model

The data from the training set circuits were fitted with the NBM (see Section 2.1). A prediction performance summary is reported from a Boolean logic and a quantitative point of view (see Figure 2.2(b) and 2.3(a), respectively). The logic behavior of all the circuits is accurately captured for all the training set circuits (Figure 2.2(a)) and for all except two test set circuits:  $X_1TLr$  and  $X_{rep}TLr$  showed an unexpectedly non-increasing and non-decreasing HSL-dependent output, anticipated above, that was not predicted by the model (see Figure 2.2(b)). The overall quantitative predictions on test set circuits showed a 0.88 correlation coefficient (see Figure 2.3(a)).



### 2.3. Application of the models



**Figure 2.2: Overall prediction performance by the two models analyzed in this work.** a-d) Logic behavior of the circuits in terms of RFP output level in vivo (yellow bars) and in silico (cyan bars) in absence of HSL and at the maximum HSL concentration tested. Results are shown for training set (a-b) and test set circuits (c-d), considering NBM (a,c) and BM (b,d). Red squares surrounding the sub-panels indicate a circuit configuration with unexpected in vivo behavior. Red edges in the in silico-predicted output bars indicate that the model is not able to predict the observed logic behavior of the circuit. Yellow bars represent the average output value and error bars represent the 95% confidence intervals of the mean; cyan bars the median predicted value and error bars represent the 95% confidence intervals calculated via Monte Carlo simulations.

## 2. Metabolic Burden modelling

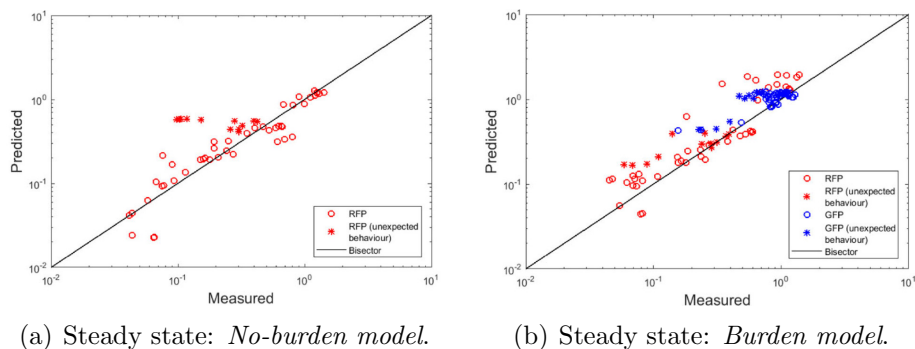


Figure 2.3: **Measured output of the circuits at all the HSL concentrations tested in this work plotted against the values predicted by the NBM (a) and BM (b).** Red and blue points represent RFP and GFP output, and are expressed as  $\text{AU}_R \text{ cell}^{-1} \text{ min}^{-1}$  and  $\text{AU}_G \text{ cell}^{-1} \text{ min}^{-1}$ , respectively. Asterisks correspond to the data of the two circuits showing unexpected in vivo behavior, while circles correspond to the data of all the other test set circuits. The solid line is the bisector line. Each point represents the average value of the in vivo measured condition, versus the median value of the corresponding model prediction.

The NBM fits the training set data accurately (see Figure B.9), and the estimated parameter values showed a relatively contained uncertainty (see Table 2.1).

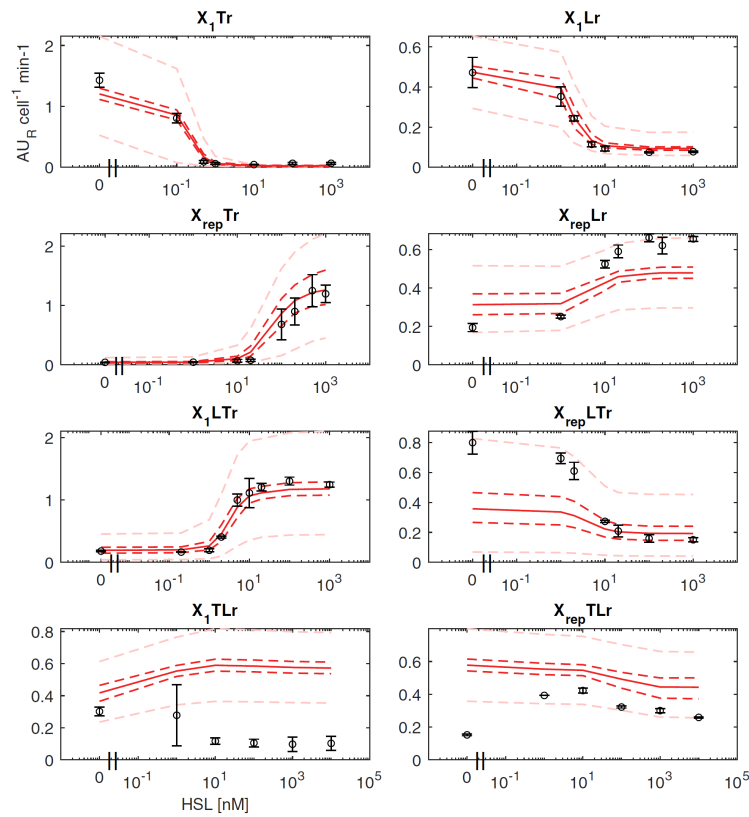
Test set data could be accurately predicted by the identified model in the  $X_1Tr$ ,  $X_1Lr$ ,  $X_{rep}Tr$  and  $X_1LTr$  constructs (see Figure 2.4). The  $X_{rep}Lr$  and  $X_{rep}LTr$  showed qualitatively correct predictions, but they underestimated the experimental data at maximum output level by up to 2 fold (see Figures 2.2 and 2.4). On the other hand, as expected, the two remaining circuits ( $X_1TLr$  and  $X_{rep}TLr$ ) did not show a correct prediction even qualitatively: a simple Hill function-based model is not able to describe their observed HSL-dependent RFP output (see Figure 2.4).

## 2.3. Application of the models

Table 2.1: Parameter description and estimated values.

Parameter	Units -==dimensionless	Estimated value (NBM, training set)	Estimated value (BM, training set)	Estimated value (BM, global fitting)
$\alpha_{X_1}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	14.63(3%)	36.17(7%)	24.33(4%)
$K_{X_1}$	nM	4.16(9%)	5.39(9%)	6.71(6%)
$\eta_{X_1}$	-	1.42(3%)	1.51(2%)	1.19(2%)
$\delta_{X_1}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.14(3%)	0.16(3%)	0.2(2%)
$\alpha_{X_2}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	9.06(1%)	26.76(3%)	20.36(3%)
$K_{X_2}$	nM	15.06(7%)	17.26(6%)	31.39(7%)
$\eta_{X_2}$	-	1.24(3%)	1.25(3%)	0.97(2%)
$\delta_{X_2}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.14(2%)	0.26(5%)	0.18(5%)
$\alpha_{X_3}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	15.81(3%)	36.9(5%)	35.67(5%)
$K_{X_3}$	nM	4.37(16%)	7.64(17%)	8.9(14%)
$\eta_{X_3}$	-	1.45(6%)	1.41(6%)	1.34(6%)
$\delta_{X_3}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.16(5%)	0.18(3%)	0.19(2%)
$\alpha_{X_{rep}}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	2.85(1%)	4.6(4%)	8.22(2%)
$K_{X_{rep}}$	nM	6.67(12%)	5.26(11%)	1.86(5%)
$\eta_{X_{rep}}$	-	1.32(12%)	1.21(9%)	0.86(2%)
$\delta_{X_{rep}}$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.13(17%)	0.22(14%)	0.09(15%)
$\alpha_T$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	3.1(1%)	3.45(3%)	4.56(2%)
$K_T$	$AU_R \text{ cell}^{-1}$	6.47(5%)	15.6(7%)	6.92(2%)
$\eta_T$	-	1.59(7%)	8.28(31%)	2.57(2%)
$\delta_T$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.03(19%)	0.22(3%)	0.21(2%)
$\alpha_L$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.63(6%)	0.56(9%)	0.76(2%)
$K_L$	$AU_R \text{ cell}^{-1}$	56.39(17%)	52.13(22%)	34.92(2%)
$\eta_L$	-	1.91(16%)	1.93(42%)	1.93(3%)
$\delta_L$	$AU_R \text{ cell}^{-1} \text{ min}^{-1}$	0.11(26%)	0.22(19%)	0.22(3%)
$\Sigma_{X_\lambda}$	-	NA*	0.2(8%)	0.36(2%)
$\Sigma_{X_{lac}}$	-	NA	1(6%)	0.56(2%)
$\Sigma_{X_{tet}}$	-	NA	0.07(23%)	0.12(11%)
$J_{RFP}$	$AU_R^{-1} \text{ cell min}$	NA	0.04(3%)	0.04(3%)
$J_{ter}$	$AU_R^{-1} \text{ cell min}$	NA	0.07(15%)	0.31(2%)
$J_{lac}$	$AU_R^{-1} \text{ cell min}$	NA	0.01(13%)	0.01(7%)
$S_m$	$AU_G^{-1} \text{ cell}^{-1} \text{ min}^{-1}$	NA	1.75(2%)	1.75(2%)

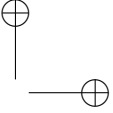
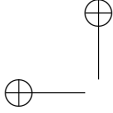
## 2. Metabolic Burden modelling



**Figure 2.4: NBM prediction of the measured HSL-dependent output in all the test set circuits without Monitor cassette.**

Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.

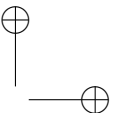
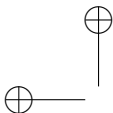
To consider the effect of parameter uncertainty on the output prediction, uncertainty was propagated via Monte Carlo approach during the fitting and simulation procedure (see Appendix B). In the training set, the resulting confidence bands of circuit outputs were very narrow,



### 2.3. Application of the models

---

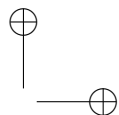
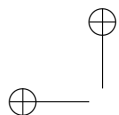
demonstrating a low uncertainty in model output, given the distribution of the estimated parameters (see Figure B.9). In the test set, the uncertainty of parameter values does not considerably affect many of the circuits: only  $X_{rep}Tr$  and  $X_{rep}LTr$  show relevant confidence bands around the central tendency value (see Figure 2.4). Steep and sensitive genetic switches, i.e., biological devices in which the transfer function shows a steep response and starts increasing (or decreasing) for very low values of its input, have been proved to promote high-entity noise propagation throughout cascades of interconnected devices [95, 96]. As a consequence, in the latter situation the output curve is sensitive to small variations of parameters and activity of the input block. In the analyzed circuits, the tetR-based block is a highly sensitive switch, since even a small activity of the upstream block can result in an output value that is significantly lower than its maximum. This is demonstrated by the  $P_{LtetO1}$  output that, in presence of an upstream block in the off-state, is considerably lower than in absence of it (see Figure B.10). Conversely, the lacI-based block exhibits a similar output value in presence or absence of an upstream device in the off-state (see Figure B.10). The described situation may explain the large output uncertainty of  $X_{rep}Tr$  and  $X_{rep}LTr$  for high RFP levels (see Figure 2.4). To confirm this effect on variability, univariate sensitivity analysis was carried out on the  $\delta$  parameters of all the used devices, considering plausible variability range values for such parameters (see Section 2.1). Results showed that a relatively small variation of the basic activity of promoters was sufficient to cause high variability in the output curves in all the devices containing  $P_{LtetO1}$  as output promoter ( $X_1Tr$ ,  $X_2Tr$ ,  $X_{rep}Tr$ ,  $X_1LTr$  and  $X_{rep}LTr$ , see Figure B.11). The results described above suggest that the robustness of the quantitative behavior of the analyzed circuits can be low in some cases, due to relevant output variations in response to small variations of the parameters. Univariate sensitivity analysis was also carried out on the other three parameters of the Hill equation describing the de-



## 2. Metabolic Burden modelling

---

vices transfer functions,  $\alpha$ ,  $K$  and  $\eta$ , to understand their effect on circuit outputs. The results, reported in Figures B.12-B.14, showed that a variation of  $K$  and  $\eta$  could explain the experimental output of  $X_{rep}LTr$  within confidence bands, but not the one of  $X_{rep}Lr$ , while the variation of  $\alpha$  is able to capture the output of both  $X_{rep}LTr$  and  $X_{rep}Lr$ . On the other hand, as expected, the parameter variations applied during sensitivity analysis could not describe the experimental output of  $X_1TLr$  and  $X_{rep}TLr$ , even by allowing the variation of all the four Hill function parameters (see Section 2.1 and Figure 2.4). Evolutionary instability issues, such as mutations occurring in the genes or regulatory parts of the circuits, may cause alterations in their output [97]. To evaluate if the output trend of  $X_1TLr$  and  $X_{rep}TLr$  was due to such alterations, phenotypic and genetic stability was assayed via specific experiments (see Figure B.20). The on- and off-state output of both circuits were found to be reversible, i.e., cultures could reproducibly change RFP output level from low to high upon induction or de-induction, depending on the circuit (Figure B.20). Only  $X_1TLr$  showed stability mutants occurring at high HSL concentrations ( $10\mu M$ ), but not at intermediate ones, although the RFP output decreases also in presence of  $10nM$  of HSL (Figure B.20). As anticipated above (and confirmed later in this work), tetR gene expression represents a burden for the cell, compared to the other proteins in the circuit, and this may explain the observed, yet low, instability occurring at high TetR protein synthesis levels. The XrepTLr circuit, on the other hand, did not show mutants. The output reversibility and the reduced presence of stability mutants only in one circuit and condition suggest that the unexpected RFP output is not due to evolutionary instability. Another issue might be enzymatic queuing, in which protein degradation complexes become a limiting resource and causes a slower degradation of all the proteins including the same specific degradation tag [98]. Although the  $X_1TLr$  and  $X_{rep}TLr$  circuits both include two proteins (TetR and LacI) with the same fast-degradation tag, simple



---

## 2.3. Application of the models

in silico simulations showed that queuing effect could not explain the observed RFP output (see Figure B.21).

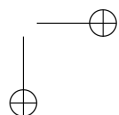
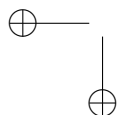
In summary, the NBM is able to successfully fit the experimental data from the training set and, by applying small variations to some of the model parameters, to quantitatively predict the output of all the set circuits with expected logic behavior. On the other hand, the output of the two test set circuits that show unexpected behavior was not captured by the NBM in any of the in-silico experiments.

The experimental data of RFP output coming from the circuit collection with the Monitor cassette was also fitted with the NBM and analogous conclusions can be drawn (see Figure B.15).

### 2.3.4 Circuit predictability with the Burden Model

Among the available models describing circuits output considering cell burden [53, 55, 56], the one proposed by Qian et al. [53] (see Section 2.1), that was also adopted in other works [19, 54], was selected. This model includes a low number of burden-related parameters, i.e., one for each gene in the circuit, while the other models, although successful in the in-silico study of different situations [51, 55, 56], required the estimation or assumption of a larger number of parameters. The model by Qian et al. can be integrated into a simple Hill function model by introducing a protein synthesis-dependent factor, which has a global negative effect on the protein expression of all the circuit. The weight of each protein synthesis term quantifies the contribution of each circuit module to the global cell load, and has been previously used as a mechanistic model-derived lumped parameter measuring resource usage [53].

Analogously to what was performed for the NBM, the data from the training set circuits were fitted with the BM, by considering both RFP and GFP, representing the circuit output and the burden measures, respectively, and the output of test set circuits was finally predicted





## 2. Metabolic Burden modelling

---

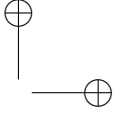
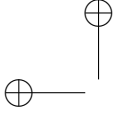
(see Figure 2.2(d) and 2.3(b) for an overview of the logic and quantitative prediction performance).

Among the training (see Figure 2.2(b)) and test set circuits (see Figure 2.2(d)), only  $X_{rep}TLr$  shows a non-correct logic behavior prediction, while all the other circuit configurations could be captured. In particular, it is worth noting that the BM is able to predict the output of  $X_1TLr$ , which could not be predicted by the NBM.

The training set data were fitted by the BM with reasonable accuracy (see Figures B.16-B.17). In particular, the model showed excellent quantitative accordance with RFP experimental data (see Figure B.16), as it was observed above for the NBM; GFP data were all well fitted by the BM, except  $X_{reprg}$  and  $X3Lrg$ , for which the model showed a slightly lower descriptive capability to capture the measured data than for the other circuits (see Figure B.17).

The estimated parameter values showed significant deviations from the ones obtained via NBM (see Table 2.1). The most remarkable differences can be observed for the  $\alpha$  parameter values of almost all the devices. These values are, in general, higher when estimated via the BM. This trend was expected, since in a limited cell resources framework the devices are globally burdened, and the estimated  $\alpha$  values are linked to the maximum achievable activity, which may not be reached in any of the tested conditions [53]. For instance, the  $X_1rg$  device shows a 2.5 fold difference in the  $\alpha$  values between NBM and BM, meaning that the observed activity at full induction reached by the  $P_{lux}$  promoter in this device is much lower than maximum attainable one, which was 2.5 fold higher, due to the high RFP expression rate in this induction condition. A lower fold change is observed for devices characterized by lower activity in the on-state (e.g.,  $X_{reprg}$ ), since the  $\alpha$  values estimated via the NBM are close to the maximum attainable ones, estimated by the BM. In addition to the Hill-related parameters, the BM includes resource usage parameters for each gene in the circuit [53]. The estimated values of the LuxR protein contribution



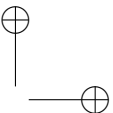
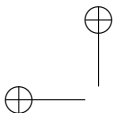


### 2.3. Application of the models

---

to cell burden ( $\Sigma_{X\lambda}$ ,  $\Sigma_{lac}$  and  $\Sigma_{tet}$ , corresponding to the expression systems driven by  $P_R$ ,  $P_{LlacO1}$  and  $P_{LtetO1}$ , respectively) showed that LuxR expression alone in the input devices causes cell burden and decreases circuit output by up to 50%, with the cassette driven by  $P_{LlacO1}$  giving the highest burden and the one driven by  $P_{LtetO1}$  giving the lowest one (see Table 2.1). This effect can be observed in the GFP output curves of the four input blocks in absence of HSL (see Figure B.17). In these four circuits in this condition, RFP expression is negligible and the only protein having a significant contribution to cell burden is LuxR; as a result, GFP output level is inversely correlated with the corresponding  $\Sigma$  value. This result was unexpected, since  $P_{LlacO1}$  has a lower activity than  $P_{LtetO1}$  in the used chassis, and a lower burden value for it was expected (given identical RBSs upstream of luxR gene). The estimated resource usage parameter values of RFP, TetR and LacI proteins ( $J_{RFP}$ ,  $J_{tet}$  and  $J_{lac}$ , respectively) enable to conclude that, with the used RBSs, TetR expression causes the highest cell load, while LacI the lowest one. Such values can be useful to evaluate the working boundaries in the bottom-up design of reduced-burden circuits, as demonstrated previously via different approaches [51].

Quantitative prediction results on the RFP output of test set circuits showed that the overall performance of the BM (correlation coefficient of 0.87, see Figure 2.3(b)) is analogous to the one of the NBM. Considering individual circuits,  $X_1Lrg$ ,  $X_{rep}LTrg$  and  $X_1TLrg$  showed a good prediction with data consistent with the confidence bands of the model (see Figure 2.5). Importantly,  $X_1TLrg$  was one of the two circuits whose output could not be correctly predicted by the NBM; the other circuit ( $X_{rep}TLrg$ ), however, still behaved unpredictably. A slight over- ( $X_1Trg$  and  $X_{rep}Trg$ ) or under-estimation ( $X_{rep}Lrg$ ) of the experimental data maximum output level was observed for three circuits, with an error up to 1.4 fold. The predicted output of  $X_1LTrg$  showed a slightly anticipated switch point and an over-elongation at

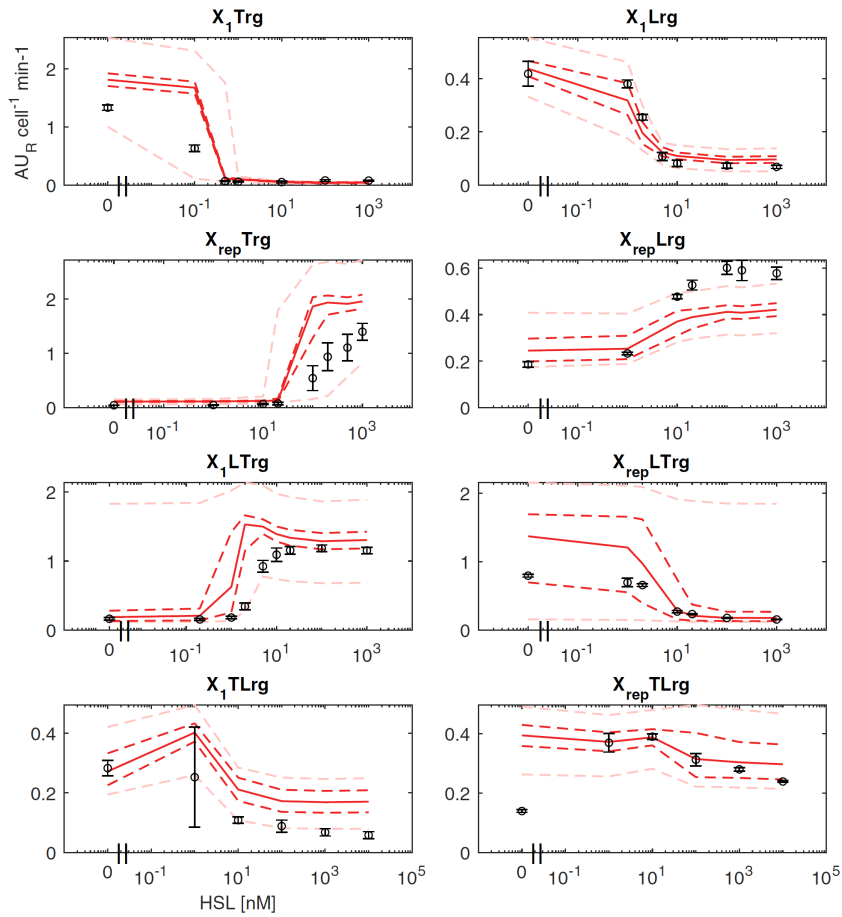


## 2. Metabolic Burden modelling

---

intermediate HSL concentrations that is not observed in experimental data. Considering the Monitor output of the same circuits (see Figure 2.6), GFP showed an overall over-estimation of the experimental data, with a lower prediction performance than RFP (0.71 correlation coefficient, see Figure 2.3(b)), suggesting that additional modelling work is needed to improve the predictive capability of burden-related models.

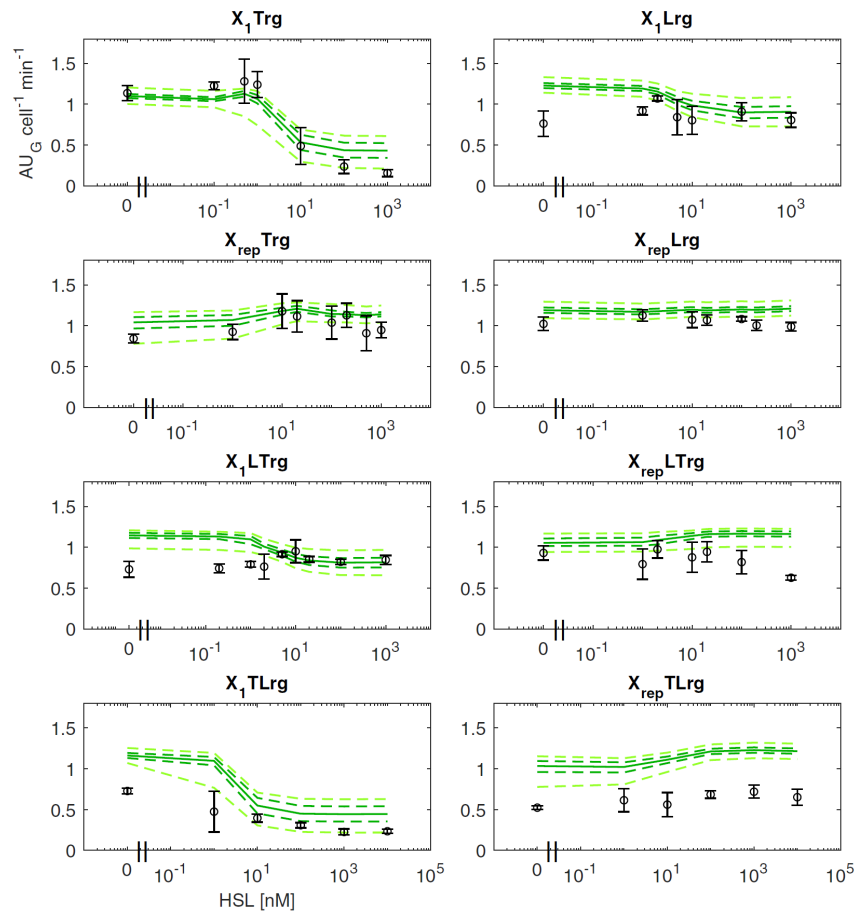
### 2.3. Application of the models



**Figure 2.5: BM prediction of the measured HSL-dependent RFP output in all the test set circuits with Monitor cassette.**

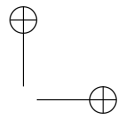
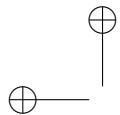
Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis

## 2. Metabolic Burden modelling



**Figure 2.6: BM prediction of the measured HSL-dependent GFP output in all the test set circuits with Monitor cassette.**

Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark green lines are the 95% confidence bands of the output distribution. Dashed light green lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis



---

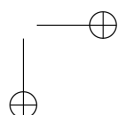
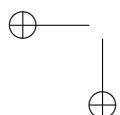
## 2.3. Application of the models

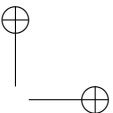
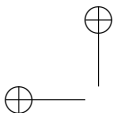
As it was carried out for the NBM, sensitivity analysis was performed. Results are shown for a multivariate sensitivity analysis (see Figures 2.5,2.6). As expected, confidence bands are higher than for the NBM, since the BM has more parameters that can vary in a multivariate fashion. Results showed that all the RFP data can be explained by confidence bands, except  $X_{rep}TLrg$ , leading to analogous conclusions drawn for the NBM: parameter variations of plausible entity can capture all the experimental data except for circuits showing qualitatively inconsistent predictions.

In summary, the BM does not improve the quantitative prediction performances of the analyzed circuits with expected behavior, over the NBM. However, it correctly predicted the output of one of the two circuits with unexpected output behavior, not predicted by the NBM, and, in addition, enabled the estimation of burden-related parameters that support the rational design of synthetic circuits.

### 2.3.5 Model fitting using all the available experimental data

The BM was also identified by using all the available data of the training and test sets, in order to demonstrate the descriptive capability of the model to simultaneously fit all configurations, considering both RFP and GFP data as before (Figures B.18-B.19). The NBM was used to fit the RFP experimental data as a term of comparison, but its descriptive capability was significantly worse than the one of the BM (Figure B.18). Model comparison demonstrated a significantly higher fitting performance for the BM (LR test, p-value < 0.05). The parameter estimates resulting from the BM fitting are reported in Tab 2.1. Their values confirm the conclusions about burden contribution levels for all the circuit proteins, since the  $\Sigma$  and  $J$  parameters have the same ranking as before, despite a large variation was observed for





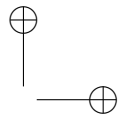
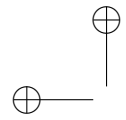
## 2. Metabolic Burden modelling

---

some of them (e.g.,  $J_{tet}$  which was  $> 4$  fold higher than before). Fitting results in Figures B.18-B.19 show that the output of some configurations are over- or under-estimated up to 1.5 and 2.1 fold for RFP and GFP, respectively, although the outputs of all circuits is qualitatively captured. Parts activity variation upon interconnection, an open problem in synthetic circuit design, can explain the observed changes between individual configurations. Specifically, despite cell burden modeling can explain some unexpected phenomena in bottom-up designed circuits, other context-dependent effects still have to be quantitatively elucidated, e.g., promoter transcription variation caused by diverse flanking DNA sequences in different configurations [20]. The obtained results suggest that the modeling of cell burden significantly improves circuit output description capability, but the context-dependent behavior of the assembled devices must be taken into account in future studies to predict new designed configurations more accurately.

### 2.3.6 Fixing non-functional cascades via rational design

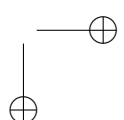
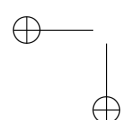
To further demonstrate the usefulness of BM in the rational design of circuits, new variants of the two cascades with unexpected behavior were designed and constructed to correct their HSL-dependent logic function. Based on the estimated resource usage parameters via BM (see Table 2.1), *tetR* was identified as the gene causing the highest load among the three regulated modules. For this reason,  $X_1T_{Lr}$  and  $X_{rep}T_{Lr}$  were mutagenized to decrease the translation efficiency of *tetR*, obtaining  $X_1T_{wLr}$  and  $X_{rep}T_{wLr}$  (Figure 2.8). Analogously, the training set circuit  $X_2Tr$  was mutated, obtaining  $X_2T_{wr}$ , to enable the learning of the new *tet*-based NOT gate transfer function (Figure 2.7). The use of a weaker RBS (BBa\_B0033 instead of BBa\_B0031) up-



---

### 2.3. Application of the models

stream of tetR successfully modified the individual NOT gate transfer function (Figure 2.8(a)), resulting in a less sensitive switch, as indicated by the  $K_T$  parameter that increased by 10 fold. Cascades with this modified tet-based NOT gate are expected to exert a lower cell load than their previous design when TetR protein is over-expressed, thereby restoring the correct functioning of the interconnected gates. Experimental results showed that this RBS change yielded circuits with expected increasing ( $X_{1TwLr}$ ) and decreasing ( $X_{repTwLr}$ ) behavior as a function of HSL (Figures 2.8(b)-2.8(c)). In addition, considering all the HSL concentrations tested, the new circuits had about 2 fold higher growth rate than the previous ones (data not shown). The RFP output of the realized circuits was also accurately predicted by the NBM, using the same training set as above (see Figure 2.1) except  $X_2Tr$  that was used instead of  $X_2Tr$  (Figures 2.8(b)-2.8(c)). This result demonstrated that, after the attenuation of the main burden source, the two circuits with previously unexpected behavior could be not only fixed in terms of qualitative behavior, but also their quantitative HSL-dependent output could be successfully captured via traditional NBM.



## 2. Metabolic Burden modelling

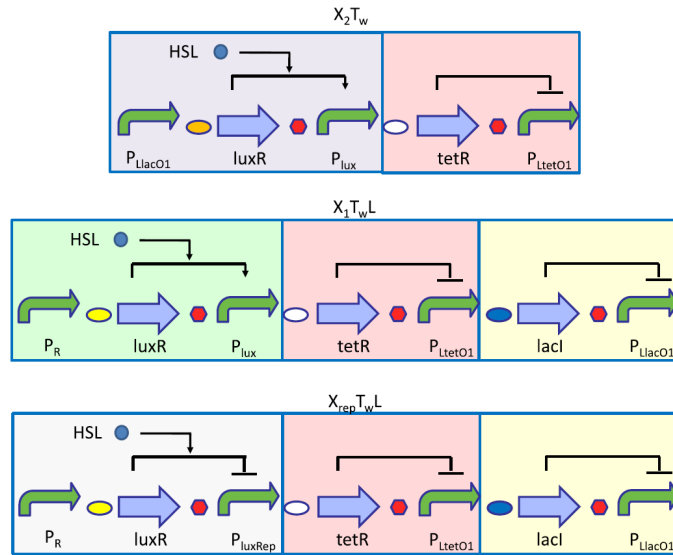


Figure 2.7: **Re-designed circuit variants schema, without Monitor cassette.** All of them are available with an RFP expression system downstream of the output promoter (indicated in the text with the r suffix). Symbols are described in Figure 2.1, except white ovals that represent a weak RBS (BBa\_B0033), used in these circuits to decrease  $tetR$  expression.



### 2.3. Application of the models

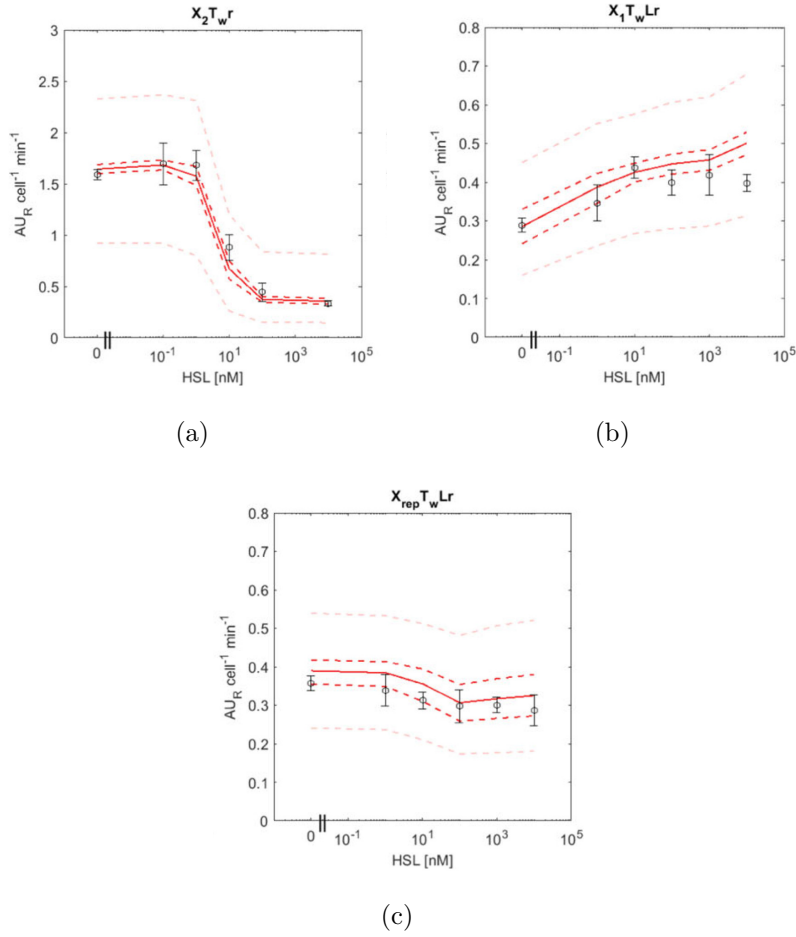
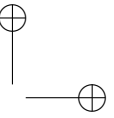
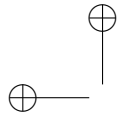


Figure 2.8: **Analysis via NBM of the measured HSL-dependent output in the re-designed circuit variants, without Monitor cassette.** **b-c)** Prediction of the test set circuit  $X_1\text{TwLr}$  and  $X_{rep}\text{TwLr}$ , respectively. In panels **a-c**, circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.

## 2. Metabolic Burden modelling

---

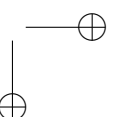
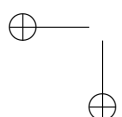


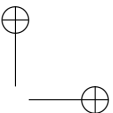
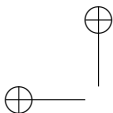
# Chapter 3

## Engineering low-burden synthetic circuits via CRISPRi

In the previous chapter, a mathematical approach to handle metabolic burden in rational design was presented, as well as its application to identify and replace components with high resource usage. To increase the probability of obtaining a working synthetic circuit meeting the design specifications, a novel toolbox for rational design of synthetic genetic circuits with reduced metabolic load, based on the CRISPR-interference mechanism, is here proposed instead. The constructed and tested devices aim to exploit the reduced translational demand of the CRISPR-dCas9 system as an alternative to repress promoter transcriptional activity, allowing to both avoid the usage of resource-demanding transcriptional regulators and expand the pool of repressible promoters to possibly every DNA region bound by an RNA polymerase (encoding a PAM sequence).

Strain, cloning protocols and reagents used, as well as circuit characterization and data processing procedures are mostly the same as the ones described in Appendix B.1. Exceptions included GFP analysis,





### 3. Engineering low-burden synthetic circuits via CRISPRi

---

inducers, microscopic assays, sgRNA<sup>1</sup> design and mutagenesis protocols; their description along with additional materials and methods are reported in Appendix C), along with a list of the circuits and parts used in this study with relative names and BioBrick<sup>TM</sup> identifiers (see Table C.1). In this chapter, a deep characterization of CRISPRi devices is provided by first investigating a suitable expression level for dCas9 (Section 3.1), then by selecting proper expression systems for the sgRNAs (Section 3.2), by characterizing the sgRNAs repression capability in different contexts (Section 3.3) and finally by studying the interplay between sgRNA repression and promoter strength (Section 3.4).

## 3.1 dCas9 expression cassette

### 3.1.1 dCas9 toxicity

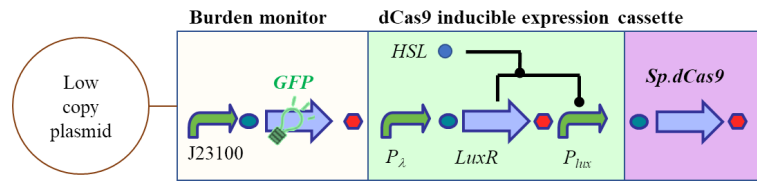
For the interference system to be feasible, a threshold in dCas9 expression level needed to be found, avoiding detrimental effects to cell growth; this means that its expression should not have a high translational demand and the synthesized protein itself should not be toxic to the cell. An HSL-inducible dCas9 expression cassette was therefore assembled downstream the GFP-based *capacity monitor* described in Section 2.1.1 (namely AEdCas9, see Table C.1 and Figure 3.1), and growth and GFP expression were compared to a control strain that contained solely the capacity monitor as exogenous DNA, named A37. To achieve a strong transcription rate of the DNA downstream its transcription start site (TSS), the  $P_{lux}$  promoter required binding of **LuxR** protein that was under constitutive expression, and which itself needed activation by an exogenous inducer, **HSL**. In this way, the

---

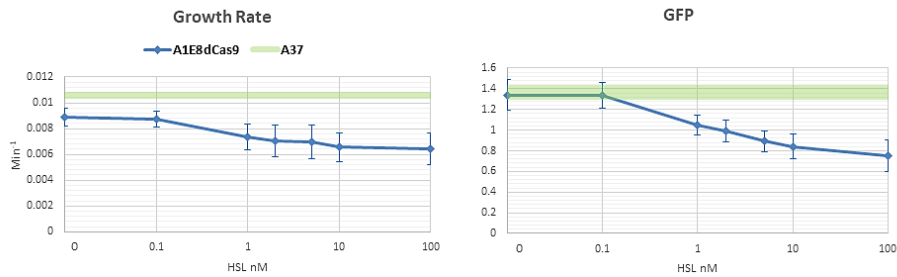
<sup>1</sup>As of now, *tracr* is intended as the entire sgRNA, excluding the complementary 20 nucleotides called gRNA.

### 3.1. dCas9 expression cassette

addition of HSL was necessary and sufficient to bind and activate the LuxR protein and promote transcription from the inducible promoter.



(a) Circuits schema.

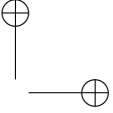
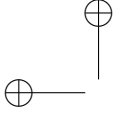


(b) Growth rate and GFP expression comparison.

**Figure 3.1: Toxicity measurements of dCas9 expression.** **a.** A capacity monitor (A37) expressed GFP alone indicating the maximum possible expression level. AEdCas9 expresses dCas9 under the  $P_{lux}$  promoter, which is activated upon binding of LuxR:HSL. **b.** The x-axis represents the HSL induction proportional to the amount of dCas9 inside the cell. Data are reported as mean values over at least 3 biological replicates while error bars (line width for A37) represent the 95% confidence intervals of the mean.

A decrease in growth rate and GFP expression with increasing expression of dCas9 can be seen starting from an induction level between 0.1 and  $1nM$  of HSL. The decrease in fitness was assumed to be a result of cellular resource limitation, while a certain level of toxicity could be deduced by the decreased growth rate compared to A37, even with a leakiness of dCas9 production (i.e., without HSL induction).

To determine whether the burden imposed from our constructs also



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

lead to toxic effects for the cell, like changes in cell morphology, images of:

- a no burden strain (the host strain TOP10);
- a control construct used to measure maximal reporter gene expression with a constitutive expression level of dCas9 close to the threshold found as described above (AYgPtet+J116dCas+J119H);
- the maximally induced dCas9 cassette (AEdCas9 - 100nM $HSL$ );

were taken and compared.

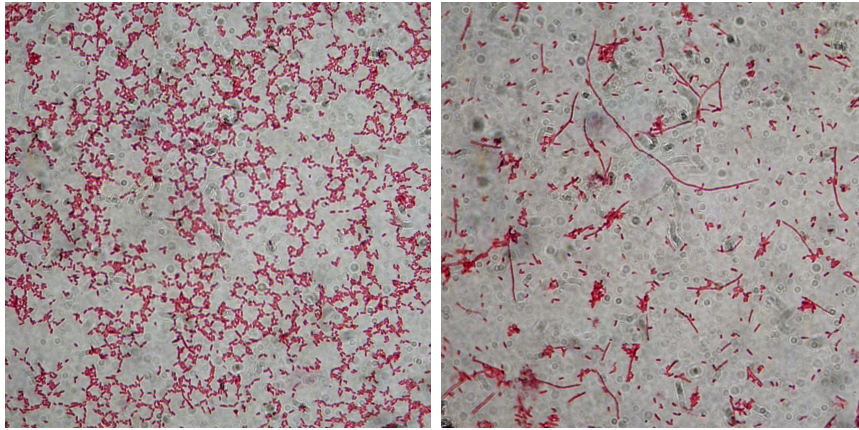
Microscope images showed that indeed high expression of dCas9 results in a change in morphology of the cells, from rod-shaped to filamentous. This is in accordance with data recently published in literature [99], which also concludes that high level expression of dCas9 causes abnormal morphological changes in *E.coli*. A construct expressing dCas9 through an optimized constitutive expression cassette (described above) showed very few filamentous cells surrounded by colonies of rod-shaped bacteria, resulting in a phenotype almost identical to the control strain and thus showing that a non-toxic synthesis rate was reached (see Figure 3.2).

It was confirmed that dCas9 is toxic to the cell when expressed at high concentrations, affecting growth rate, expression levels (related to available cellular resources for translation), and morphology. However, there was a window of inductions that shown negligible toxic effects and in which the inhibition of transcription capability could still be studied.

#### 3.1.2 Repression efficiency

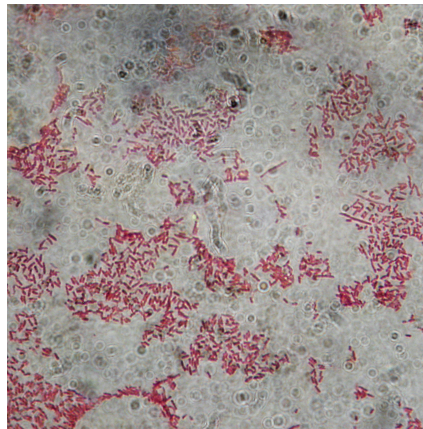
Once the potential toxicity of dCas9 was characterized, its efficacy at repressing target reporter genes depending on the amount of protein and guide needed to be evaluated. Despite in [46] a similar study

### 3.1. dCas9 expression cassette



(a) *TOP10*.

(b) *AEdCas9 full induction*.

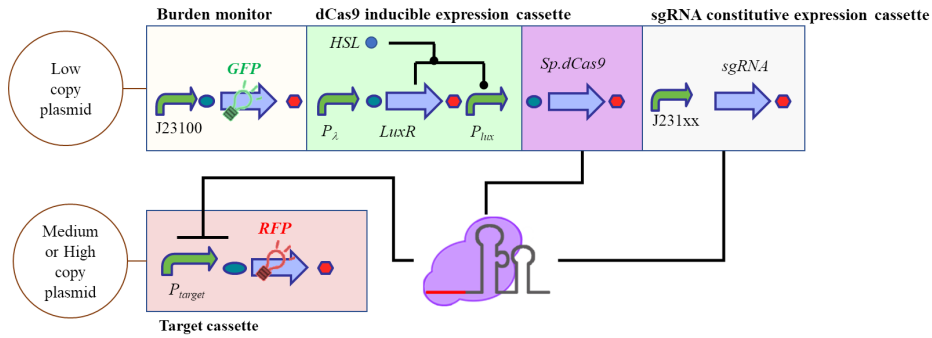


(c) *AYgPtet+J116dCas+J119H*.

**Figure 3.2: Microscopic Images of *E. coli* Strains.** Images were taken using the Leica bright field microscope using the 100x/1.25 oil immersion objective representing a no burden **a.** and high burden **b.** strain, and subsequently compared to a strain constitutively expressing dCas9 in an optimized amount **c** to visualize the burden imposed.

### 3. Engineering low-burden synthetic circuits via CRISPRi

was conceived, no standard units to express promoter strength were used nor constitutive optimal dCas9 expression cassette were eventually obtained, hampering the re-usage of those results for this study. Six new constructs were combinatorially built on that aim, by combining upstream of the dCas9 expression cassette described above three different constitutive promoters - with graded strengths but identical transcription start sites at +1 (giving rise to identical transcripts) - with two different sgRNAs: gPtet and gPlac (see Figure 3.3).



**Figure 3.3: CRISPRi with inducible dCas9 and constitutive sgRNA configuration.** General configurations of the CRISPRi systems with inducible dCas9 and constitutive sgRNA expression cassettes; in the sgRNA expression cassette, xx stands for 16/00/19 which are the codes of the promoters used in this study.

Guides were designed to target the previously described  $P_{LtetO1}$  and  $P_{LlacO1}$  promoters respectively, which were assembled on co-transformed plasmids in medium or high copy, driving the expression of an RFP reporter gene. As negative controls, the same system with switched sgRNAs were used (i.e., gPtet to target  $P_{LlacO1}$  and vice versa); due to the orthogonality of the two sgRNA binding sites, the response was supposed to be indicative of the maximum possible expression of RFP in the absence of a complementary sgRNA, but with the same proteins and sgRNAs amounts synthesized in the cell. Through these circuits,



### 3.1. dCas9 expression cassette

it was possible to monitor growth rate, GFP, and RFP expression in a single time course; growth rate and GFP signal were considered as representative of cellular burden and general cell status, while RFP was dependent to the repression capability of the sgRNA:dCas9 complex. The choice of the constitutive promoters with different strengths (such that BBa\_J23116 < BBa\_J23100 < BBa\_J23119, [84]) allowed to study the effect of increasing guide expression for several levels of dCas9 in the cell, which was tuned by varying the concentration of HSL in the medium, as before.

#### Inducible dCas - $P_{LtetO1}$ medium copy target

Firstly, a significant decrease in RFP output, was observed even at no induction of dCas9, indicating the high repression efficiency by the CRISPRi complex. This was hypothesized to be a result of the transcriptional/translational leakage of the HSL-based expression cassette system driving the expression of dCas9, resulting in a basal unavoidable synthesis of the protein (see Figure 3.4). Nonetheless, an increase

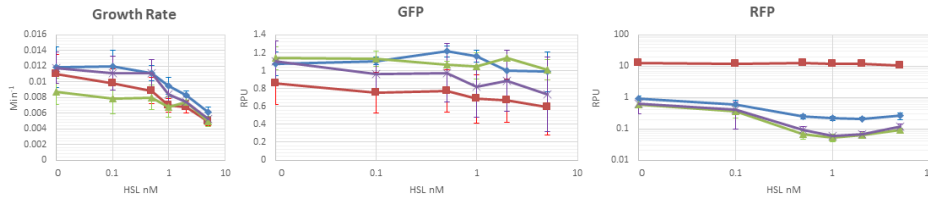
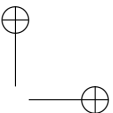
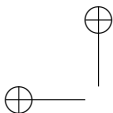


Figure 3.4: **HSL-inducible dCas9 expression system targeting medium copy  $P_{LtetO1}$** . Growth rate, GFP and RFP expression of the construct. Lines in blue, green, violet and red represent respectively weak, medium, strong sgRNA expression and control. On the x-axis, the HSL induction proportional to the amount of dCas9 in the cell. The relative amount of the sgRNA, constitutively expressed, depends on the number in the name of the circuit, namely  $116 < 100 < 119$ ; the control circuit bore a non-targeting sgRNA (gPlac) driven by the medium strength promoter BBa\_J23100. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

of RFP repression was observed for increasing HSL concentrations (re-



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

lated to dCas9 level) and strength of constitutive promoter (related to sgRNA level). However, repression was so strong that, on a linear y-axis scale, all three constructs exhibited almost no RFP signal (data not shown). Switching the scale to logarithmic, the expected conservation of repression capability in terms of strengths ranking of the different sgRNA-driving promoters was confirmed for several levels of dCas9 in the cell.

### 3.1. dCas9 expression cassette

In other words, the strongest promoters BBa\_J23100 and BBa\_J23119 showed the lowest RFP expression, while BBa\_J23116 (the weakest) showed the highest expression of RFP. This confirmed the hypothesis that availability of both dCas9 and gRNA affected the strength of repression.

#### Inducible dCas - $P_{LlacO1}$ medium copy target

For the  $P_{LlacO1}$  driven target in medium copy (namely, E52), the trend were confirmed and strengthened, being the ranking repression exactly the same of the strength of the three constitutive promoters used (Figure 3.5). However, in this case, the difference between con-

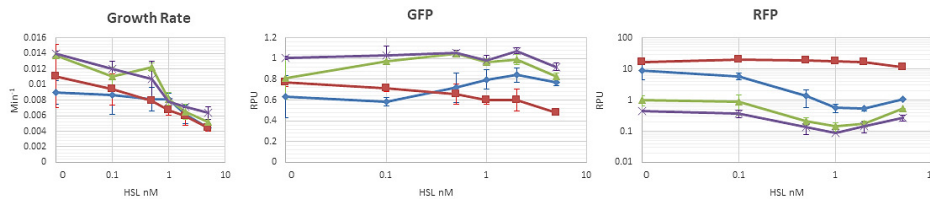
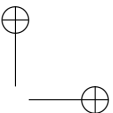
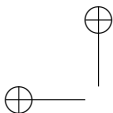


Figure 3.5: **HSL-inducible dCas9 expression system targeting medium copy  $P_{LlacO1}$ .** Growth rate, GFP and RFP expression of the construct. Lines in blue, green, violet and red represent respectively weak, medium, strong sgRNA expression and control. On the x-axis, the HSL induction proportional to the amount of dCas9 in the cell. The relative amount of the sgRNA, constitutively expressed, depends on the number in the name of the circuit, namely  $116 < 100 < 119$ ; the control circuit bore a non-targeting sgRNA (gPtet) driven by the medium strength promoter BBa\_J23100. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

trol and non induced functional strains was less marked (despite the needs to plot RFP data again in log-scale); this added to our knowledge that guide sequence, and thus energy required for DNA:RNA heteroduplex formation, also played a role in the repression capability of the CRISPRi complex, which needed to be taken into account in the design of possible circuitry. Nonetheless, for dCas9 levels correspondent to HSL inductions higher than 0.5 - 1nM, it was observed that



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

growth rate decreased and the RFP increased. This fits with what observed in the previous Section 3.1.1 where both GFP and growth rate measurements started to decrease around the same threshold. The GFP expression of AEd116gPlac targeting E52 was informative of the relation between burden and gene expression. Very low constitutive guide expression by the BBa\_J23116 promoter, allowed us to observe the burden imposed by RFP expression with increasing repression. At no induction, RFP expression was sufficiently high to affect GFP expression, however when the lux promoter was half activated ( $[HSL] = 1 \div 2\mu M$ ), RFP reached a minimum value while GFP reached a maximum value. This suggested that indeed the system set up was capable of conceptualizing burden as well as repression efficiency.

#### Inducible dCas - $P_{LtetO1}$ high copy target

To better understand the CRISPRi system in more cases and with data observable on a wider, more comprehensive scale, target reporter genes were also placed in high copy vectors; these system were designed to produce larger quantities of RFP and therefore to impose a more observable burden in terms of growth rate and GFP expression levels (Figure 3.6). The control of  $P_{LtetO1}$  in high copy (I13521) generated an RFP output of around  $20RPU$ , and as expected the lowest GFP signal of the group.

GFP expression exhibited a slight increase at half induction of dCas9, partially attributed to the decrease in RFP expression. In other words the burden imposed by RFP expression was relieved by the CRISPRi complex, allowing increased GFP expression. However, further induction of dCas9 resulted again in the characteristic decrease of GFP expression and growth rate due to the high metabolic load of the dCas9 protein. At maximum induction of  $5nM$ , a complete repression of the the two constructs that bore the intermediate (BBa\_J23100)

### 3.1. dCas9 expression cassette

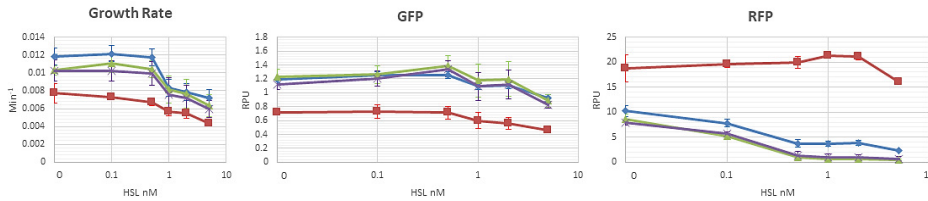


Figure 3.6: **HSL-inducible dCas9 expression system targeting high copy  $P_{LtetO1}$ .** Growth rate, GFP and RFP expression of the constructs. Lines in blue, green, violet and red represent respectively weak, medium, strong sgRNA expression and control. On the x-axis, the HSL induction proportional to the amount of dCas9 in the cell. The relative amount of the sgRNA, constitutively expressed, depends on the number in the name of the circuit, namely  $116 < 100 < 119$ ; the control circuit bore a non-targeting sgRNA (gPlac) driven by the medium strength promoter BBa\_J23100. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

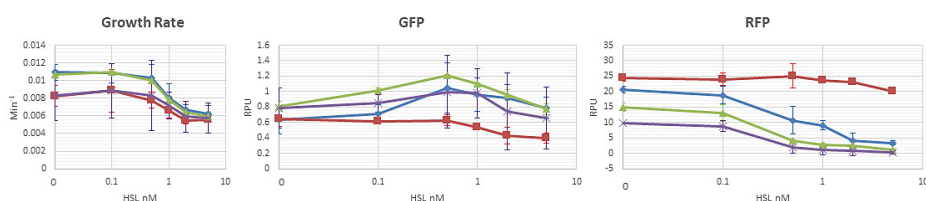
and strong (BBa\_J23119) promoters upstream the sgRNA was observed, while for the weak promoter BBa\_(J23116) complete repression was not attained, even if the decreasing trend suggested that a higher dCas9 induction could lead to complete repression (in absence of possible toxic effects). The absence of an increasing RFP level for inductions higher than  $1nM$  that was seen in medium copy target circuits could be explained though the general high burden cell state and therefore with a possible variation in RFP protein expression.

#### Inducible dCas - $P_{LlacO1}$ high copy target

Testing the the system with the reporter gene RFP driven by  $P_{LlacO1}$  in high copy plasmid (A33) as CRISPRi target (Figure 3.7), more conclusions could be drawn due to the higher strength of the  $P_{LlacO1}$  compared to  $P_{LtetO1}$ : the negative control for this system reached a maximum value of approximately  $25RPU$ , against the  $20RPU$  achieved by  $P_{LtetO1}$ .

Growth rate followed the general trend of the other circuits described above, confirming the toxicity that high dCas9 expression imposed on

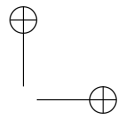
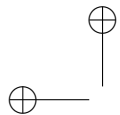
### 3. Engineering low-burden synthetic circuits via CRISPRi



**Figure 3.7: HSL-inducible dCas9 expression system targeting medium copy  $P_{LlacO1}$ .** Growth rate, GFP and RFP expression of the construct. Lines in blue, green, violet and red represent respectively weak, medium, strong sgRNA expression and control. On the x-axis, the HSL induction proportional to the amount of dCas9 in the cell. The relative amount of the sgRNA, constitutively expressed, depends on the number in the name of the circuit, namely  $116 < 100 < 119$ ; the control circuit bore a non-targeting sgRNA (gPtet) driven by the medium strength promoter BBa\_J23100. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

the cells. As expected, GFP at zero induction is below the one observed with the medium copy target (see Figure 3.5); its expression increased steadily with increasing induction of dCas9, and reached a maximum value at an induction of  $[HSL] = 0.5nM$  for all, which corresponded to a high enough repression of the RFP target. Although expression of RFP was repressed and led to an increase in GFP signal output, the high expression of dCas9 still caused an imbalance in cellular growth and synthesis (GFP signal decreased together with RFP and growth rate for inductions higher than  $0.5nM$  of HSL). Again, for RFP expression, promoter strength that drove the sgRNA determined repression efficacy, as depicted in the graph showing RFP signal output. Moreover, in this case RFP levels of the functional constructs corresponding to a zero induction of HSL were closer to the maximum output generated by the control; they then decreased substantially in a small range if dCas9 induction, and finally reached full repression for sgRNAs driven by BBa\_J23100 and BBa\_J23119, while BBa\_J23116 driven guide attained a 10x decrease in RFP expression.

From the data gathered in these experiments, it could be firstly con-



---

## 3.2. sgRNA expression cassette design

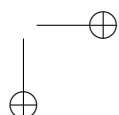
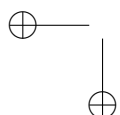
cluded that the CRISPRi system was functional and efficient even for low expression levels of dCas9. Also, a higher efficiency was shown from  $P_{LtetO1}$  sgRNA in repressing its target promoter compared to the  $P_{LlacO1}$ -gPlac sgRNA pair.

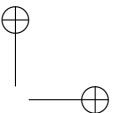
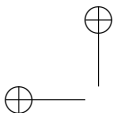
It is worth to notice that through these experiments, a range of inductions in which dCas9 expression level showed a great efficiency without excessively affecting GFP signal and growth rate (namely  $0.2 \div 0.5 nM$  HSL) were highlighted. However, the system was very sensitive to HSL variations with a range of inductions to achieve strong repression without burden effect that was too small to be useful in rational design of complex circuits in a tunable manner.

A new dCas9 expression cassette was therefore conceived by comparing data obtained in Figure 3.1 with expression potentials of the three constitutive promoters (BBa\_J23116, BBa\_J23100, BBa\_J23119) in low copy (pSB4C5) and medium copy (pSB3K3) vectors [data not shown]. The combination of BBa\_J23116 and medium copy plasmid to express dCas9 gene (namely J116dCas, dCas9 constitutive expression cassette in Figure 3.13) was expected to lead a strong enough constitutive expression to fully repress a promoter in a high copy plasmid, without negatively affect growth rate or overload the cell (also confirmed by cell morphology imaging - see Figure 3.2).

## 3.2 sgRNA expression cassette design

Once the dCas9 expression level was fixed at a minimum effective concentration, tunable and modular expression cassettes to be used as genetic knobs - acting on the amount of sgRNA expressed in the cell - were necessary to obtain a regulated sgRNA production device. Two different logic systems were chosen to effect the expression of sgRNAs upon the addition of an inducer. The first system, *M12 Inv* (Figure 3.8), expressing RFP and bearing the capacity monitor always in a





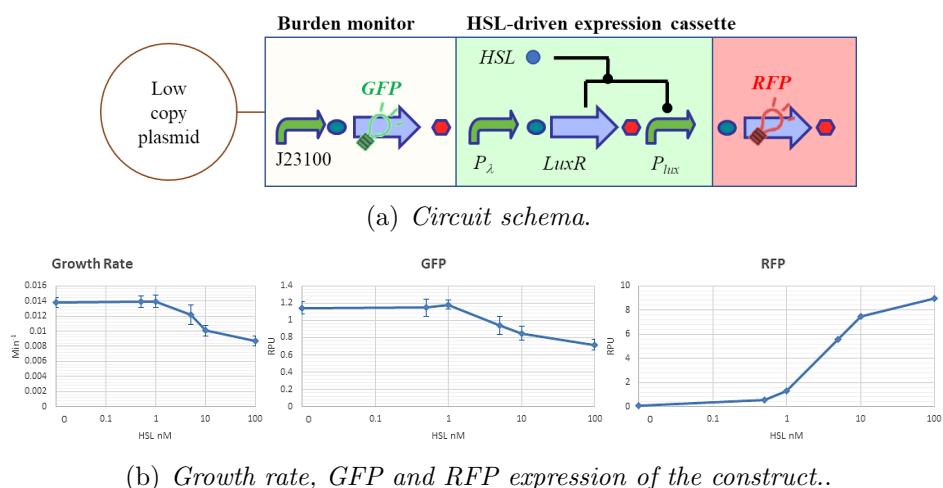
### 3. Engineering low-burden synthetic circuits via CRISPRi

---

low copy plasmid vector (pSB4C5), was based on the luciferase quorum sensing network of *Vibrio fischeri*, analogously to the one used to induce dCas9 in Section 3.1.1. Growth rate and GFP expression remained stable with increasing inductions with HSL; however, a slight decrease in both signals at highest inductions was observable. This might be explained by the increasing expression of RFP, which reached burdensome levels at  $[HSL] = 5nM$ , corresponding to the concentration at which both growth rate and GFP signal started decreasing. RFP expression ranged from zero to 9 RPU and started to give a readable output at  $0.5nM$  HSL concentration.



### 3.2. sgRNA expression cassette design

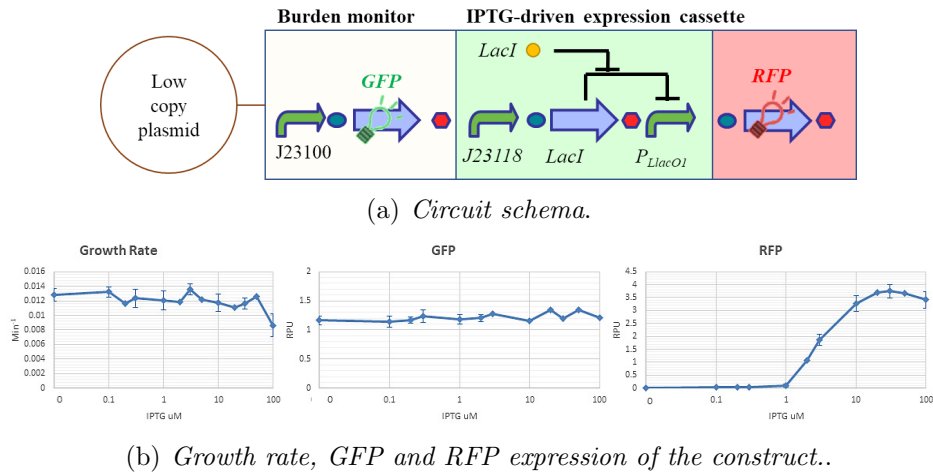


**Figure 3.8: LuxR:HSL inducible promoter characterization.** a) The systems express an RFP to characterize the synthesis rate of the HSL system driving its transcription, along with the usual GRP monitor cassette. b) The x-axis represents HSL inducitor concentration. LuxR protein was activated by the exogenous molecule HSL. Subsequently, the activated transcription regulator was able to bind operator regions of the  $P_{lux}$  and recruit polymerases for transcription. The expression levels of downstream genes is thus dependent on the concentration of HSL present in the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

The second inducible system, named *AY*, assembled on a low copy plasmid as well, affected the expression of the downstream DNA relying on parts isolated from the *E.coli* Lac operon. In specific, the Lac operon repressor, *LacI*, was under constitutive expression and kept transcription levels of the  $P_{LlacO1}$  promoter very low. *LacI* was also able to bind an exogenous regulator, *IPTG*, which changed its conformation and relieved repression on  $P_{LlacO1}$  thus allowing downstream gene transcription (see Figure 3.9).

The maximum RFP signal output  $P_{LlacO1}$  was  $4RPU$ , a value that was low enough that growth rate and GFP expression were unaffected; both remained constant for the all inductions of *IPTG*. RFP signal

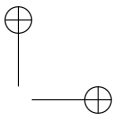
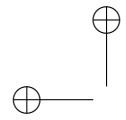
### 3. Engineering low-burden synthetic circuits via CRISPRi



**Figure 3.9: IPTG-inducible promoter characterization.** a) The systems express an RFP to characterize the synthesis rate of the LacI system driving its transcription, along with the usual GFP monitor cassette. b) On the x-axis, IPTG induction is proportional to the amount of RFP inside the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

output shown a readable output at IPTG concentrations above  $1\mu\text{M}$ ; compared to M12 Inv, the increase in RFP output was sharper and reached the maximum of around  $3.7\text{RPU}$  at an IPTG induction of  $30\mu\text{M}$ .

This suggests a higher Hill coefficient in the transfer function of the circuit. The data obtained from these constructs showed that both systems are functional at the used copy numbers and expression levels of regulatory genes. Moreover, the promoters are able to drive the downstream gene over a wide range of transcriptional levels, which, in case of RFP, do not lead to high metabolic burden. However, as discussed in the following two sections, additional data are needed to consider the two systems as reliable “knobs” for sgRNAs.



---

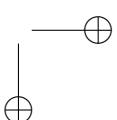
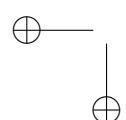
## 3.2. sgRNA expression cassette design

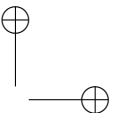
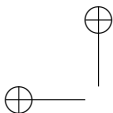
### 3.2.1 Flow cytometry analysis of IPTG-induced system

Luciferase quorum sensing parts were known to result in the unimodal behavior of transformed cells [84], therefore these parts were used with confidence in the efforts of building the corresponding sgRNA inducible device.

Lactose operon parts, on the other hand, offered a few more hurdles to face in circuit and experiment design: lactose requires the expression of *Lac permease* - a gene naturally included in TOP10 strain genome - to enter the cell and activate the system [100], leading to bimodal behaviour of cell population (i.e., fully induced and not induced without graded response). This problem was bypassed using IPTG, a lactose derivative that was able to diffuse across the plasma membrane in *E. coli* [101] (see Section C.2).

The hypothesis was confirmed through single cell analysis of bacterial populations bearing the IPTG-inducible GFP expression cassette Y35+Y3 (see Figure 3.10(a)), which consisted of the constitutive expression of LacI protein targeting  $P_{LlacO1}$  driving GFP expression (GFP expression cassette is in a medium copy plasmid); therefore, upon IPTG addition, repression on the Lac promoter was relieved and a GFP signal was recovered; those cells were cultured in media representing non-induced, half-activated and fully-activated device configurations. Differently from the construct previously illustrated in Figure 3.9, Y35-Y3 is activated at lower IPTG concentrations, most probably due to the different copy number of the Plac promoter compared to the LacI expression cassette, which thus exerts a weaker repression. Population-based characterization in microplate reader experiments is reported in Figure 3.10(b) for comparison. As expected, in flow cytometric analyses it was noticed the increase of mean GFP values recorded with increasing IPTG concentration, confirming the dose-response curve in Figure 3.10(b) and implying a well-tunable system;



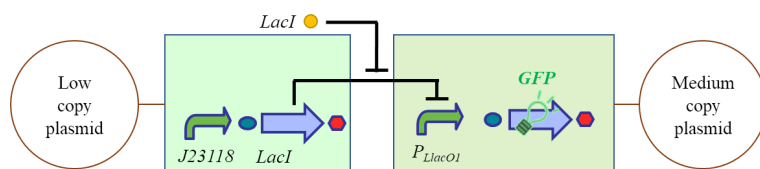


### 3. Engineering low-burden synthetic circuits via CRISPRi

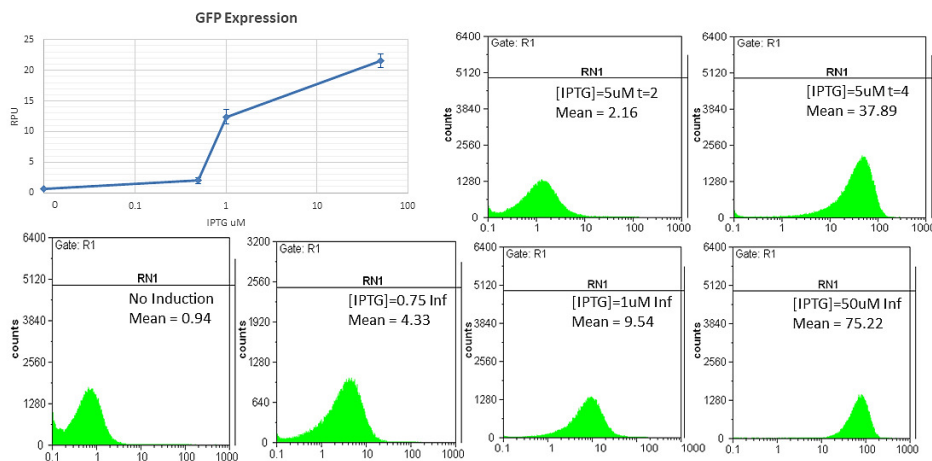
---

the majority of cells in the population responded in a unimodal fashion. However, it is worth noting that a slow IPTG-dependent activation dynamics has been observed: a representative experiment is reported in Figure 3.10(b), in which a culture was analyzed over time after induction with  $5\mu M$  of IPTG, and a different fluorescence distribution was detected between  $2h$  and  $4h$  after inducer addition (corresponding to  $t = 2$  and  $t = 4$ ). For this reason, all the experiments involving IPTG were carried out by adding the inducer to the medium since pre-inoculum, to avoid the slow induction dynamics, and to rely on a tunable system giving a homogeneous cell behavior.

### 3.2. sgRNA expression cassette design

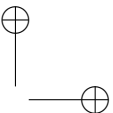
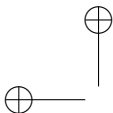


(a) Circuit schema.



(b) Response of the system.

**Figure 3.10: Flow Cytometer analysis of IPTG-inducible device.** **a)** The configuration of the system is depicted. **b)** On the top left, microplate reader experiments of the circuit at different inductions indicated the GFP expression potential of the promoter in medium copy. The two boxes on its right are a representative experiment of the slow dynamics of the lac system constructed.  $t = 2$  and  $t = 4$  represents the numbers of hours the bacterial populations were cultured with the inducer before sample analysis. In the four boxes on the bottom of the figure, to better understand the behavior of the newly constructed IPTG-inducible cassette, flow cytometric analysis of bacterial populations induced with diverse IPTG concentrations are reported, showing that the population behaved in a unimodal fashion. There existed however a small population of cells that exhibited different GFP signal output than the rest, indicating that the lower stability of the system at intermediate concentrations of IPTG. Data on flow cytometer were acquired for at least 100,000 events. Cells were gated based on SSC vs FSC readings to eliminate outliers and improve signal output. Data of microplate reader experiment are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

#### 3.2.2 TSS Optimization

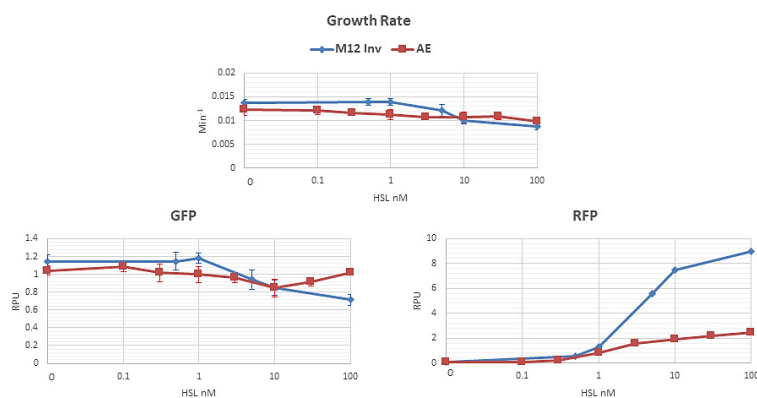
CRISPRi complex is guided by the sgRNA whose binding with desired DNA depends on its sequence complementarity to the target; hence, the presence of nucleotides downstream the promoter transcription start site driving the expression of the sgRNA (i.e., TSS > +1) can result in elongation and insertions of mismatching nucleotides of the desired 20 nucleotide annealing sequence of the sgRNA. Therefore, promoter driving sgRNA expression needed to be optimized to exclude possible undesired effects.

The most probable start sites of both of the chosen promoters were identified from the literature [40, 102]: for  $P_{lux}$ , the TSS fell upstream three adenine nucleotides, while for  $P_{LlacO1}$ , only one adenine was present downstream the TSS. This meant that the sgRNA transcribed from these promoters would include mismatches at the 5' end of the sgRNA (i.e., an extra adenine to compose a 21 nucleotide guide in the case of gPlac, and 3 extra adenines added to the 5' end of gPlux). From other investigations on the effect of mismatches on sgRNA:DNA complementarity [64, 103, 104], it was known that while the effect of a single additional nucleotide to the 5' end of a sgRNA is marginal, three additional nucleotides showed result in a substantial effect on the specificity and efficacy of a sgRNA. To that end, the mutagenesis of promoters driving sgRNA expression were carried out, to remove any nucleotide(s) that were initially part of the promoter but situated beyond the TSS. The two novel inducible devices obtained (AE with  $P_{lux-3A}$  promoter and AY-A with  $P_{lac-A}$ ), specifically implemented for sgRNA transcription, were subsequently compared to their original template (M12 Inv and AY respectively) by expressing the same mRNA encoding a strong RBS with RFP coding sequence used as reporter gene; this for a better understanding of the effect caused by these deletions.

## 3.2. sgRNA expression cassette design

### HSL-inducible sgRNA expression cassette

$P_{lux}$  cassette was mutagenized starting from M12 Inv to obtain the  $P_{lux-3A}$  promoter of AE strain: the two promoters exhibited similar growth rates and GFP expression for concentrations lower than  $10nM$ ; for higher concentrations, GFP expression for AE was higher than for M12 Inv, along with a lower RFP expression. In fact, the deletion of three adenine nucleotides downstream the  $P_{lux}$  TSS showed a 3-fold decrease in maximum RFP expression achievable (see Figure 3.11, circuit schema is the same as Figure 3.8(a) except for the  $P_{lux}$  promoter switched to  $P_{lux-3A}$ ). While RT-pcr experiments are still ongoing to evaluate variations in transcriptional efficiency for the mutagenized promoter, the dependence of this behavior is still uncertain since it could be addressed both by an altered promoter strength or a decreased efficiency of the RBS.



**Figure 3.11: Lux promoter mutagenized for optimal sgRNA expression.** M12 Inv and AE are quasi-identical constructs that differ solely by the presence or absence of three adenine nucleotides downstream the  $P_{lux}$  transcription start site. Measurements of growth rate, GFP, and RFP are shown. The x-axis represent HSL induction and is proportional to the amount of RFP inside the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

### 3. Engineering low-burden synthetic circuits via CRISPRi

#### IPTG-inducible sgRNA expression cassette

Similarly, deletion of the single nucleotide downstream the  $P_{LlacO1}$  TSS had an effect on its strength, however it was less relevant than that of  $P_{lux}$ . It was found that the modified promoter ( $P_{lac-A}$ ) of strain AY-A becomes responsive at a higher induction of IPTG, and reaches a maximum of  $3RPU$  compared to  $3.7RPU$  generated by the unmodified promoter of strain AY. The trend of the response curve was well conserved, but the deletion seemed to have an effect on the growth rate of the strain, resulted lower than the non-mutagenized strain (see Figure 3.12, circuit schema is the same as Figure 3.9(a) except for the  $P_{LlacO1}$  promoter switched to  $P_{lac-A}$ ).

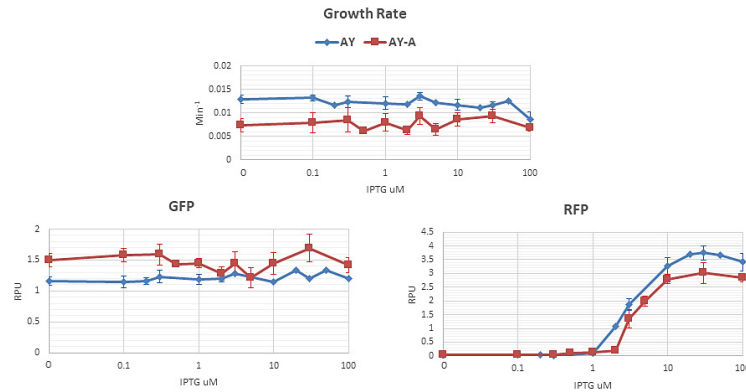
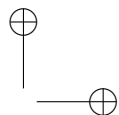
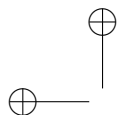


Figure 3.12:  $P_{LlacO1}$  mutagenized for optimal sgRNA expression. AY and AY-A differ only for an extraneous nucleotide after the TSS of  $P_{LlacO1}$ , which has been mutagenized via PCR to obtain the new synthetic promoter named  $P_{lac-A}$ . Measurements of growth rate, GFP, and RFP extrapolated from data obtained are shown. On the x-axis, increasing IPTG inductions determine higher amounts of RFP inside the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

Considering the overall wide tunability ranges of expression and rel-





---

### 3.3. Characterization of inducible sgRNA systems

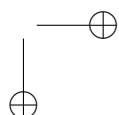
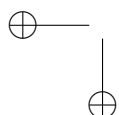
ative induction obtained, the two novel systems resulted suitable for sgRNA expression; however, while the modified  $P_{lux-3A}$  promoter (from AE) was the only HSL-inducible device used for express the guides, both IPTG-inducible ones ( $P_{LlacO1}$  and  $P_{lac-A}$ ) were used in the following studies due to the minor change in sgRNA expression (only one additional nucleotide).

The AE and AY-A promoters implemented in the study have been uploaded to the Registry of Standard Biological Parts as BBa\_J107202 and BBa\_J107203, respectively.

## 3.3 Characterization of inducible sgRNA systems

Once an optimized dCas9 and two different sgRNA expression devices were developed, the general standardized CRISPRi-based repression device architecture was conceived. The system, as reported in Figure 3.13 was implemented, composed by the following common modules:

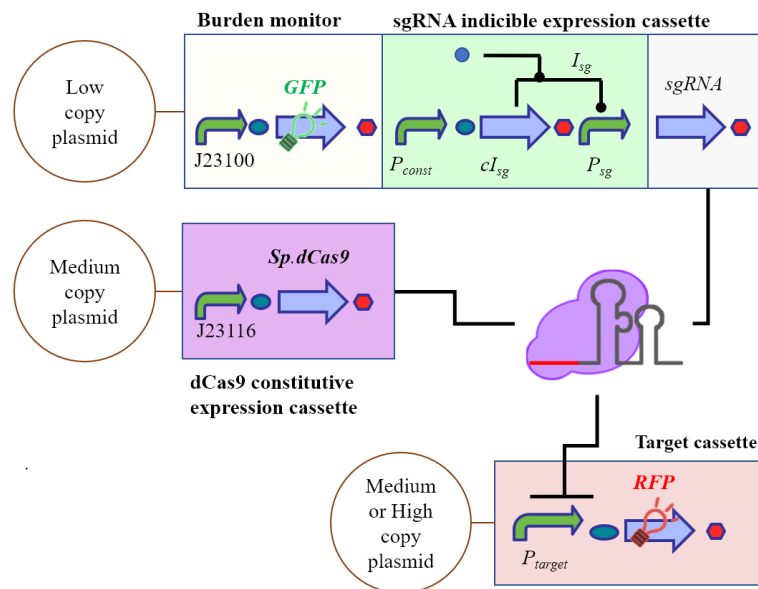
- constitutive GFP in low copy plasmid, acting as burden monitor as illustrated above;
- inducible sgRNA expression cassette in low copy plasmid composed by a constitutive promoter  $P_{const}$  driving the expression of a protein ( $cI$ ) that can bind an exogenous inducer ( $Isg$ ) and whose active complex ( $cI:Isg$ ) tunes the strength of the downstream promoter ( $P_{sg}$ ), thereby regulating the transcription rate of the sgRNA;
- constitutive dCas9 expression cassette in medium copy plasmid;



### 3. Engineering low-burden synthetic circuits via CRISPRi

- constitutive target promoter driving the expression of RFP in medium or high copy plasmid that can be bound and repress by dCas9:sgRNA complex.

Again, for each configuration characterized, a control circuit bearing an orthogonal non-targeting sgRNA under the same expression system was built and tested.



**Figure 3.13: CRISPRi model system schema.** The system is composed of: constitutive GFP burden monitor in low copy plasmid, an inducible sgRNA expression cassette in low copy plasmid, a constitutive dCas9 expression cassette in medium copy plasmid and a constitutive target promoter driving the expression of RFP in medium or high copy plasmid. **Control circuits:** bear an sgRNA specifically targeting an orthogonal promoter under the same expression system.

### 3.3. Characterization of inducible sgRNA systems

#### 3.3.1 Constitutive dCas9 expression: MC targets

In relation to the schema reported in Figure 3.13, in the following experiments  $P_\lambda$  was used as  $P_{const}$ , LuxR as  $cI$ , HSL as  $Ig$ ,  $P_{lux-3A}$  as  $P_{sg}$  and both  $P_{LtetO1}$  and  $P_{LlacO1}$  as  $P_{target}$  in medium copy plasmid.

##### $P_{LtetO1}$ medium copy target analysis

The optimized HSL-inducible cassette that drove the transcription of the sgRNA targeting  $P_{LtetO1}$  co-transformed with the target gene in medium copy displayed almost no RFP expression even at no induction (see Figure 3.14). A logarithmic scale applied on the  $y$ -axis showed approximately a 100-fold decrease in RFP at no induction up to a 1000-fold decrease in expression at HSL induction of  $1nM$ .

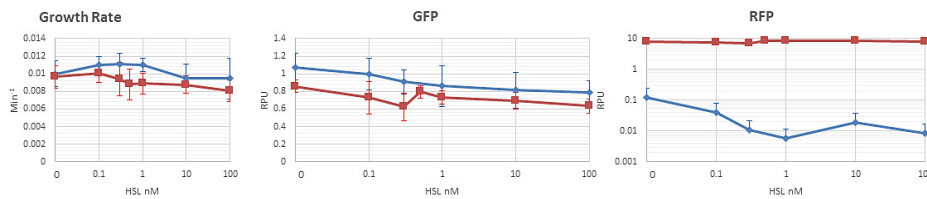


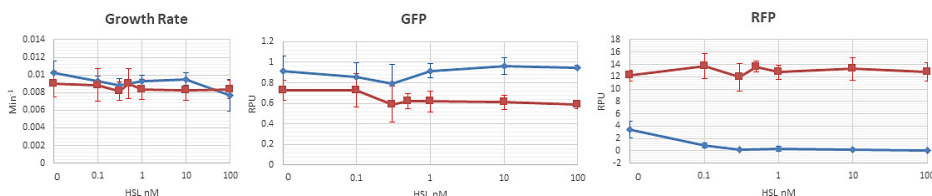
Figure 3.14: **AEgPtet co-transformed with medium copy target and constitutive dCas9.** The growth rate, GFP and RFP expression of the constructs are shown. The x-axis represents the HSL induction proportional to the amount of sgRNA transcribed inside the cell. The control (in red) bore a sgRNA targeting  $P_{LlacO1}$ , absent in the system. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

Although the medium copy  $P_{LtetO1}$  driven RFP target imposed a small burden onto bacterial cells, the integration of a functional CRISPRi system exhibited an improved growth rate and GFP expression profile in comparison to a control expressing a non-specific guide probably due to RFP expression shut down.

### 3. Engineering low-burden synthetic circuits via CRISPRi

#### $P_{LlacO1}$ medium copy target analysis

The system bearing an HSL-driven sgRNA targeting  $P_{LlacO1}$  driving RFP in medium copy showed to be less strong, with an RFP value of around  $3.5RPU$  at no induction to be compared to a signal output of  $12RPU$  for the negative control. However, small HSL inductions ( $< 1nM$ ) were sufficient to saturate the repression, corroborating the high efficiency of CRISPRi (Figure 3.15).



**Figure 3.15: AEGPlac co-transformed with medium copy target and constitutive dCas9.** The measurements obtained for growth rate, GFP and RFP. The x-axis represents the HSL induction that is proportional to the amount of sgRNA present in the cell. The control (in red) bore the sgRNA targeting  $P_{LtetO1}$ , absent in the circuit. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

In consideration of the efficiency and functionality of the system, it was decided to test the system with targets in high copy vectors, in an effort towards confirming and elucidating the actual repression capability of the studied systems.

### 3.3. Characterization of inducible sgRNA systems

#### 3.3.2 Constitutive dCas9 expression: HC targets

As before, in relation to the schema reported in Figure 3.13, in the following experiments  $P_\lambda$  was used as  $P_{const}$ , LuxR as  $cI$ , HSL as  $Ig$  and both  $P_{LtetO1}$  and  $P_{LlacO1}$  as  $P_{target}$ , but now in high copy plasmid.

##### $P_{LtetO1}$ high copy target analysis

The system contained by the strain AEGPtet+HC shown in Figure 3.16 confirmed the improved GFP expression profile of the functional system versus the control, due to a higher RFP expression in the latter. However, growth rate was negatively affected by the increasing inductions of HSL in both AEGPtet+HC and control; the study of this aspect is beyond the scope of this study, despite a deeper investigation on this behavior will be addressed in future works.

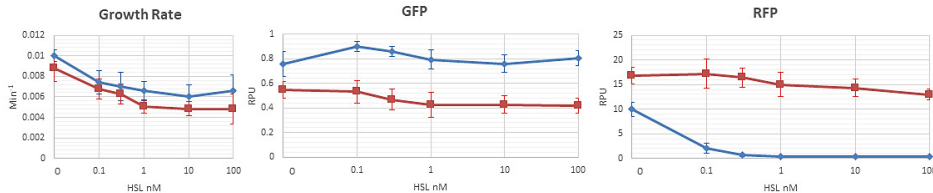


Figure 3.16: **High copy  $P_{LtetO1}$  target with HSL-inducible sgRNA and constitutive dCas9.** Signal output of growth rate, GFP and RFP. On the x-axis, HSL induction was proportional to the amount of sgRNA transcribed. The control (in red) bore a non-targeting sgRNA ( $P_{LlacO1}$ ). Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

RFP expression shown a behavior similar to the medium copy target one, which resulted in the complete repression of the system at low inductions of HSL; in fact, although fold repression was lower at no induction, the signal disappeared at inductions above  $1nM$ .

### 3. Engineering low-burden synthetic circuits via CRISPRi

#### $P_{LlacO1}$ high copy target analysis.

The system implemented in the strain AEgPlac+HC (see Figure 3.17) was the first to exhibit no repression, compared to control, when not induced with HSL, a feature contributing to the success of a finely tunable repression system.

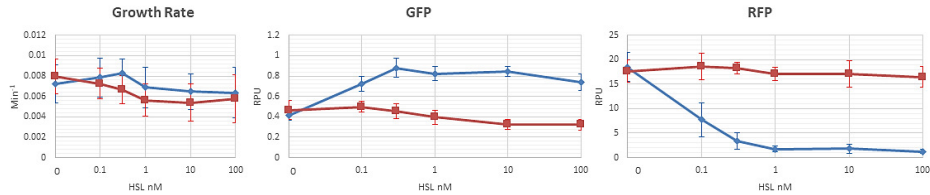
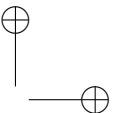
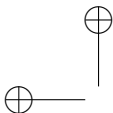


Figure 3.17: **High copy  $P_{LlacO1}$  driven target with HSL-inducible sgRNA and constitutive dCas9.** Signal output of the constructs at diverse HSL inductions. The x-axis represents the concentration of HSL induction that is proportional to the amount of sgRNA in the cell. The control (in red) bore a sgRNA targeting  $P_{LtetO1}$  promoter that was not present in the circuit. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

Both functional strain and control displayed the same RFP value at zero induction (18 RPU), which remained constant for the control and decreased readily for the AEgPlac+HC and reached the minimum value with an HSL induction of 1 nM. GFP levels were initially low for both constructs, owing to the high burden of RFP expression in high copy, but increased substantially when RFP expression was silenced by the CRISPRi complex. Growth rate on the other hand was more variable with relevant error bars, but the trend of the two graphs followed a similar pattern to other constructs where the functional system's growth rate exceeded that of the control and remained higher for all inductions of HSL (excluding the no induction value).



---

### 3.3. Characterization of inducible sgRNA systems

#### 3.3.3 IPTG-driven sgRNA: $P_{LtetO1}$ target analysis

As further investigation to test whether an improved repression curve could be achieved in a different circuit configuration, the IPTG-based inducible sgRNA expression cassette described in Section 3.2 was adopted to express sgRNA targeting  $P_{LtetO1}$  in medium and high copy. In relation to the schema in Figure 3.13, J23118 was here used as  $P_{const}$ , LacI as  $cI$ , IPTG as  $Ig$ , both  $P_{LlacO1}$  and its optimized version  $P_{lac-A}$  as  $P_{sg}$  and lastly  $P_{LtetO1}$  as  $P_{target}$  in medium and high copy plasmids.

#### IPTG-driven sgRNA: $P_{LtetO1}$ on medium copy target

In the former case, AYgPtet+dCasE62 was tested alongside AY-AgPtet+MC, representing the non-optimized and optimized promoter respectively, driving the expression of the sgRNA for  $P_{LtetO1}$  target. Since the non-optimized and optimized versions of the IPTG-inducible system showed similar (though non-identical) dose-response curves, an analogous repression curve in the tested system was expected, being the additional nucleotide downstream of TSS in non-optimized promoter negligible in terms of repression capability [105]. From a more detailed analysis regarding promoter optimization effects on the specific guide sequence of the guide targeting  $P_{LtetO1}$  promoter and the target sequence itself (refer to Table C.4), the problem was clarified. During the preparation of the optimized promoter, an adenine nucleotide after the TSS that would have been part of the sgRNA targeting the promoter, was removed; however, looking at the promoter and modified sgRNA sequences, it was found that the extraneous adenine nucleotide was complementary to the promoter sequence, which resulted in a 21 nucleotide perfectly annealing guide. On the other hand, the optimized promoter transcribed a sgRNA of 20 nucleotides that was perfectly complementary as well, but of shorter

### 3. Engineering low-burden synthetic circuits via CRISPRi

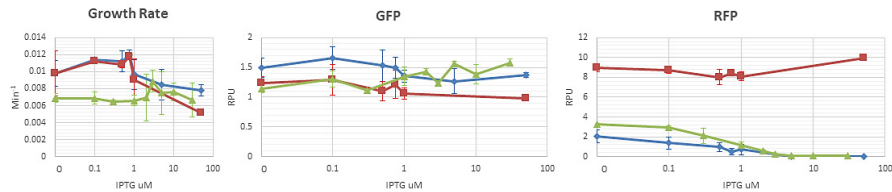


Figure 3.18: **Medium copy  $P_{LtetO1}$  target with IPTG-inducible sgRNA and constitutive dCas9 cassettes.** Signal output of the constructs at diverse IPTG inductions. The optimized construct (in green) presented as AY-AgPtet contained the optimized  $P_{LlacO1}$  with the deletion of the single adenine nucleotide present after the TSS. The x-axis indicates the concentration of IPTG, which is dependent on the amount of sgRNA inside the cell. The control (in red) bore sgRNA targeting  $P_{LuxRep}$  absent in the circuit. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

length. Consistently with the dose-response curves of IPTG-inducible systems (Figure 3.9), it was observed that the sgRNA repression capability was slightly higher for the optimized system. Despite the additional nucleotide is not expected to change guide efficiency, additional studies are needed to confirm it by decoupling inducible promoter transcriptional activity and repression capability as a function of guide length, despite systematic studies are available in literature in different systems [65]. Looking at the growth rate, the optimized construct exhibited a significant decrease, similar to the construct used to characterized the optimized promoter expression potential (Figure 3.12); the reason of this unexpected behavior is still under investigation. GFP expression on the other hand behaved as expected, with both constructs expressing higher levels of the reporter gene in comparison to the control. However, the GFP output of the optimized construct behave as the control for low inductions and then increased to levels similar to the non-optimized construct with increasing IPTG inductions, along with a decreased expression of RFP. Again, both constructs achieved maximal RFP repression at IPTG inductions of  $5\mu\text{M}$ , whereas the negative control had an average RFP output of



### 3.3. Characterization of inducible sgRNA systems

around  $10RPU$ .

#### IPTG-driven sgRNA: $P_{LtetO1}$ on high copy target

Trends observed for the medium copy version were confirmed for constructs co-transformed with high copy target plasmid, despite growth rate of the optimized promoter construct AY-AgPtet+HC was lower than both control and non-optimized construct AYgPtet+HC.

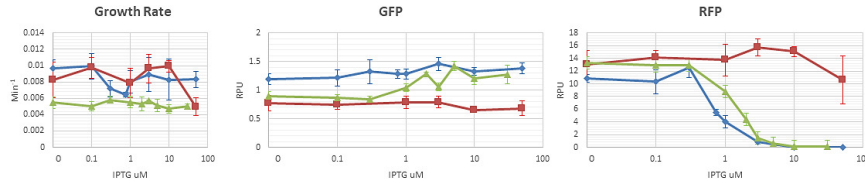
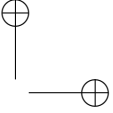
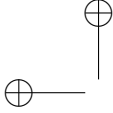


Figure 3.19: **High copy  $P_{LtetO1}$  target with IPTG-inducible sgRNA and constitutive dCas9.** Signal output of the constructs at diverse IPTG concentrations. The x-axis represents the concentration of IPTG that is proportional to the amount of sgRNA transcribed in the cell. The control (in red) bore a sgRNA targeting  $P_{LuxRep}$  absent in the circuit while the optimized one is reported in green. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

GFP expression for AYgPtet+HC remained higher than the control, slightly increasing at the highest IPTG concentrations. Similarly, the optimized construct AY-AgPtet+HC (see Figure 3.19) had a GFP signal similar to the control at low IPTG concentrations, but increased significantly with IPTG concentrations greater than  $1\mu M$ , reaching the expression levels of the non-optimized construct. RFP expression was similar as in the previous case in which the target was in medium-copy plasmid. The weaker repression obtained from the optimized construct allowed an initial RFP signal output equal to the control, similarly to AEGPlac+HC seen in Figure 3.17. AYgPtet+HC instead resulted in reduced initial RFP output. The two constructs exhibited



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

an overall similar trend, with AY-AgPtet+HC RFP values higher than AYgPtet+HC for several of the initial IPTG concentration points, and reached a minimal RFP output at the same concentration of  $5\mu M$ .

#### 3.3.4 CRISPRi Mathematical model

Despite several models have been proposed for describing CRISPR/Cas9 mechanism via both empirical and mechanistic approaches [106, 105, 104, 64], those results are still not completely accepted or validated enough to be directly used as golden standards in rational design.

While a mechanistic analysis of the system developed is beyond the aim of this study, here an empirical approach has been adopted to preliminarily describe the system, aiming to obtain a proof-of-concept characterization and a feasibility evaluation of the implemented circuitry to be used as a low burden module for gene regulation. The mathematical model used was an adapted version of Equations 2.7-2.14 in which sgRNAs were considered as promoter repressor in a steady state system.

It is worth to notice that, under the hypotheses of overabundance of dCas9 in the cell, it is possible to neglect the description of the dCas9:sgRNA relative components amount and binding dynamics, avoiding the detailed description that would lead to the usage of a greater amount of parameters and equations.

### 3.3. Characterization of inducible sgRNA systems

In general, under the same assumptions described in Section 2.2.1 and 2.2.2, the output at steady-state of the system in Figure 3.13 can be described by the following set of equations:

$$S_{cell,norm,GFP} = \frac{a_G}{a_G + \mu} \cdot \frac{S_x}{D} \quad (3.1)$$

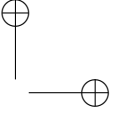
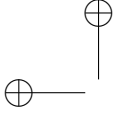
$$sgRNA = \delta_{sg} + \frac{\alpha_{sg}}{1 + \left(\frac{K_{sg}}{[I_{sg}]}\right)^{\eta_{sg}}} \quad (3.2)$$

$$S_{cell,norm,RFP} = \frac{a_R}{a_R + \mu} \cdot \frac{\left( \delta_{P_{target}} + \frac{\alpha_{P_{target}}}{1 + \frac{[sgRNA]}{K_{P_{target}}}} \right)}{D} \quad (3.3)$$

$$D = 1 + \Sigma_{cI_{sg}} + \Sigma_{HC} + J_{RFP} \cdot \left( \delta_{P_{target}} + \frac{\alpha_{P_{target}}}{1 + \frac{[sgRNA]}{K_{P_{target}}}} \right) \quad (3.4)$$

In this model:

- $a_R$  and  $a_G$  are maturation rates of RFP and GFP respectively;
- $\mu$  is the cell growth rate of the from which mature reporter protein dilution is dependent;
- $S_x$  is the constitutive expression rate of immature GFP reporter protein of burden monitor cassette;
- $\delta, \alpha, K, \eta$  are parameters of Hill or Michaelis Menten ( $\eta = 1$ ) functions used to describe transcriptional activities of regulated promoters;



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

- $sgRNA$  amount is described by a Hill function without maturation neither degradation terms since the molecule does not need to be translated into a functional protein and being its degradation rate (assumed the same for each  $sgRNA$ ) faster than cell division [107];
- $I_{sg}$  is the inducer (i.e., HSL or IPTG) of the regulated promoters driving  $sgRNA$  expression;
- $cI_{sg}$  is the overabundant constitutively expressed transcriptional co-factor activated by the inducer  $I_{sg}$ ;
- $P_{target}$  is the target promoter;
- Michaelis Menten functions were used to describe RFP synthesis rate, being  $dCas9:sgRNA$  complex repressing target promoters transcriptional activity and acting on a single binding site;
- $D$  is the same denominator described in Section 2.2.2, adapted with proper terms<sup>2</sup> to describe the system including its structural differences from the ones described in chapter 2;
- $\Sigma_{cI}$  is the burden contribution given by the expression of transcriptional co-factors used in  $sgRNA$  constitutive expression cassette;
- $\Sigma_{HC}$ , which is set to zero when the target promoter is in medium copy number, describes the burden contribution given by the presence of 3 plasmids in the cells (i.e., High copy target cases);

---

<sup>2</sup>All the configurations tested and modeled in this work bore the same genetic structure and were implemented in the same host strain in presence of the usual GFP-based burden monitor, hence normalization steps as described in Equations 2.9-2.12 were still valid.

### 3.3. Characterization of inducible sgRNA systems

- $J_{RFP}$  is the burden contribution given by the expression of RFP used in target cassettes.

It is worth to notice that circuits bearing the dCas9 constitutive expression cassette unexpectedly showed a higher GFP level compared to cultures bearing the burden monitor alone (see Figure 3.20).

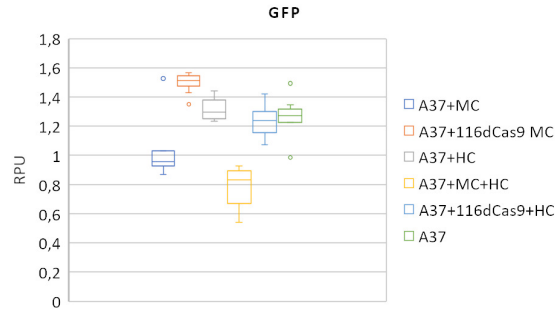
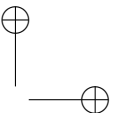
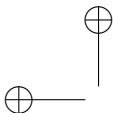


Figure 3.20: **Effects of co-transformed plasmids and dCas9 expression on GFP monitor level.** In the boxplot, GFP output of circuits bearing **burden monitor** in low copy co-transformed with an empty medium copy plasmid, **burden monitor** in low copy co-transformed with constitutive dCas9 expression cassette in medium copy plasmid, **burden monitor** in low copy co-transformed with an empty high copy plasmid, **burden monitor** in low copy co-transformed with two empty plasmids (medium and high copy respectively), **burden monitor** in low copy co-transformed with constitutive dCas9 expression cassette in medium copy and an empty high copy plasmid and **only the burden monitor** in low copy plasmid are shown in RPU.

The use of the same baseline immature GFP expression level  $S_m$  from Section 2.1.3 for all the circuits would therefore impose the usage of a hypothetical negative burden term for those strains bearing a dCas9 expression cassette in medium copy plasmid; however, being a negative contribution not biologically meaningful, the immature GFP synthesis rate  $S_x$  of a strain bearing the burden monitor and dCas9 expression cassettes was chosen as baseline instead of  $S_m$  value since the modeled circuits were all implemented with the same 2-plasmid architecture (except for gRNA expression cassette characterization circuits for



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

which  $S_m$  was kept). Clearly, while this interesting and unexpected behavior of the system needs deeper investigations, this aspect was beyond the scope of the study, aiming to firstly find a general evidence of its worth of usage. It is worth to notice that, comparing the value of  $S_m$  in Table 2.1 with the one reported in Table C.2 the value of the parameter is different, due to the use of a slightly varied architecture of the monitor cassettes used in the two studies: in Chapter 3 the GFP expression device was placed downstream the tested circuits and therefore the chosen reference encoded a terminator upstream the constitutive promoter; in this Chapter, the monitor cassette were always placed as first transcriptional element of the plasmid, so the reference strain did not bear the terminators upstream the expression cassette as described in Section 2.1.3.

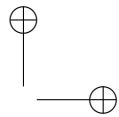
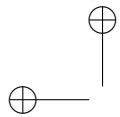
Once the base GFP level for circuits bearing the constitutive dCas9 expression cassette in medium copy plasmid along with the GFP monitor in low copy was fixed, it was possible to analytically calculate the metabolic contribution exerted by the maintenance of a high copy plasmid, namely  $J_{HC}$ . Being the GFP signal for a circuit bearing dCas9 expression cassette, burden monitor and an empty high copy plasmid given by:

$$S_{cell,norm,GFP} = \frac{a_G}{a_G + \mu} \cdot \frac{S_x}{1 + \Sigma_{HC}} \quad (3.5)$$

the evaluation of  $\Sigma_{HC}$  is feasible.

A summary of the fitted parameters and constants appearing in this model is reported in Appendix (see Table C.2); all the fitting steps and approaches adopted were performed by using the Matlab (Matworks) *lsqnonlin* routine with residuals weighted on experimental standard deviations.

Under this assumptions, several approaches have been evaluated to validate the applicability of the model. In all the figures shown below, where three dashed lines are shown in RFP and GFP graphs, a sensitivity analysis on  $\delta_x$  parameters has been performed multiplying

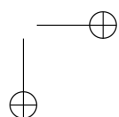
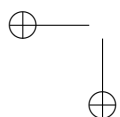


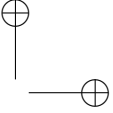
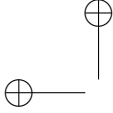
### 3.3. Characterization of inducible sgRNA systems

---

by 0.5 and 1.5 the estimated value; this because data showed a huge repression effect in presence of a leaky repressor concentration (i.e., for zero induction of sgRNAs, systems were already repressed) and small variation on evaluation of this parameter provided sensitive effects in the overall system response prediction.

1. **One step Global Fitting:** A global fitting on all the data was performed to find a parameter set including the available information by adopting both the model described above (*Burden model*) and a *No Burden* version neglecting the  $D$  term of Equations 3.1 and 3.3.
2. **Two-step Global Fitting:** A two-step fitting procedure was performed to evaluate the feasibility of a modular approach: firstly, sgRNA expression cassettes parameters were estimated along with burden contribution of transcriptional factors included in their architecture; then, repression systems in both medium and high copy of the three combination of sgRNA expression cassette-target promoter pairs were fitted, given the value of the parameters estimated in the first step. Both steps were performed with and without burden contribution, as before. In this analysis, the  $J_{RFP}$  chosen value was the one derived in the Global fitting procedure with burden model; since neither the HSL-based optimized promoter  $P_{lux-3a}$  nor  $P_{LlacO1}$  were strong enough to provide a working condition in which RFP expression were high enough to properly display a variation in GFP, the fitting procedure would have been affected, leading to poor-quality estimates. It is worth to notice, however, that the  $J_{RFP}$  value used in this study was different from the one reported in Table 2.1 (i.e 0.021 in the former case, 0.04 in the latter); despite the model structure adopted in the two studies as well as hosts and most of the biological parts used were the same, this difference





### 3. Engineering low-burden synthetic circuits via CRISPRi

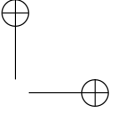
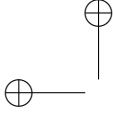
---

could probably be addressed to the hidden effects of copy number variations and co-transformation of different plasmids in the same cell. In fact, while the pool of strains used to estimate parameters in the study of Chapter 3 included circuits borne only on low copy plasmids, this study relies on modules encoded in 2- or 3-plasmid architectures. This point is not trivial and further investigations on copy number variation of circuits with “multi-plasmid architecture” effects on gene expression are still ongoing.

- 3. Two-step approach with Training set and Test set for prediction:** Lastly, given the fitting results of the first step of the previous analysis (i.e., characterization of the two sgRNA expression cassette and therefore guide amount dependency on induction and burden contribution of LacI and LuxR protein at their constitutive expression rate) the experiments used in the second step of the fitting were divided into a training and a test set: the former, including HSL-driven gPlac and IPTG-driven gPtet targeting promoters in medium and high copy, was used to fit all the parameters of the system (after sgRNA cassette characterization of the first fitting step); the latter, including HSL-driven gPtet targeting systems in medium and high copy, was used as testbed to evaluate the predictive power of the model given a relatively small amount of data. Again, both versions of the model (i.e., with and without burden description) were used.

Figures relative to the former procedures are reported in Appendix section C.7.





---

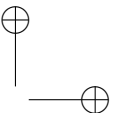
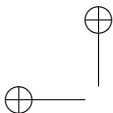
### 3.3. Characterization of inducible sgRNA systems

#### 3.3.5 Overall observations

Observing the global fitting, both Burden and No Burden models seemed to be able to properly describe the system; this is true, at least, when the sensitivity analysis is included, highlighting the expected uncertainty on  $\delta$  parameters and therefore the importance of deeply understand how and why leaky expression of sgRNAs exerts such a strong repression in the system. Considering the repression curves and in particular  $\alpha$  and  $k$  parameters, a stronger efficiency of gPtet regard to gPlac can be seen; while this result was not unexpected, it highlights the necessity in rational design of synthetic genetic circuits of a full characterization of each designed guide, waiting for a functional and applicable model to quantitatively predict guide efficiencies to be developed.

Moving to the two step fitting analysis, the overall fitting results were slightly worse, as expected being the dataset to fit the sgRNA expression functions smaller. However, parameter estimation were quite reliable regard to Global Fitting versions of burden and no burden models adopted. Moreover, while GFP signal were not completely caught in case of the Burden Model, RFPs are still well described in both cases. Also, predictions obtained appeared overall successful: small but appreciable differences could be observed just for gPtet repressing systems (compare Figures C.9(a)/C.14(a) - C.10(a)/C.15(a) - C.9(b)/C.14(b) - C.10(b)/C.15(b) - C.9(c)/C.16(a) - C.9(c)/C.16(a) for Burden model and C.12(a)/C.17(a) - C.13(a)/C.18(a) - C.12(b)/C.17(b) - C.13(b)/C.18(b) - C.12(c)/C.19(a) - C.12(c)/C.19(a) for No Burden model.)

Considering all the data obtained in Table C.2, none of the parameter sets showed clear evidence of effects of copy number variation in target cassette. More precisely, it was expected that ratios between  $\alpha$  and  $k$  parameters of sgRNA related repression curves on medium and high copy number target cassette (E62-A33 for gPlac and E52-I13521 for



### 3. Engineering low-burden synthetic circuits via CRISPRi

---

gPtet respectively) were linked to ratios between relative plasmid copy number ( $\sim 5$  [84]); however, this was not the case.

Most importantly, while the usage of burden model allowed the description of two signals at once (i.e., GFP and RFP) supporting the results achieved in Chapter 3, the fact that all the tested systems were also explained by a model that did not take into account any load contribution, evidenced the lack of excessive resource demand exerted by these modules, supporting the aimed worth of this study.

#### 3.4 Dependence on promoter strength: $P_{luxRep}$ library analysis

After sgRNA repression capability dependences was studied as a function of its sequence and level, a system to characterize repression dependence on target promoter strength using the same sgRNA sequence was implemented. On that direction, a sgRNA (gPluxH) able to target a library of promoters created by mutagenizing the  $-10$  box from the same HSL:LuxR-repressible promoter - created by our lab in a previous study [84] - was designed. The sgRNA gPluxH (see Table C.4 in Appendix) was able to anneal to all  $-35$  boxes of the promoters leading to their repression and were expressed through an IPTG-inducible cassette (AY) (see Figure 3.21). Promoters were chosen according to their transcriptional strengths, in particular:

- BBa\_J107100 (named  $J119H$  or more generally with the interchangeable term  $P_{luxRep}$ ) was the strongest and used as template to create the library via PCR mutagenesis;
- BBa\_J107111 (named  $P122$ ) was an intermediate-high strength promoter

### 3.4. Dependence on promoter strength

- BBa\_J107101 (named  $P_2$ ) was a weak-intermediate promoter;
- BBa\_J107105 (named  $P_{44}$ ) a weak promoter.

Since their sequences differed only at the level of the  $-10$  box, this allowed the study of the repression a single guide can have on promoters with different strengths.

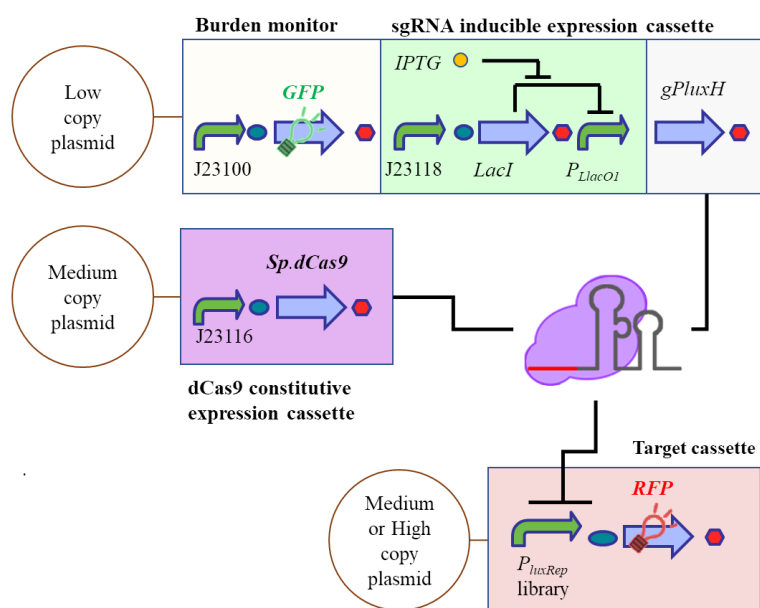


Figure 3.21: **Schema of CRISPRi repressing system on  $P_{luxRep}$  promoter library.** The system configuration is presented. Target promoters were identical excluding the  $-10$  box that was mutagenized to obtain promoters of different strengths such that BBa\_J107100>BBa\_J107111>BBa\_J107101>BBa\_J107105.

### 3. Engineering low-burden synthetic circuits via CRISPRi

#### 3.4.1 $P_{luxRep}$ library with MC target

When the target promoters were in medium copy plasmids (see Figure 3.22), all constructs maintained a stable growth rate over the range of IPTG concentrations tested; a rather stable growth rate and GFP expression over the range of IPTG concentrations. RFP expression was minimal for all promoters. However, considering a logarithmic y-axis scale, it was possible to notice that the level of RFP expressed was conserved in terms of the ranking of promoter strengths [84].

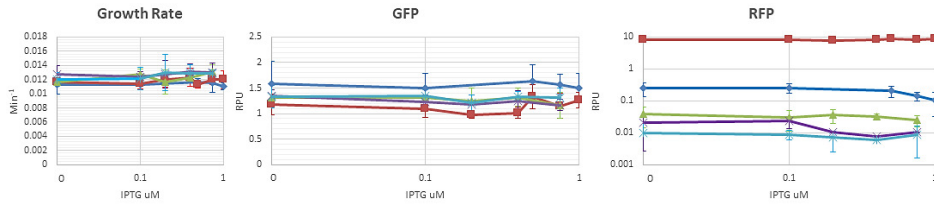


Figure 3.22: **IPTG-inducible sgRNA and constitutive dCas9 repressing medium copy  $P_{luxRep}$  promoter library.** Measurements of growth rate, GFP and RFP expression of the transformed strains. Lines in blue, green, violet, cyan and red represent, respectively 119H, P122, P2, P44 members of the library - in descendant strength order - and control. The x-axis represents the concentration of IPTG in micromolar and was proportional to the amount of sgRNA in the cell. The control bore a sgRNA targeting  $P_{LtetO1}$  that was absent in the circuit. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

#### 3.4.2 $P_{luxRep}$ library with HC target

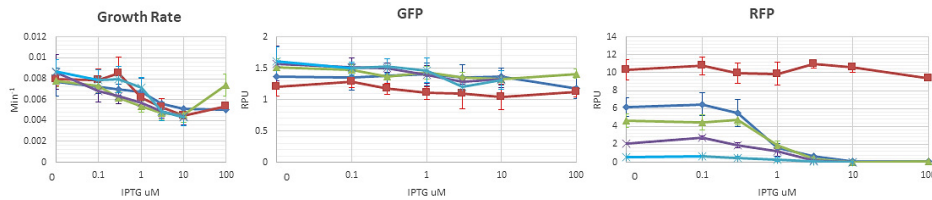
To increase the range of RFP expression and to better study the system with different configurations, the medium copy vector carrying the target promoters were substituted with high copy vectors, leading to a higher expression level of RFP (see Figure 3.23).

The high copy targets did indeed show increased RFP measurements,

### 3.4. Dependence on promoter strength

compared with medium copy ones, that was accompanied by a decrease in growth rate (see Figure 3.20 B). In particular, the promoters showed ranking consistent with Figure 3.22 and their relative expression (compared with negative control) was consistent with the original publication previously illustrating their strength, with the *J119H* promoter being the strongest one. At high IPTG concentrations, RFP becomes strongly repressed. This means that gPluxH in the tested condition can exert a high repression strength and range as a function of IPTG (corresponding to the amount of sgRNA transcribed), leading to well tunable RFP levels. Nonetheless, RFP expression from AYgPluxH+J119H showed slightly lower level than the control (which includes the same setup, but with a non-specific sgRNA), meaning that the basic activity of  $P_{LlacO1}$  in the off state (i.e., without IPTG) was sufficient to exert a relevant repression.

The respective descending order of promoter strengths were measured to have RFP values of 6.1RPU, 4.6RPU, 2.1RPU, and 0.6RPU, respectively.



**Figure 3.23: IPTG-inducible sgRNA and constitutive dCas9 repressing high copy  $P_{luxRep}$  promoter library.** Measurements of growth rate, GFP and RFP expression of the transformed strains. Lines in blue, green, violet, cyan and red represent, respectively 119H, P122, P2, P44 members of the library - in descendant strength order - and control. The x-axis represents the concentration of IPTG in micromolar and was proportional to the amount of sgRNA in the cell. The control bore a sgRNA targeting  $P_{LtetO1}$  that was absent in the circuit. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

### 3. Engineering low-burden synthetic circuits via CRISPRi

To find a possible transcriptional strength-dependent sgRNA-mediated repression pattern among the tested constructs, their RFP expression were normalized to the expression level reached at no IPTG concentration and the maximum RPU reached by each curve (see Figure 3.24).

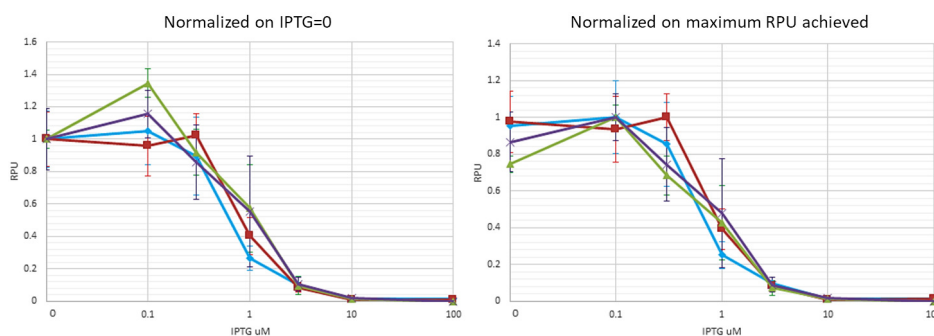
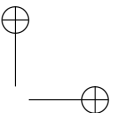
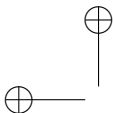


Figure 3.24: **IPTG-inducible sgRNA and constitutive dCas9 repressing high copy  $P_{luxRep}$  promoter library.** RFP expression of the transformed strains normalized to values obtained at no IPTG concentration (left box) and to the maximum RFP level achieved by each curve (right box). Lines in cyan, red, green and violet represent, respectively 119H, P122, P2, P44 members of the library, in descendant strength order. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

The normalized expression showed no relevant change in repression function shape among the four promoters, thereby suggesting that transcriptional strength does not play a role in sgRNA efficiency in terms of percent activity repression. On the other hand, as expected, absolute RFP expression depends on promoter strength when targets are both in medium- and high-copy context and in both non-repressed and repressed state (see Figure 3.22 and 3.23). Specifically, the normalized curves showed similar IPTG concentration leading to half-maximum repression and slope. Growth rate and GFP expression did not show relevant IPTG-dependent and promoter strength-dependent variation, except for the control strain, in which GFP is lower than the



### 3.4. Dependence on promoter strength

---

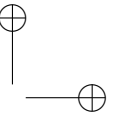
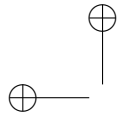
other tested constructs, consistently with the highest RFP expression shown by the control due to its non-repressed condition.

Taken together, the results suggested that the characterization of sgRNAs in one context can be used to quantify their repression strength, which can be generalized to other target promoters with the same sgRNA binding region. Despite results showed that no percent repression difference can be seen, they again showed that repression strength is highly dependent from the specific sgRNA: the ones designed to repress  $P_{LtetO1}$  and  $P_{luxRep}$  showed higher repression strength the one targeting  $P_{LlacO1}$ .

### 3. Engineering low-burden synthetic circuits via CRISPRi

---





# Chapter 4

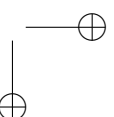
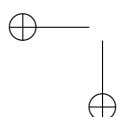
## Further investigations

This chapter contains preliminary results regarding some improvements to the works described in the previous chapter; strains, cloning procedures, biological parts used to implement new synthetic circuits, materials and methods are the same as the ones used and described in Section 2.1 and Appendix C.

### 4.1 CRISPRi-Transcription Factor interference

Considering the possible applications of CRISPRi repression systems in the construction of complex synthetic circuits, further investigations were performed to see how the sgRNA:dCas9 complex would behave in the presence of transcriptional regulator competing for the same promoter to be bound. To that end, a circuit bearing:

- a constitutive expression cassette synthesizing a transcriptional regulator (TF, specific for the sgRNA-targeted promoter) that





## 4. Further investigations

---

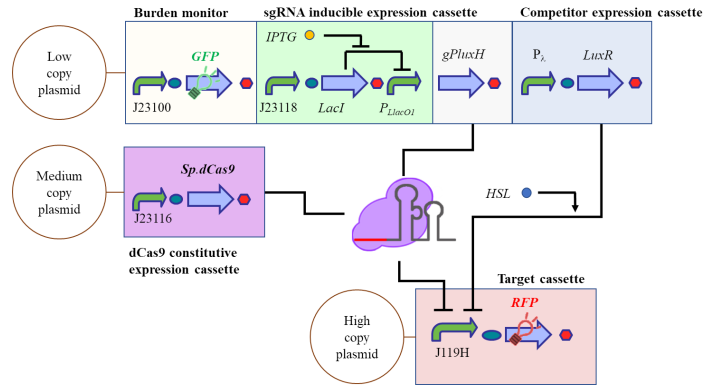
could be inhibited or activated by a specific inducer supplemented to the medium;

- an inducible sgRNA expression cassette tuned by an inducer orthogonal to the one used for the TF to be regulated and targeting the same promoter;
- a constitutive dCas9 expression cassette in medium copy plasmid;
- a module expressing an RFP driven by the promoter targeted by both dCas9:sgRNA complex and TF, assembled in high copy plasmid.

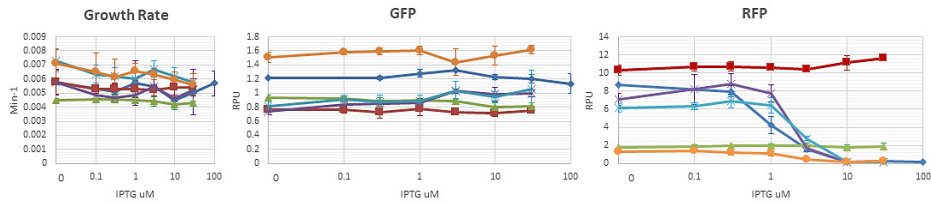
was built and tested starting from the circuits described in the previous chapter.

This systems was therefore tunable through 2 different inducers, one tuning the expression of sgRNA while the other regulating the repression exerted by the transcription regulator. The LuxR protein (that binds HSL) was used as a competitive repressor of  $P_{luxRep}$  BBa-J107100; experiments were set up to achieve different levels of repression by varying the concentration of HSL depending on the amount of active LuxR:HSL complex while expression of sgRNA was under the control of another supplemented inducer - IPTG - whose amount was dependent to the sgRNA:dCas9 complex repressing the target promoter (see Figure 4.1). Control circuits bore a sgRNA targeting  $P_{LtetO1}$ , absent in the circuit, tested at the zero and (100nM) concentration of HSL to activate LuxR protein.

## 4.1. CRISPRi-Transcription Factor interference



(a) *Circuit schema.*



(b) *Growth rate, GFP and RFP expression .*

**Figure 4.1: Competition between LuxR protein and sgRNA:dCas9 complex for repression of *J119H*.** a) The system setup described composed by an inducible sgRNA and constitutive dCas9 form a complex that competes with LuxR transcription factor for the repression of the BBa\_J100107 driving RFP expression. Both the activated LuxR:HSL and dCas9:sgRNA complexes acted to repress target promoter and downstream gene expression. b. The x-axis represents IPTG induction that determines the amount of sgRNA present and thus the level of repression by the CRISPRi system. Curves in blue, violet, cyan and orange represent 0nM, 0.5nM, 5nM and 50nM of HSL induction respectively, while green and red curves represent zero and full induced (50nM HSL) controls. The amount of activated LuxR protein depends to the concentration of HSL supplemented to the medium, given in the legend. The control bore a sgRNA targeting *P<sub>LtetO1</sub>* absent in the circuit, tested at zero and (100nM) induction of HSL to activate LuxR protein. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

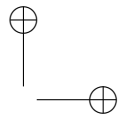
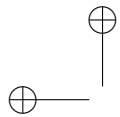


## 4. Further investigations

---

An analysis of the strain bearing  $P_{luxRep}$  with the illustrated setup showed that RFP expression can be successfully repressed by either IPTG (driving gPluxH) or HSL (which activates LuxR that acts as a repressor) (Figure 4.1). In particular, the control strain (containing a non-specific sgRNA instead of gPluxH) had RFP level of about 10 and 2 RPU without HSL and with 100 nM of HSL, respectively, demonstrating a successful 5-fold transcription factor-mediated repression. Only by expressing gPluxH,  $J119H$  promoter activity becomes completely repressed for all the added HSL concentrations. This result suggests that the repression effect of sgRNA is stronger than the one exerted via protein repressor. The repression curves in presence of different HSL concentrations showed an IPTG-dependent maximum activity and switch point. Specifically, an increase of IPTG concentration results into lower RFP expression, as expected by the gPluxH-mediated repression, and into higher IPTG concentration corresponding to RFP drop. This effect may be due to the lower gPluxH efficiency in presence of a protein complex (LuxR:HSL) which binds to the promoter region and their binding sites overlap. Considering growth rates and GFP signals, they showed a low variation; the strain with the most repressed RFP output (maximum HSL) showed the highest growth rate and GFP signal as expected.

Although the results of these sets of experiments were informative of the behavior that a CRISPRi complex can have inside the cell when it encountered a transcription regulator, the high metabolic load of the circuit lead to a scarcity of data available for definitive conclusions to be drawn. Inasmuch, further studies are needed involving other system configurations using different type transcriptional regulators (for example one that promotes gene expression) in order to elucidate and eventually model interactions among the CRISPRi complex and competitive transcription regulators.



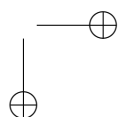
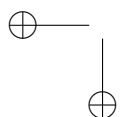
## 4.2 Approaches to improve CRISPRi tunability and repression range<sup>1</sup>

### 4.2.1 Altered sgRNAs

In literature [65, 104, 103] many examples of relationship between alteration in sgRNA affinity (due to mismatches or altered lengths of the annealing region of the guide) and repression strength have been reported. An exploitation of this effect was thought to be useful to tune the repression capability of the CRISPRi device systems described in the previous chapter; by changing the 20 nucleotides composing the annealing region of the sgRNA, a decrease of the already proven high efficiency of the system (i.e., obtaining a null repression for very low synthesis of the sgRNA) and a widening of the induction range of the expression module has been sought. This improvement is necessary to use the developed system in rational design of those synthetic genetic circuits where a fine modulation of an already low signal without other possible regulation systems is required. As a preliminary investigation of the worth of this approach, one of the previously described circuits showing a very high efficiency was chosen as model circuit to work on; the same circuit of Figure 3.21, with J119H target in medium copy plasmid, were taken and the sgRNA encoded in its sequence (gPluxH, targeting the promoter  $P_{luxRep}$ ) mutagenized to obtain several versions as described in Figure 4.2.

---

<sup>1</sup>Part of this study (Section 4.2.2) has been started and carried out at the *Del Vecchio Laboratory - Massachusetts Institute of Technology (Cambridge, MA)* and is still ongoing. Details of the building blocks of this work are reported in [108].



## 4. Further investigations

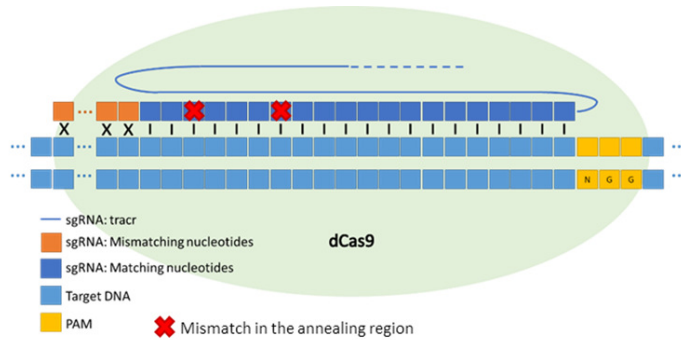


Figure 4.2: **Design of sgRNAs containing curtailments, elongations and mismatches.** sgRNAs can be mutagenized in order to change their structure by curtain or mismatching elongations of their 5' or including mismatches in the annealing region.

For the construction of the strains, AYgPluxH was used as template and then mutagenized by combining primers RV\_Plac-A, RV\_Plac and modified versions of FW\_gPluxH\_tracr with curtailments, elongations and mutations<sup>2</sup> on the first 20bp of the sgRNA (again, following the same protocol described in the former chapters) in order to obtain the desired gRNA sequence, reported in Table 4.1. Control strains always bore a non-targeting sgRNA (namely gPtet) as did for the characterization of  $P_{luxRep}$  library in Section 3.4.

<sup>2</sup>Point mutations choices were initially designed by using a python script based on [64], kindly provided by Prof. H.M. Salis (Penn State University). The tool was supposed to predict the relative decrease in strength of altered sgRNAs, compared with a perfectly matching 20-nt guide. While its reliability on independent datasets, different from the ones used as training sets, is beyond the scope of this work, the tool was used as starting point to obtain suggestions about designing mutations with diverse repression effects.

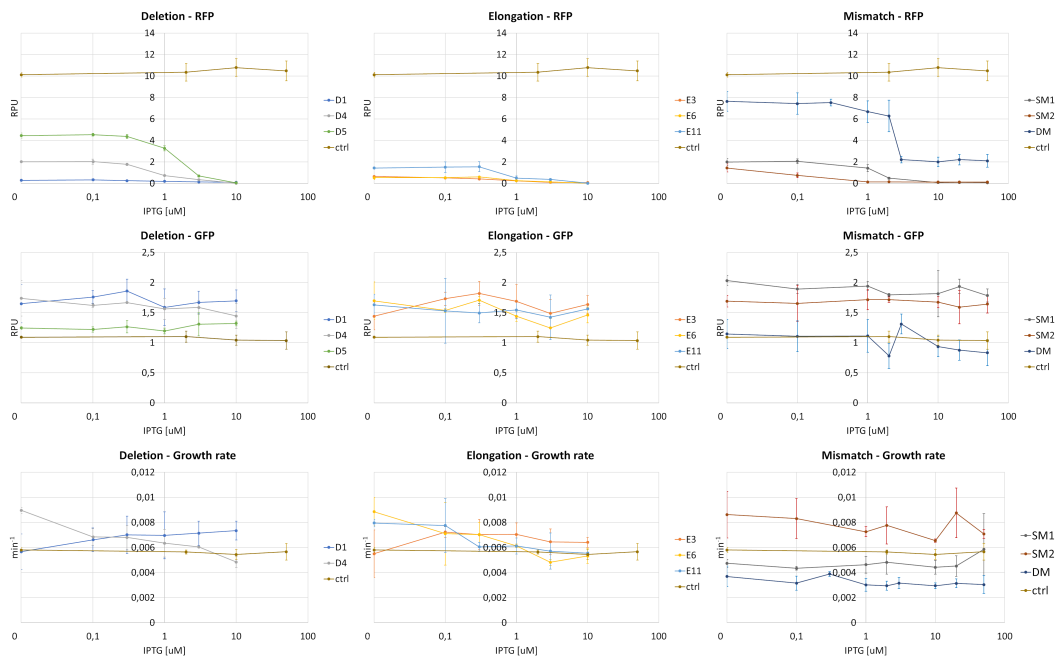
## 4.2. Approaches to improve tunability and repression range

**Table 4.1: List of altered gRNAs with relative sequences and feature.** Annealing region of the sgRNAs obtained via PCR mutagenesis using gPluxH as primer. Deleted bases are reported in **red** while substituted mismatching (MM) nucleotides are in **blue**; annealing adenines at 5' obtained by conserving the adenine after TSS of  $P_{LlacO1}$  of the template (i.e., by using RV\_Plac primer instead of RV\_Plac-A) are underlined.

Name	gRNA Sequence	Modification at/from 5'
<b>gPluxH</b>	TGACACCTGTAGGATCGTAC	Original 20bp annealing guide
<b>D1</b>	<b>T</b> GACACCTGTAGGATCGTAC	Curtailed with 1 deletion
<b>D4</b>	<u>A</u> <b>TGACA</b> CCTGTAGGATCGTAC	Curtailed with 4 deletions
<b>D5</b>	<b>TGACA</b> CCTGTAGGATCGTAC	Curtailed with 4 deletions
<b>E3</b>	<u>TCAT</u> GACACCTGTAGGATCGTAC	Elongated with 3 MM inserts
<b>E6</b>	<u>ATCTCA</u> TGACACCTGTAGGATCGTAC	Elongated with 6 MM inserts
<b>E11</b>	<u>ATAGGATCTCA</u> TGACACCTGTAGGATCGTAC	Elongated with 11 MM inserts
<b>SM1</b>	TGACACCTGTAGGAT <b>GG</b> TAC	Mutated with 1 MM at the 16 <sup>th</sup>
<b>SM2</b>	TGACACCTGTAG <b>T</b> ATCGTAC	Mutated with 1 MM at the 13 <sup>th</sup>
<b>DM</b>	TGACACCT <b>CTAGT</b> ATCGTAC	Mutated with 1 MM at 13 <sup>th</sup> and 10 <sup>th</sup>

Looking at the results reported in Figure 4.3, some interesting observation could be drawn. First, looking at the RFP, it was clear that elongation with mismatching nucleotides does not affect repression efficiency while deletions and mismatches allow higher improvements in repression tunability, both in terms of maximum RPU achievable and induction range. GFP signals, on the other side are quite stable for most of the constructs, with conserved ranking, except for the DM (double mismatch) strain. The same conclusions can be drawn for growth rates, stable for all the constructs except for the mismatching ones, where the stronger modifications (SM1 and DM2 respectively) shown a very low growth rate.

## 4. Further investigations



**Figure 4.3: Behaviour of gPluxH sgRNA containing curtailments, elongations and mismatches.** Upper row represent RFP data thus repression achieved by system expressing sgRNA dependently on the IPTG concentrations, central row is GFP which is in turn dependent to the cellular metabolic load, bottom row shows Growth rates of the constructs tested in this study. Data are grouped by type of modification per column: curtailments (D1, D4, D5) in the first column, elongations (E3, E6, E11) in the central one and mismatches (SM1, SM2, DM) on the right. The brown curves in all the diagrams represents control strains, bearing a non-targeting sgRNA.



## 4.2. Approaches to improve tunability and repression range

Overall, despite some of the modifications applied to the sgRNA sequence led to significant widening of repression ranges (e.g., D5 and DM, see Figure 4.4), none of the tested sgRNA allowed to completely unblock the repression exerted by the CRISPRi complex on the target promoter for un-induced systems.

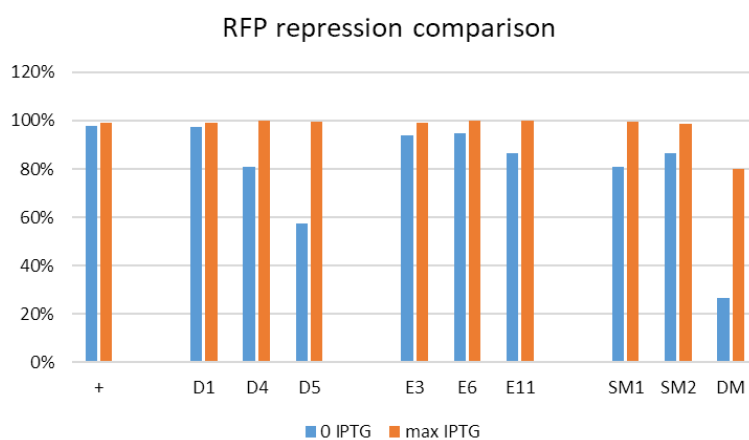


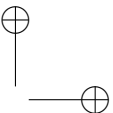
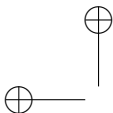
Figure 4.4: **Repression efficiency comparison between several modification approaches.** Data relative to minimum (zero IPTG induction) and maximum (full induction) repression of  $P_{luxRep}$  achieved by the tested guides are reported; percentage of repression are obtained normalizing the RFP minimum and maximum values of Figure 4.3 on a control strain bearing a non-targeting sgRNA. Each pair of columns represent one single type of modification made on the initial gPluxH sgRNA, reported in the bar diagram as “+”.



## 4. Further investigations

---

Deletions seemed to achieve the first rank in terms of efficiency and linearity of the modification effects; however, it is worth to notice that curtailing the annealing sequence implies a decrease in guide specificity, thus possibly leading to unwanted and hard to predict off-target effects. Mismatches also provided a sensible effect on repression strength tunability; still, the variation on the affinity of the sgRNA due to mismatch, thus repression strength is hard to predict and several trial-and-error experiment could be necessary to possible obtain an eventually optimal guide. With the used tool, for example, while a quantitatively predicted efficiency of the mutations in terms of efficiency decrease was provided, this output was not confirmed by experimental data. The software outcomes relied on a biophysical model estimated the mutation-induced variation in Gibson's Free energy ( $\Delta\Delta G$ ) and were calculated using parameter sets derived from two different datasets: *in vitro* and *all data*. SM1 analysis gave a value of 0.23 using the *in vitro* database and 0.2185 using the *all data* one, SM2 gave a 0.7461 with the *in vitro* and 1.4727 with the *all data* and DM2 gave the highest energies with scores of 1.286 and 1.472 for *in vitro* and *all data*, respectively. However, the mutation applied on gPluxH to obtain SM2 showed to be less effective than the one used for SM1 and anyways not comparable with DM2 results. Nonetheless, the sgRNA showing the wider repression tunability (DM) does not allow a full repression of the system on the other hand. While this approach seems to reserve promising perspectives, the lack of an effective rational to predictably design guides as well as the high nonlinearity in sgRNA sequence modification and repression alteration are however still harsh hurdles to overcome for properly exploiting this approach without an accurate computational tool to select suitable guide mutations.



---

## 4.2. Approaches to improve tunability and repression range

### 4.2.2 CRISPRi competition

As an alternative approach to optimize the repression range of CRISPRi systems, as well as to study the feasibility of using more than one guide sharing the same pool of dCas9 in a genetic circuit, a study on CRISPRi competition has been carried out.

Combining the parts used in Section 3.3.2 (i.e., HSL inducible sgRNA with constitutive dCas9 expression cassette and  $P_{LtetO1}$  target in high copy plasmid) with the constitutive sgRNA expression cassette exploited in Section 3.1.2, six new constructs were combinatorially assembled and tested (see Figures 4.5(a) and 4.6(a)): three of them drove the gPtet targeting guide transcription via HSL-inducible sgRNA expression system along with the synthesis of a competitive non-targeting gPlac guide through 3 kind of constitutive expression cassette with different strength (namely  $J23116 < J23100 < J23119$ ); the others, did the opposite (i.e., constitutive targeting sgRNA expression with HSL-inducible competitor guide). All the circuits bore the usual GFP metabolic burden monitor in low copy plasmid, constitutive dCas9 expression cassette in medium copy plasmid and  $P_{LtetO1}$  driving RFP expression in a target module carried by a high copy plasmid. While the response in absence of a competitor was studied in the previous chapter, a variation in the response of the system was expected in presence of a competitor. In particular, no variations in GFP and maximum RFP reachable were expected since the addition of the constitutive sgRNA expression cassettes does not load the circuits with further proteins to be translated nor should directly alter other protein levels. All the experiments were performed along with two control strains: the first one bearing (apart for dCas9 and target expression cassettes) only a HSL-inducible sgRNA expression device encoding the targeting gPtet guide; in the second one, bearing the same HSL-inducible cassette in which the sgRNA were substituted with the orthogonal non targeting gPlac (“only targeting” and “only competitor” controls,



## 4. Further investigations

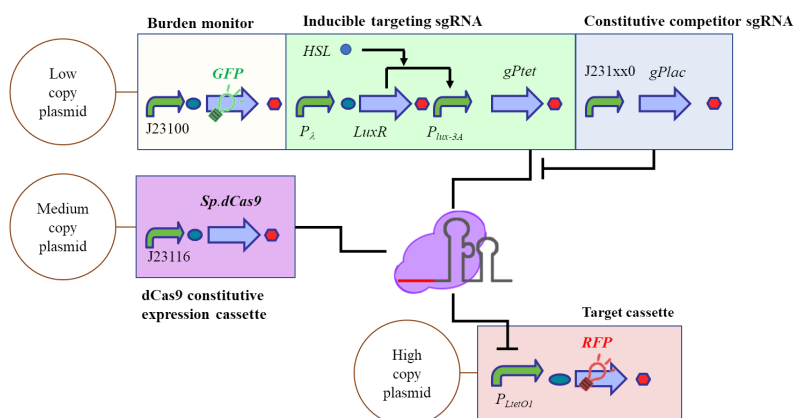
---

see Figures 4.5(b) and 4.6(b) respectively).

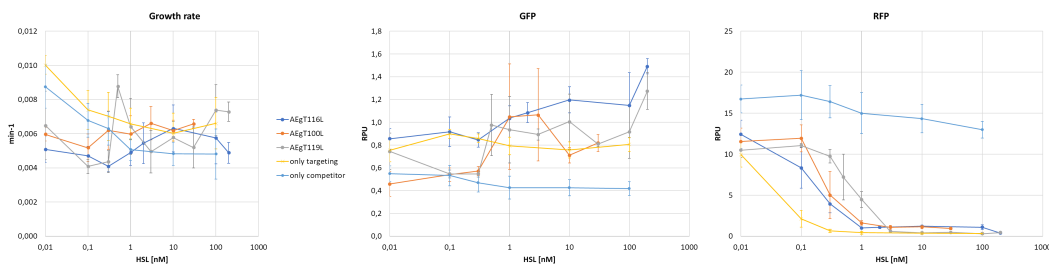
### **Inducible targeting sgRNA and constitutive competitor**

The effect of competition is clearly observed (Figure 4.5). While the only targeting control shows an early decrease in RFP signal, the response of circuits expressing competitor sgRNAs showed decrease for higher HSL concentrations in a trend overall conserving the strength of the constitutive promoter driving their transcription. On the other hand, both GFP and Growth rate signals were very noisy; growth rate in particular showed a remarkable decrease for high HSL concentrations, so further conclusions cannot be drawn in absence of further data.

## 4.2. Approaches to improve tunability and repression range



(a) Circuits schema.

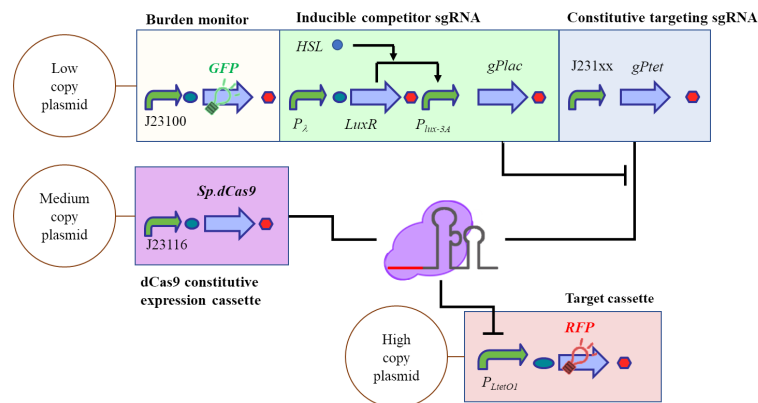


(b) Growth rate GFP and RFP expression comparison.

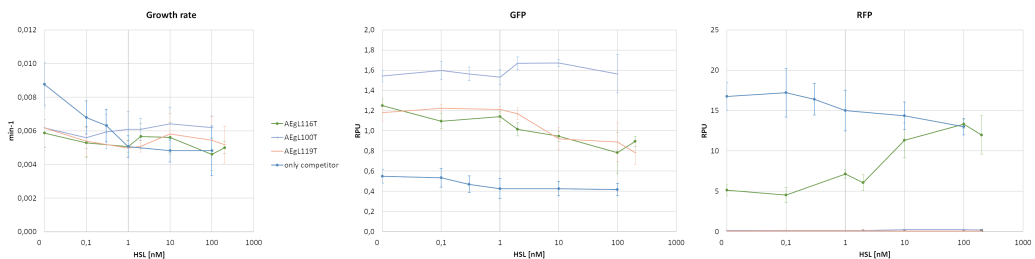
**Figure 4.5: CRISPRi competition with constitutive competitor expression.** a) The system configuration depicted by a genetic circuit in which a HSL-inducible targeting sgRNA and a constitutive competitor expression cassettes synthesize two guides competing for the same pool of free dCas9 proteins. In the constitutive competitor sgRNA expression cassette, xx stands for 16/00/19 which are the codes of the promoters used in this study, listed following increasing strengths. b) The graphs of growth rate, GFP and RFP expression are shown. The yellow “only targeting” curve represents a circuit bearing only the HSL-inducible targeting sgRNA expression cassette without competitor while the blue “only competitor” one represents a circuit with only an HSL-inducible sgRNA expression cassette but with the competitor guide coded in the sgRNA. The x-axis indicated the induction of HSL that is dependent on the amount of targeting sgRNA in the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## 4. Further investigations

### Constitutive targeting sgRNA and inducible competitor



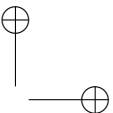
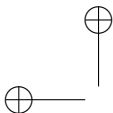
(a) *Circuits schema.*



(b) *Growth rate GFP and RFP expression comparison.*

### Figure 4.6: CRISPRi competition with inducible competitor expression.

**a)** The system configuration depicted by a genetic circuit in which a constitutive targeting sgRNA and a HSL-inducible competitor sgRNA expression cassettes synthesize two guides competing for the same pool of free dCas9 proteins. In the constitutive targeting sgRNA expression cassette, xx stands for 16/00/19 which are the codes of the promoters used in this study, listed following increasing strengths. **b)** The graphs of growth rate, GFP and RFP expression are shown. The blue "only competitor" one represents the circuit with only the competitor HSL-inducible sgRNA expression cassette without targeting guides. The x-axis indicated the induction of HSL that is dependent on the amount of targeting sgRNA in the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.



## 4.2. Approaches to improve tunability and repression range

---

In this configuration, the effect of dCas9 sharing is less explicit but still conserved (Figure 4.6); the RFP signal was completely repressed when targeting sgRNA was expressed under the two stronger constitutive promoters J23100 and J23119. The leaky expression of the specific sgRNA was enough to halve the repression exerted by the targeting sgRNA, when the expression of competitor was driven by the weak J23116; moreover, the system in this configuration reached a completely un-repressed state for high HSL concentrations hence high level of competitor. Curiously, while noise of growth rate and GFP signals is reduced compared with what seen for the circuit with inverted sgRNA expression cassettes, the unexpected ranking of GFP signal denied again a deeper interpretation of the results, as well as the again marked decrease in growth rate.

Overall, the finite amount of dCas9 inside a bacterial cell was proved to be an aspect that needs to be taken into account in design of genetic circuits. Despite these results need to be reinforced with more case studies and an appropriate mathematical analysis (ongoing), the feasibility of using sgRNA competition as a degree of freedom to tune the efficiency of CRISPRi system, as well as the unavoidable necessity of a formal description of this aspect to use CRISPRi system in rational design of synthetic genetic circuits, has been highlighted.



## 4. Further investigations

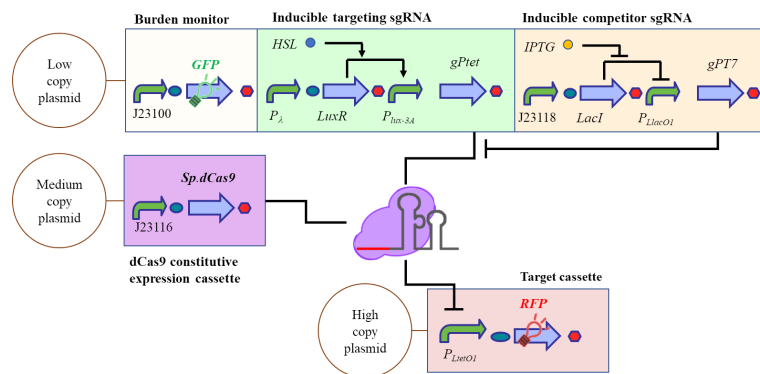
---

### Double knob circuitry

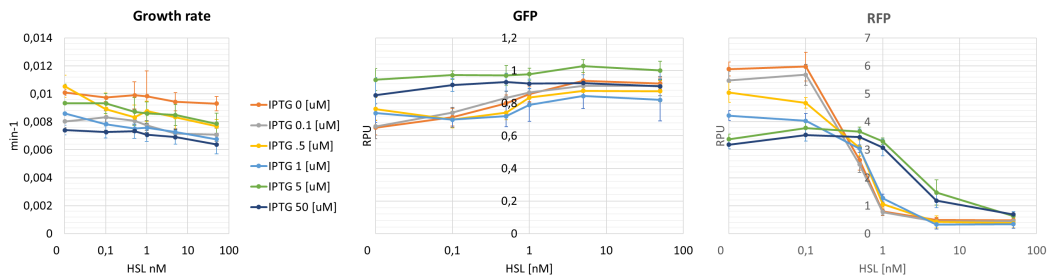
In order to thicken the grid of combinatorial concentrations of sgRNAs, a “double knob” version of the CRISPRi competition model were developed, by expressing both of the guides under the two inducible sgRNA expression cassette described in Section 3.2.2. The circuit, as reported in Figure 4.7(a) bore a HSL-inducible cassette to express the targeting guide and an IPTG-inducible one to express the competitor. Since the guides used in the previous works had a possible target in the parts composing this circuit (i.e., gPtet used to repress the target, gPlac could repress the competitor expression, gPluxH could repress the targeting sgRNA expression due to the high homology between  $P_{lux}$  and  $P_{luxRep}$ ), a new sgRNA to be used as orthogonal competitor was developed; this guide, called gPT7 with sequence 5'-TACTAGAGAAAGAGGAGAAA-3', was obtained by taking the sequence targeting a T7-phage family promoter, with no homologies with *E. coli* genome nor with any of the promoters used to build the circuits. Such system was supposed to provide RFP responses with higher switch points (i.e., higher HSL induction needed to fully repress the system) for increasing IPTG inductions, due to the competition effect exerted by the competitor guide when synthesized in amounts high enough to saturate the free dCas9 protein pool. Despite the expected behavior was observed for increasing IPTG inductions in terms of HSL needed to achieve a strong repression, a monotonically decreasing trend of maximum RFP synthesis rate was observed for increasing IPTG concentrations with low HSL in the media, while the competitor guide increases RFP repression for small amounts of targeting sgRNA. GFP on the other hand maintained the slightly increasing trend correspondent to a decrease in RFP expression and an overall reduced level due to the expression of 3 plasmids and 4 expression cassettes.



## 4.2. Approaches to improve tunability and repression range



(a) Circuits schema.

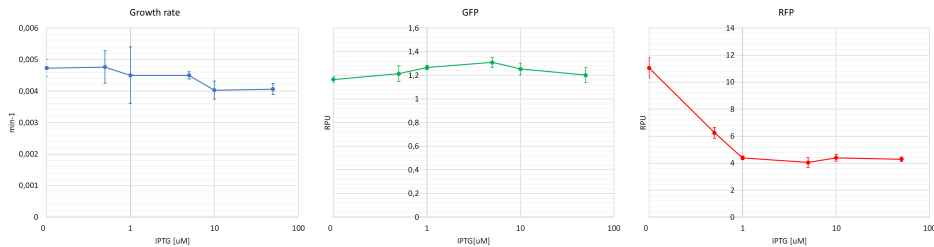


(b) Growth rate GFP and RFP expression comparison.

**Figure 4.7: CRISPRi competitor with double inducible system.** **a)** The system configuration, showing the two sgRNA expression devices driving the transcription of two competitive guides sharing the same pool of dCas9 protein. **b)** The graphs of growth rate, GFP and RFP expression are shown. Curves colors represent different IPTG induction and thus different amounts of competitor sgRNA. The x-axis indicated the induction of HSL that determines the amount of targeting sgRNA in the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## 4. Further investigations

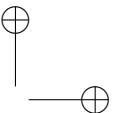
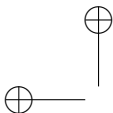
In order to evaluate how the competitor guide affects the system response without any other sgRNA expressed in the cell, a strain bearing only the competitor, constitutive dCas9 and the  $P_{LtetO1}$ -driven RFP target expression cassettes was tested at different IPTG concentrations (see Figure 4.8). From this analysis it was clear that the competitor



**Figure 4.8: Effects of competitor expression on RFP expression.** The graphs of growth rate, GFP and RFP expression are shown. The schema of the circuit is the same as above but lacking of the HSL inducible sgRNA expression cassette. The x-axis indicated the induction of IPTG that determines the amount of theoretically non-targeting sgRNA in the cell. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

guide was indeed able to exert a certain repression on the target promoter: in presence of a leakage of the targeting sgRNA (HSL= $0nM$  in Figure 4.7) in the double knob version of the circuit, competitor provided a repression leading to a RFP from  $6RPU$  (at IPTG= $0\mu M$ ) to  $3RPU$  (at IPTG= $50\mu M$ ); the repression given by solely the competitor sgRNA led from an almost un-repressed state when non induced (as did the control bearing gPluxH in Figure 3.19 in an almost identical configuration except for non-targeting sgRNA used<sup>3</sup>) to  $4RPU$  for IPTG= $50\mu M$ . Moreover, while the synergistic effect of the two guides could be seen in the decreasing of the maximum synthesis rate

<sup>3</sup>In this study, the already proved non-targeting gPluxH sgRNA could not be used due to its affinity with  $P_{lux-A}$  driving the expression of the targeting guide in the double knob configuration.



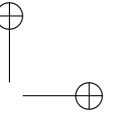
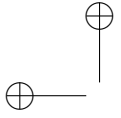
## 4.2. Approaches to improve tunability and repression range

---

of the two strains at different IPTG concentrations and no HSL, a sort of competitive effect can be seen looking at the switch points of the curves that increase as the IPTG concentration thus competitive sgRNA concentration rises.

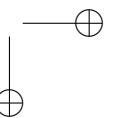
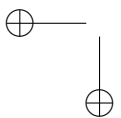
The reason of the unexpected efficiency of gPT7 was found aligning its sequence against the whole plasmid. Indeed, during the design of the competitor sgRNA, the sequence were checked to lead no-off target effects just on the host genome and other promoters bore by the plasmids composing the circuits; however, 12 nucleotides at the 3' part of the sgRNA annealing part were found complementary to a region - upstream a previously unnoticed PAM sequence - covering the junction scar and part of the RBS downstream the  $P_{LtetO1}$  target promoter driving the expression of the RFP. As shown in [65], CRISPRi exerted on a non promoter region is less effective but still functional despite the severe decrease in efficiency due to the remaining mismatching nucleotides at the 5' of the sgRNA.

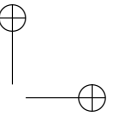
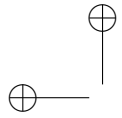
Overall, despite the double knob circuitry led to results not looked for, all the experiments carried out in both single and double knob configurations of the CRISPRi competition model circuits led to the evidence of a hidden regulation layer that must be taken into account and possibly exploited in rational design of synthetic genetic circuits expressing more than one sgRNA and a limited amount of dCas9 enzyme, confirming the hypothesis introduced in [108]. Moreover, even if more experiments must be performed (e.g., by using a proper non-targeting guide in the double knob configuration) this work lays the foundations for further and deeper investigations about yet unexplored dCas9 sharing mechanism and CRISPRi repression efficiency control via multiple sgRNA competition effects.



#### 4. Further investigations

---



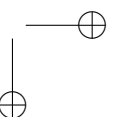
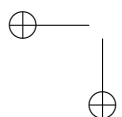


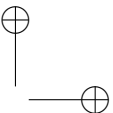
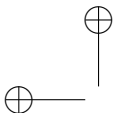
# Chapter 5

## Discussion and conclusions

The bottom-up characterization of genetic devices is essential for the interpretation and forward engineering of complex genetic circuitry. Although obstacles still exist, the currently persisting unpredictability sources are being tackled by different groups. On that point, metabolic burden resulting from the expression of recombinant proteins in bacterial cells is still one of the major issue hampering a successful rational design of synthetic circuits.

In this work, a model explicitly including parameters describing the load from each synthesized gene has been evaluated and tested on a case study obtaining overall good results: while a quantitative prediction of all the tested circuits has not been completely achieved and further improvements need to be done to improve the metabolic load monitoring system, the use of the model allowed to predict circuits whose behavior could not be described otherwise, due to a high metabolic load in the cell. Discussions below highlight the worth, necessity and feasibility of modeling metabolic load in synthetic biology. Since the mathematical description of transcriptional load highlights the working boundaries of a design circuit but does not provide an





## 5. Discussion and conclusions

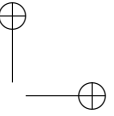
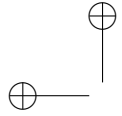
---

actual solution to overcome its effects, a CRISPR interference based system has been developed and characterized in many versions, as a design alternative to overcome the hampering effect due to metabolic burden. Despite improvements to widen the tunability of the system are still ongoing, the desired overall little load has already been achieved; the promising results discussed above, include also a reinforcement of the observed “low-burden trend” obtained by exploiting the same mathematical model studied adopted before.

**Metabolic burden study.** In the first part of this work, recently proposed in vivo and in-silico methodological approaches have been adopted to face the long-standing issue of biological devices predictability for the rational design of synthetic circuits. Namely, a reporter expression cassette was used to quantify cell load and a mathematical model, which explicitly describes global burden-related effects on protein synthesis levels, was adopted for predictions. A set of ad-hoc constructed genetic circuits implementing transcriptional cascades was used as a testbed by following a rigorous bottom-up design process, including the learning of individual modules function (using a training set) and the evaluation of model predictions on a previously unseen circuit collection (using a test set).

It is worth to notice that despite other mathematical models could explain a similar behavior by including hidden regulations and feedbacks, it is hardly unlikely that the response of the systems came from other mechanisms (see Appendix B.8).

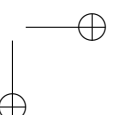
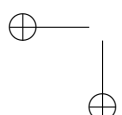
From a qualitative point of view, considering the inducible/repressible behavior of the circuits, the used model systems collection included circuits exhibiting expected HSL-dependent trend and a small set of configurations showing apparently unexpected outputs. The circuits spanned a wide range of RFP levels, corresponding to circuit output, and also showed diverse growth rates and GFP levels, indicating a

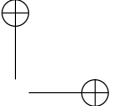
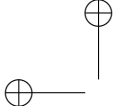


---

variation of cell burden for different circuits and HSL inducer concentrations. Such statistics suggest that the considered collection includes sufficient variations in the observed variables to test the model descriptive and predictive capabilities.

A model-free correlation analysis of the measured data showed a strong negative correlation between GFP and RFP levels in the 1-block circuits, including only the input module with RFP downstream. This was expected, since high RFP expressions cause an increase of cell burden [19, 54]. Circuits with more than one block, on the other hand, did not show such trend, since the entity of cell load is expected to be not only RFP-dependent, but also function of the expression of the other circuit-borne proteins. On the other hand, the previously observed strong correlation between GFP (expressed via chromosomal constitutive cassette) and growth rate [51] was not observed in our data, which showed a weak correlation. This difference could be due to several factors, discussed below. By analyzing only the circuits exhibiting the lowest GFP values (corresponding to higher cell burden), their correlation with growth rate is relevant, while for the others it is non-significant. Circuits exhibiting a lower cell burden did not show a relevant variation in growth rate, but a clear HSL-dependent GFP trend could be observed. These results suggest that a GFP-growth rate correlation can be only observed if the circuits are affected by a relevant cell burden, and that the use of a constitutively expressed reporter protein has a clear superior performance, in terms of sensitivity, over the traditional use of growth rate for cell burden monitoring. Moreover, in this work we analyzed a different genetic context (Monitor cassette placed in plasmid instead of chromosome), and used different experimental protocols (HSL addition and growth to reach the steady-state of intracellular species) and data analysis methods (growth rates computed over the whole exponential growth phase, and GFP computed as fluorescent protein synthesis rate per cell) than the previous work [51].





## 5. Discussion and conclusions

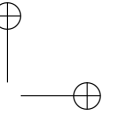
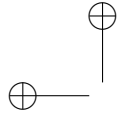
---

Two models were compared in this work: NBM and BM. While the former uses RFP and growth rate data, the latter also uses GFP to eventually estimate and exploit the burden-related resource usage parameters. Both models were able to accurately fit the data of the training set and their overall quantitative prediction capability was comparable. Nonetheless, the BM allows to predict the output trend of one non-functional circuit exhibiting an unexpected output which could not be predicted by the NBM. Indeed, the behavior of this circuit can be successfully explained only by modelling cell load.

However, another circuit exhibiting an unexpected output trend could not be predicted even by the BM.

It is worth noting that the two circuits with unexpected logic behavior analyzed in this study are not robust logic gates, since the difference predicted by both the NBM and BM between on-state and off-state is very low. In fact, despite the transfer functions of all the devices have a wide induction range (see Figs.B.9,B.10), their output range may not be entirely spanned in the interconnected configurations. In these two circuits, the transcriptional input provided to the lac-based NOT gate is predicted to exert a detectable but not tight repression on  $P_{LlacO1}$ , thereby covering a small part of its available output range. Moreover, the on-state has a transcriptional output activity comparable with a medium-strength promoter and the off-state has a high basic activity. While software tools able to guide the design of robust functions have been proposed [46], the design of robust gates was beyond the scope of this work: we limited our study to the analysis and predictability of the qualitative and quantitative output observed from the interconnection of pre-characterized modules. The previously proposed software system also considers cell load by identifying the part configurations in which a relevant impairment of cell growth was observed. The BM used in our study could be a further support in the rational engineering of genetic circuits, also from the knowledge of the resource usage of all the involved proteins (see below).

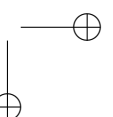
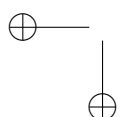


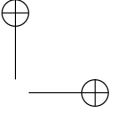
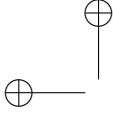


---

The NBM and BM were also systematically compared in terms of descriptive capabilities. To carry out this task, all the available data were considered and fitted with the NBM and BM. The BM captures circuits output with higher accuracy than the NBM, which is unable to capture the experimental output exhibited by some of the circuits even qualitatively. An over- or under-estimation of RFP and GFP outputs, up to 1.5 and 2.1 fold, respectively, were still present; such fold change values are reasonably contained, and comparable to other studies focused on predictability [13, 40, 84]. Context-dependent variability might cause such variation. Previous studies proposed a linear model-based method to score the quality of part collection members, relying on the characterization of their activity in different genetic constructs [45]. Analogously, because of context-dependent variability of parts, the BM may be unable to explain all the variation observed in the experimental data, and the same devices measured in different context can show diverse activity. In case of simultaneous fitting, the estimated parameter values represent the average values that best describe parts behavior considering all the circuits. On the other hand, if a parameter value is estimated on a single training set circuit, it represents the specific value of the analyzed one. In the training set circuits used in this study, all the model parameters have been estimated considering a single circuit, except the burden-related parameters, which were the result of a simultaneous fitting considering all the circuits including the protein causing resource usage.

The fitting of the training set data by the BM enabled to estimate resource usage parameters for the proteins involved in the cascades. With the used RBSs, TetR was found to cause more cell load than LacI (i.e., the other repressor used) and RFP. This resource usage ranking was confirmed through the parameter estimation results of the simultaneous fitting of all the available data (not only training set circuits) by the BM. The contribution of LuxR to cell load was estimated as a non-dimensional parameter lumping the product of resource usage pa-





## 5. Discussion and conclusions

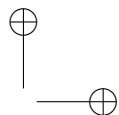
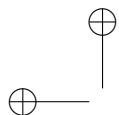
---

parameter and maximum LuxR synthesis rate per cell, since, differently from the other circuit proteins, the latter was not estimated in this study. For this reason, in the present work the resource usage of LuxR cannot be directly compared with the other considered proteins.

Based on the resource usage estimation via BM, we finally fixed the two circuits with unexpected behaviour by decreasing the translation of the gene causing the main load to the cell. To this aim, a 10 fold decrease in TetR translation efficiency successfully restored a correct function for both circuits, and the NBM was able to accurately predict the observed output. The obtained results demonstrate the usefulness of BM in the identification of the modules causing excessive cell load and the successful utilization of resource usage knowledge to drive the rational re-design of predictable circuits.

Based on the cell burden measurements and modeling, in both a bottom-up and global fitting fashion, and on the subsequent study of low-burden variants, we enabled to confirm that the unexpected behavior of the  $X_1TL$  and  $X_{rep}TL$  configurations was due to TetR over-expression, which caused excessive cell load. Despite tetR expression affects the two circuits by breaking their logic behavior, it is worth noting that cell burden also affects other circuits without breaking their function. In particular, according to GFP measurements and BM predictions, TetR exerts a high load in the  $X_2T$  and  $X_1T$  configurations, even higher than in  $X_{rep}TL$ . However, this affect is not visible, since the high-load condition persists when the circuit output is low, thereby masking any burden-induced effects on the expected logic function of these circuits.

Taken together, our results showed that the use of the BM has advantages over using the NBM, in terms of predictability of some configurations in bottom-up approach, descriptive power of circuit and Monitor output, estimation of load-corrected transfer function parameters and estimation of the resource usage parameters, which can support the rational design of circuits with predictable function.

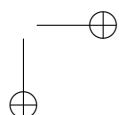
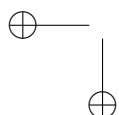


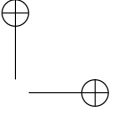
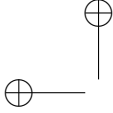
---

However, several steps still need to be carried out to improve the predictability of models like the ones used in this study. For instance, here the growth rate was fixed in the model for each circuit and HSL concentration, without any effort to predict it from the specific circuits used. The knowledge of growth rate could affect protein dilution, and recombinant strains with different dilution rates may exhibit diverse quantitative circuit behaviors. The growth rate prediction task is hampered by its poor predictability as a function of GFP value by the Monitor. In addition, cell growth rate may have a relationship with the amount of resource pools (which are assumed to be constant in the BM used in this study). New models considering this aspect could give significant benefits in the description and prediction of experimental data from synthetic circuits, as well as improve the understanding of biological systems features.

Overall, this results show the outcome achieved via bottom-up design process considering limited cell resources and demonstrate the need of further efforts to improve models for biological engineering, to disclose hidden interactions among biological systems elements.

**CRISPRi toolbox development.** In the second approach presented in this work, different investigations focused on the common goal of facilitating the use of CRISPRi to build novel customized gene regulatory networks with low cell load for the host were provided. Specifically, the first investigation was aimed at finding a low-toxicity high-efficiency trade-off for the constitutive expression of dCas9, to obtain an optimized fully functional minimal burden dCas9 expression cassette. The results showed that a low transcription rate on a medium-copy plasmid, which is co-transformed in engineered bacteria in all the quantitative experiments done in this work, was successful. After the optimization of dCas9 expression, a number of synthetic



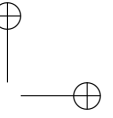
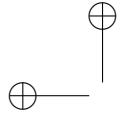


## 5. Discussion and conclusions

---

circuits in which sgRNAs were expressed by inducible promoters to target RFP expression by binding to its upstream promoter were built and characterized. Among the carried out tests, three sgRNAs (targeting  $P_{LtetO1}$ ,  $P_{LlacO1}$  and  $P_{luxRep}$  promoters), expressed by  $P_{lux}$  or  $P_{LlacO1}$  (which required optimization before being used as effective *expression knobs*), with target promoters in medium- or high-copy plasmid, showed to be all functional in terms of repression. As expected, the target copy number plays a role in the output curves by showing tighter repression by sgRNAs when target is in medium-copy compared to high-copy in which the tunability range of repression was higher. The CRISPRi system is more flexible in the choice of target than other specific protein transcription factors. It is worth noting that, despite all the tested sgRNAs designed in this work are actually repressors, the repression efficiency strongly depends on the sequence, e.g., the sgRNA designed to target  $P_{LtetO1}$  exerts a much stronger repression than the one targeting  $P_{LlacO1}$ . In some cases, repression was so strong that the output showed very low RFP expression for any sgRNA level. This means that the basic activity of inducible promoters driving sgRNAs was sufficient to produce sgRNA levels with good repression capabilities. A constitutive GFP expression cassette was used as a proxy for cell load, as recently described by other groups. GFP data showed that no relevant burden affects the cells upon sgRNA expression. Importantly, metabolic load was much lower than the one observed in some repression systems based on widely used protein repressors (e.g., TetR), thereby suggesting that the use of CRISPRi in synthetic circuits as a low-burden alternative to transcription factors can be successful.

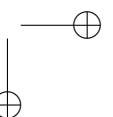
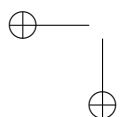
These results were confirmed by the mathematical analysis of the system: by applying an adapted version of Burden and No Burden models introduced in the previous part of the manuscript, the overall negligibility of the load exerted by the CRISPRi repression device to describe the system were proven. Interestingly, three other features

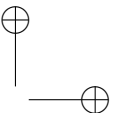
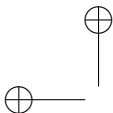


---

have been highlighted via mathematical modeling: firstly, to obtain a omni-functional model, deeper investigation are needed to describe sequence-dependent repression strength for guides; second, predicting copy number variation effects (e.g. copy number-dependent parameter variations) is still a major issue to be formally faced, developing a model that explicitly and mechanistically take into account this aspect. Lastly, since parameters describing GFP burden monitor base level and RFP load contribution showed a strong context dependency, a deeper investigation on this aspect or the development of a more reliable measurement system is still needed. To better understand the possibly occurring interplay between CRISPRi system and commonly found genetic circuit designs, we studied the effects on sgRNA repression capability as a function of transcriptional activity of target promoter. It resulted that fold repression caused by sgRNA was not dependent from promoter strength: a sgRNA targeting a small library of promoters having graded activity but sharing the same core region (in which the guide RNA binds) showed reasonably similar repression curves. This result enables the predictable re-use of previously designed sgRNAs to target different promoters sharing a conserved sequence. In general, despite no relevant cell burden was detected in the sgRNA-based circuits, some of the systems showed slight GFP and growth rate drops which were not expected. Similar effects are still under study by many groups and their full understanding is expected to improve the predictability of bottom-up designed biological systems. Having seen that the constitutive expression of dCas9 was effective in this configuration, further decreasing its expression could allow for a larger range of RFP values.

Taken together, these results suggested that further works are needed to rationalize sgRNA design with a desired repression strength, and that CRISPRi can be successfully used to design low-burden solutions. In this direction, possible methods relying on the competition between orthogonal sgRNAs sharing the same pool of dCas9 to be bound with





## 5. Discussion and conclusions

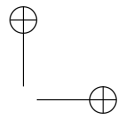
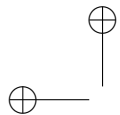
---

- assuming that it was a limited resource able to be depleted - and on altering the affinity of sgRNAs, changing their length or introducing mismatches into the complementary sequence, were explored.

In the former case, curtailments of guides and mismatching guides provided interesting preliminary results, showing relevant widening of tunability; however, whether the use of shorter guides could lead to undesired off target effects due to the decrease in sgRNA specificity, to relate mismatches number and position in the annealing region with a desired repression decrease is still a major issue. While more experiments are needed to search for a rational design of mismatches, many in-silico approaches have been published in literature based on machine learning or biophysical models [103, 109, 64] and a deeper study of the available algorithms could plot a course to standardly exploit this method in rational design.

For what concern CRISPRi competition, results are very encouraging, since they showed the possibility of using such method to indirectly alter the repression efficiency of sgRNAs, despite this method is expected to globally affect all the guides of a circuit. Moreover, the study of the mechanism per se paves the way to a better understanding of the behavior of an ever more used despite not completely yet known molecular tool. On that point, experiments with a weaker dCas9 expression cassettes are under study as well as a double knob systems expressing a truly orthogonal competitor guide. It is worth to notice that a double knob configuration is hardly going to be a ready to use module for rational design of synthetic genetic circuits with reduced metabolic load, being the usage of two different sgRNA expression cassette too much resource demanding for the related transcriptional regulators as inputs.

Lastly, we found that transcriptional protein repressors and sgRNAs, targeting overlapped regions on a promoter of interest, can be simultaneously used to exert repression. In general, sgRNA-based repression was much stronger than the protein-based one, although a tuning of



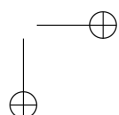
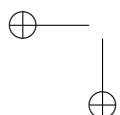
---

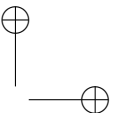
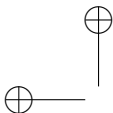
repressor expression should be done to confirm this point. The shape of the sgRNA-dependent repression curves was affected by the protein repressor in a level-dependent fashion: target expression did not only show a decrease (expected for the expression of two repressors) but also a change in sensitivity (lower for higher levels of protein repressor) in terms of sgRNA level causing expression drop. Despite additional experiments should be carried out for other expression systems to confirm this conclusion (currently evaluated on the  $P_{luxRep}$ -based system), the results suggest that the two repressors compete for the shared binding site and this effect should be taken into account when engineering new synthetic circuits.

In an effort to further facilitate CRISPRi-based circuits and enhance the reproducibility of our work in the synthetic biology community, the essential elements of the circuits (dCas9 expression cassette, tracr sequence, and mutagenized inducible devices) were cloned to conform with the BioBrick<sup>TM</sup> standard, used by labs worldwide interested in synthetic biology, and have been submitted to the open-source MIT Registry of Standard Biological Parts as standardized and pre-characterized genetic tools, now available to any lab interested in transcribing gRNAs to build novel CRISPRi circuits.

Further studies are still needed to fully elucidate and characterize the system behavior: more mathematical modeling investigations aimed at predicting the behavior of genetic circuits in all the implemented configurations and case studies can be of support to increase knowledge on the behavior of the devices herein developed. Noticeably, this work has offered an enrichment to the synthetic biology community, providing two new standardized synthetic genetic devices, a huge amount of data to characterize a system not yet in place and inspiring applications for a variety of purposes, such as the induced silencing of a gene to observe for a change in phenotype without having to create a deletion strain.

Despite investigations of CRISPRi systems in organisms other than



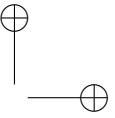
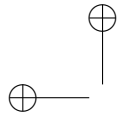


## 5. Discussion and conclusions

---

*E. coli* fall out of the scope of this study, the results of this work will greatly facilitate the future use of CRISPR/dCas9 as customizable and low-burden transcriptional regulators in synthetic circuits and complex metabolic pathways.





# Appendix **A**

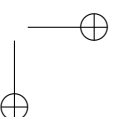
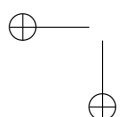
## Successful rational design example: Synthetic close-loop controller<sup>1</sup>

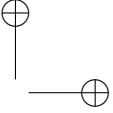
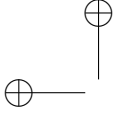
### A.1 Introduction

Feedback control is ubiquitous throughout nature [111]. Living systems exploit different control strategies to adapt and survive in changing environments, e.g., by enabling disturbance rejection to environmental fluctuations or by generating population heterogeneity [112]. Likewise, synthetic biology can take advantage from nature- and also engineering-inspired control strategies to implement novel customized functions in living systems, thereby boosting the reachable complexity of synthetic circuits design [113, 114]. Experimental and theoretical studies on negative feedback regulators were performed in both in vitro and in vivo settings. Control schemes were implemented at different levels (transcriptional [115], translational [116], or

---

<sup>1</sup>This chapter is taken from the article “*A synthetic close-loop controller circuit for the regulation of an extracellular molecule by engineered bacteria*”[110]



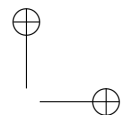
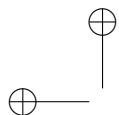


## A. Synthetic close-loop controller

---

via biochemical reactions [117]) in different chassis, from bacteria to mammalian host cells. Previous circuits realized *in vitro* exploited the advantages of negative feedback to achieve automatic regulation of gene expression and adaptive control of complex metabolic pathways [118]. Noise reduction [119, 120], noise spectra shaping [121], and rise time decrease in the dynamic response of gene networks [122] have also been reported. Gene expression close-loop control was used to automatically regulate the expression of a target protein to minimize the formation of unwanted inclusion bodies in *Escherichia coli* [52], regulate the density of a bacterial population by dynamically controlling the expression of a killer gene [123], implement RNA-based controllers to maintain target protein levels in mammalian cells [124, 125], and also implement a negative feedback RNA circuit *in vitro* [126]. Metabolic pathways for the biosynthesis of lycopene [127], fatty acids [128, 129, 130], 3-hydroxypropionic acid [131] and the bioconversion of lignin constituents [50] were improved by using toxic or accumulating intermediates, cellular stress-related molecules and nutrient availability as control signals. Negative feedback architectures have also been fundamental in the realization of gene expression oscillators [132, 133], genetic memories [132, 134] and biological device insulation [114], via disparate circuit designs.

Given a biological system, e.g., a gene regulatory network or metabolic pathway, it is possible to implement an *in vivo* feedback control of one or more target molecules by following the modular structure of a feedback-regulated system based on automatic control theory [127]. The requirements to implement such strategy in synthetic circuits are: i) a biosensing element, e.g., a transcriptional regulator protein, that is able to measure a proper signal such as the output molecule or an intermediate reaction product, ii) a regulator, e.g., a regulated promoter, that converts the biosensor output into a transcriptional signal, and iii) an actuator, e.g., an enzyme or another transcriptional regulator, that directly (via enzymatic reactions) or indirectly (via genetic

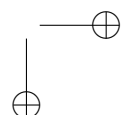
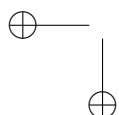


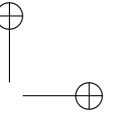
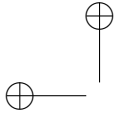
---

## A.1. Introduction

regulation of pathways) affects the target molecule according to the transcriptional signal from the regulator. While this work is focused on in vivo implementation of a feedback circuit, it is worth mentioning that ex vivo strategies have also been proposed for yeasts and mammalian cells [135, 136, 137]. In these cases, the (fluorescent) target molecule sensing is carried out via microscopy or flow cytometry and the regulation logic is implemented via external controller, which is able to close the loop by driving the injection of inducer molecules that serve as actuators. Such strategy can complement the in-cell implementation by enabling the study of different control algorithms, to tackle the intrinsic complexity and nonlinearity of in vivo regulation, and to avoid extensive genetic modifications to realize the feedback circuit [114].

Despite numerous recent studies on negative feedback in synthetic biology, two key challenges still persist, which currently hamper our ability to implement complex control systems. First, the availability of biosensors with features meeting design specifications (e.g., activity range, switch point, and induction stability) can be limited, thereby requiring significant efforts towards the search, de-novo construction, optimization and testing of candidate biosensing elements that can be successfully connected in the close-loop circuit. These aspects are extensively discussed in recent works [138, 139, 140] and will not be herein addressed. Second, the bottom-up design of accurately predictable systems is a major issue in biological engineering, leading to time-consuming trial-and-error steps during the construction of the desired circuits [23]. This rational design approach has been rigorously tested in a small number of studies, mainly involving synthetic circuits with relatively simple architectures. In case of more complex feedback-regulated circuit implementation, this issue is undeniable, since the successful realization of automatic control schemes requires the screening of different candidates [52, 128]. Mathematical modeling has been used to support the early design steps of synthetic circuits,





## A. Synthetic close-loop controller

---

as well as to guide possible debugging steps and capture the final circuit behavior [92, 141]. Models have also been used to study feedback control schemes and properties in a number of works [142]. However, a rigorous model-based engineering-inspired bottom-up design strategy for synthetic controllers is still a challenge.

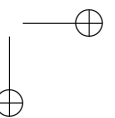
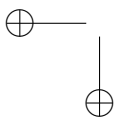
In this work, we address this key issue by testing a bottom-up approach to design a synthetic close-loop controller in engineered *Escherichia coli*, aimed to automatically regulate the concentration of an extracellular molecule, N-(3-oxohexanoyl)-L-homoserine lactone (HSL) [143]. After a preliminary selection of genetic parts and model-based circuit design, we individually characterized all the genetic circuit subsystems and fully identified a mathematical model of the complete close-loop circuit. We finally tested the full circuit to demonstrate the predictability of the behavior of the system and to evaluate the robustness of the designed architecture, compared to a related open-loop scheme, i.e., a circuit including an HSL production module without feedback control.

### A.2 Close-loop circuit design

#### Circuit description

The circuit was designed by re-wiring genetic elements of the quorum sensing and quenching mechanisms of *Vibrio fischerii* (*luxI* and *luxR* genes, and the cognate  $P_{lux}$  promoter) and *Bacillus sp. 240B1* (*aiiA* gene), respectively [133, 143, 144, 145].

The *luxI* gene, encoding an HSL synthase enzyme (LuxI), was placed under the control of the synthetic  $P_{LtetO1}$  promoter, to obtain an HSL production cassette that can be triggered by anhydrotetracycline (ATc) (see Figure A.1). The selected host strain, MG1655-Z1,





## A.2. Close-loop circuit design

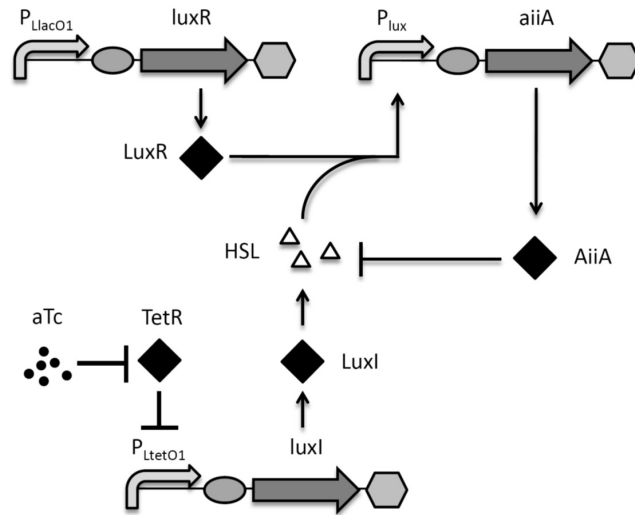
---

bears a TetR repressor expression cassette in the genome and enables  $P_{LtetO1}$  induction via ATc [88]. HSL is a small, non-toxic molecule that freely diffuses through bacterial cell membranes, and intracellular and extracellular concentrations are thereby assumed to be equal. The described system implements an open-loop HSL production system.

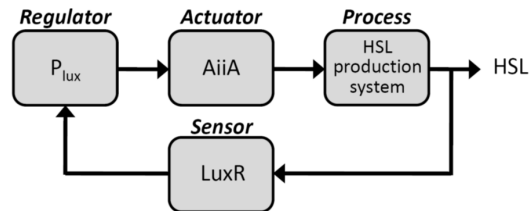
To implement a negative feedback regulation of HSL, a sensor-regulator-actuator set is needed. The constitutively produced transcriptional regulator LuxR, encoded by the *luxR* gene, binds HSL and the formed complex can induce the  $P_{lux}$  promoter transcriptional activity in an HSL-dependent fashion [13, 79]. The  $P_{LlacO1}$  promoter drives the constitutive expression of *luxR*. Although this promoter is repressed in MG1655-Z1, due to the genomic presence of an over-expression cassette of the cognate LacI repressor, the basic activity of  $P_{LlacO1}$  in the repressed state is not null and was used to express *luxR* [13]. The *aiiA* gene, encoding a lactonase enzyme (AiiA) and placed under the control of  $P_{lux}$ , is able to degrade HSL, thereby closing the loop and enabling HSL concentration control. Following the biology-electronics analogy proposed by Farmer et al. [127], in this architecture LuxR acts as HSL biosensing interface. On the other hand,  $P_{lux}$  acts as the circuit regulator, since it receives the information from HSL-bound LuxR as input and provides a transcriptional signal, driving the actuator, AiiA, as output. The ribosome binding sites (RBSs) of *luxI* and *aiiA* were selected among a list of experimentally screened candidates to maximize HSL production and degradation rates, respectively (for details, refer to A.6).

As it is observed in many existing biological elements, the transfer function of one or more modules is nonlinear, thereby making the system quantitative behavior non-trivially predictable without the help of mathematical modeling.

## A. Synthetic close-loop controller



(a) *Synthetic circuit implementing the close-loop controller in E. coli.*



(b) *Control engineering-inspired block diagram equivalent to the synthetic circuit.*

**Figure A.1: Close loop genetic controller schema.** a Curved thick arrows: promoters; ovals: RBSs; straight thick arrows: genes; hexagons: transcriptional terminators; dark diamonds: proteins; small circles: aTc; triangles: HSL; thin arrows: activation or HSL biosynthesis; thin truncated arrows: repression or HSL degradation

## Mathematical model of close-loop controller

A continuous culture mode (turbidostat) was adopted to maintain cells in exponential growth phase at a target density. The following ordinary differential equation (ODE) model was used to describe the close-loop controller system in cells cultured in a turbidostat (Eqs.A.1,A.2,A.3,A.4).

$$\frac{dX}{dt} = \mu \cdot X - D \cdot X \quad (\text{A.1})$$

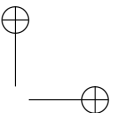
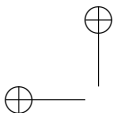
$$\frac{dL}{dt} = \alpha_{tet} \cdot \left( \delta_{tet} + \frac{1 - \delta_{tet}}{1 + \left( \frac{K_{tet}}{T} \right)^{\eta_{tet}}} \right) - (\mu + \gamma) \cdot L \quad (\text{A.2})$$

$$\frac{dA}{dt} = \alpha_{lux} \cdot \left( \delta_{lux} + \frac{1 - \delta_{lux}}{1 + \left( \frac{K_{lux}}{H} \right)^{\eta_{lux}}} \right) - (\mu + \gamma) \cdot A \quad (\text{A.3})$$

$$\frac{dH}{dt} = \frac{\beta_L}{1 + \left( \frac{\kappa_L}{L} \right)^{n_L}} \cdot X - \left( \frac{\beta_A}{1 + \left( \frac{\kappa_A}{A} \right)^{n_A}} \cdot X + D + \gamma_H \right) \cdot H \quad (\text{A.4})$$

with initial conditions:  $L(t_0) = \frac{\delta_{tet} \cdot \alpha_{tet}}{\mu + \gamma}$ ,  $A(t_0) = \frac{\delta_{lux} \cdot \alpha_{lux}}{\mu + \gamma}$ ,  $H(t_0) = 0$  and  $X(t_0) = X_0$ , set according to the initial *OD* measurement.

The  $L$  and  $A$  species represent the intracellular levels of the LuxI and AiiA enzymes, respectively, expressed as arbitrary units per cell ( $AU$ ), defined in section A.3. Transcription process was not explicitly modeled. Inducer-dependent synthesis rate dynamics was assumed to be negligible compared to the other described dynamic processes; this assumption was confirmed in this work (section A.4). Equations for



## A. Synthetic close-loop controller

---

LuxR dynamics were not included, since it is constitutively produced at constant level [133].  $X$  represents cell density, expressed as  $OD$  (see section A.3),  $H$  represents the intracellular and extracellular concentration of HSL (assumed to be identical), expressed as  $nM$ , and  $T$  represents ATc, expressed as  $ng/ml$ . The model includes 18 parameters, summarized in A.1, 16 of which were estimated in this work using data coming from several ad-hoc experiments on individual parts (see section A.3 and A.3), and the other 2 were fixed to known values from literature. Hill equations were used to describe activation transfer functions of promoters and the downstream protein synthesis:  $\alpha$  represent the maximum protein synthesis rate per cell at full induction,  $K$  represents the inducer value at which the synthesis rate is 50% of the maximum one,  $\eta$  is the Hill coefficient and  $\alpha \cdot \delta$  represents the basic synthesis rate in the off state. The activity of HSL-modifying enzymes is described analogously, via the  $\beta$ ,  $\kappa$  and  $n$  parameters, without considering basic activity. Cells were assumed to grow in exponential phase at rate  $\mu$  and diluted with rate  $D$  in turbidostat: when  $D = \mu$ , a constant cell density is maintained. Enzymes were assumed to be degraded with first-order kinetics at the same rate,  $\gamma$ , due to the LVA degradation tag (used to engineer fast degradation of the tagged proteins) present in both LuxI and AiiA, and were also assumed to be diluted at rate  $\mu$  due to cell division. In addition to enzymatic contributions,  $H$  was assumed to be washed at rate  $D$  and to spontaneously degrade at rate  $\gamma_H$ .

### Model analysis

A preliminary analysis was carried out to evaluate the steady-state in close-loop (Eqs.A.1,A.2,A.3,A.4) and open-loop (Eqs.A.1,A.2,A.3,A.4 with  $\alpha_{lux} = 0$ ) configurations, and to find which parameters had a considerable impact. For amenability reasons, mathematical analysis was carried out assuming, without any loss of generality, that all the Hill



## A.2. Close-loop circuit design

**Table A.1: Model parameters.** Parameter values used for preliminary simulations (Preliminary value) are provided with the references used to retrieve plausible values. Parameter values estimated in this work (Estimated value) are provided with the coefficient of variation (CV) of the estimate in brackets. <sup>a</sup>From previous experiments in our laboratory, in some cases in different strain, medium or vector backbone (data not shown). <sup>b</sup>Estimated in each turbidostat experiment as piecewise constant time course, as described in section A.3. <sup>c</sup>According to previous experiments (<http://2011.igem.org/Team:UNIPV-Pavia/Parts/Characterized>). <sup>d</sup>Estimated in one-plasmid context; parameter values estimated in co-transformed context, and used in the final model simulations, are:  $\alpha_{lux} = 2.92$  (4%),  $\delta_{lux} = 0.027$  (6%),  $K_{lux} = 474.1$  (12%),  $\eta_{lux} = 0.98$  (8%). <sup>e</sup>Fixed. <sup>f</sup>No literature value available:  $\kappa_L$  was arbitrarily set to obtain maximum HSL biosynthesis rate for 100ng/ml of ATc, and  $\kappa_A$  was selected considering that the steady-state level range of AiiA is 20 – 60AU, spanning a wide range of  $\kappa_A$  values (0.01 to 500AU).

Symbol	Description	Units	Preliminary value	Estimated value
$\mu$	Cell growth rate	$min^{-1}$	0.01 <sup>a</sup>	0.0126
$D$	Cell dilution rate	$min^{-1}$	0.01	Variable <sup>b</sup>
$\alpha_{tet}$	Maximum LuxI synthesis rate per cell regulated by $P_{LtetO1}$	$AU min^{-1}$	0.8 <sup>c</sup>	0.78 (5%)
$\delta_{tet}$	Relative basic activity of $P_{LtetO1}$	–	0.03 <sup>c</sup>	0.022 (6%)
$K_{tet}$	Concentration of aTc corresponding to half-maximum induction value of $P_{LtetO1}$	$ng/ml$	9 <sup>c</sup>	3.15 (9%)
$\eta_{tet}$	Hill coefficient of PLtetO1 activation function	–	4 <sup>c</sup>	3.99 (11%)
$\alpha_{lux}$	Maximum AiiA synthesis rate per cell regulated by $P_{lux}$	$AU min^{-1}$	11.2 [13]	4.99 (4%) <sup>d</sup>
$\delta_{lux}$	Relative basic activity of $P_{lux}$	–	0.014 [13]	0.016 (6%) <sup>d</sup>
$K_{lux}$	Concentration of HSL corresponding to half-maximum induction value of $P_{lux}$	$nM$	714.2 [13]	194.01 (17%) <sup>d</sup>
$\eta_{lux}$	Hill coefficient of $P_{lux}$ activation function	–	0.85 [13]	1.01 (6%) <sup>d</sup>
$\gamma$	LuxI and AiiA degradation rate	$min^{-1}$	0.0173 [79]	0.0173 <sup>e</sup>
$\beta_L$	Maximum LuxI enzymatic activity	$nM OD^{-1} min^{-1}$	12 [92]	6.83 (4%)
$\kappa_L$	Concentration of LuxI corresponding to half-maximum enzymatic activity	$AU$	1 <sup>f</sup>	6.74 (14%)
$n_L$	Hill coefficient of LuxI enzymatic activity function	–	1 [133]	3.51 (7%)
$\beta_A$	Maximum AiiA enzymatic activity	$OD^{-1} min^{-1}$	0.1 <sup>a</sup>	0.33 (9%)
$\kappa_A$	Concentration of AiiA corresponding to half-maximum enzymatic activity	$AU$	0.01, 50, 500 <sup>f</sup>	18.7 (> 100%)
$n_A$	Hill coefficient of AiiA enzymatic activity function	–	1 [133]	9.7 (> 100%)
$\gamma_H$	Spontaneous HSL degradation rate in M9pH6	$min^{-1}$	0.0002 [88]	0.0002 <sup>e</sup>

## A. Synthetic close-loop controller

---

functions of enzyme activities and AiiA synthesis rate have an  $n$  or  $\eta$  value of 1, their  $K$  or  $\kappa$  value is much higher than the specific inducer or enzyme level (thereby working in the linear region of the functions) and by neglecting the basic activity of the  $P_{lux}$  promoter ( $\delta_{lux} = 0$ ). Finally, the LuxI synthesis rate was assumed to be maximal ( $\alpha_{tet}$ ) under these hypotheses. HSL reaches a unique steady-state ( $\bar{H}$ ) in both configurations (open- and close-loop), as reported in Eqs.A.5,A.6.

$$H_{open}^- = \frac{Z \cdot X}{\mu} \quad (\text{A.5})$$

$$H_{close}^- = \frac{-\mu + \sqrt{\mu^2 + 4 \cdot X^2 \cdot Y \cdot Z}}{2 \cdot X \cdot Y} \quad (\text{A.6})$$

where  $Y = \frac{\beta_A \cdot \alpha_{lux}}{K_{lux} \cdot \kappa_A \cdot (\mu + \gamma)}$  and  $Z = \frac{\beta_L \cdot \alpha_{tet}}{\kappa_L \cdot (\mu + \gamma)}$ . In the open-loop configuration,  $\bar{H}$  is a linear function of cell density, directly proportional to LuxI expression and enzymatic activity (Eq. A.5). On the other hand, the circuit leads to an asymptotically saturating  $X$ -dependent trend (Equation A.6), which is null for  $X = 0$  (Equation A.8) and reaches a steady-state concentration that is independent from  $X$  for high cell density values (Equation A.8). By comparing  $\bar{H}_{open}$  and  $\bar{H}_{close}$  expressions, it is possible to verify that  $\bar{H}_{open} > \bar{H}_{close}$  for any  $X > 0$  (Equation A.9).

$$\lim_{X \rightarrow 0} \frac{-\mu + \sqrt{\mu^2 + 4 \cdot X^2 \cdot Y \cdot Z}}{2 \cdot X \cdot Y} = 0 \quad (\text{A.7})$$

$$\lim_{X \rightarrow \infty} \frac{-\mu + \sqrt{\mu^2 + 4 \cdot X^2 \cdot Y \cdot Z}}{2 \cdot X \cdot Y} = \frac{\sqrt{Z}}{\sqrt{Y}} \quad (\text{A.8})$$

$$\mu^4 + 4 \cdot \mu^2 \cdot X^2 \cdot U < \mu^4 + 4 \cdot \mu^2 \cdot X^2 \cdot U + 4 \cdot X^4 \cdot U^2 \quad (\text{A.9})$$



## A.2. Close-loop circuit design

---

where  $U = Y \cdot Z$ . This behavior reflects the robustness of close-loop configuration, in which the output does not change upon variations of cell density, for sufficiently high values of  $X$  with saturation trend determined by the value of the other parameters. Moreover, as expected, the steady-state value of HSL increases for increasing values of  $Z$  (linked to LuxI enzymatic activity and expression strength) and decreases for increasing values of  $Y$  (linked to AiiA enzymatic activity and expression strength).

By using biologically plausible values for all the model parameters (see Table A.1), preliminary simulations via Eqs. A.1,A.2,A.3,A.4 showed the cell density-dependent steady-state trend of HSL in response to a saturating amount ( $100ng/ml$ ) of ATc (see Figure A.2(a)). As anticipated above, the  $X - \bar{H}$  characteristic saturates at different densities depending on the parameter values (shown in FigureA.2(a) as an example by varying  $\kappa_A$ ), as expected from Equation A.8, thereby affecting the close-loop configuration robustness over density variation. The target cell density set via turbidostat affects the close-loop circuit dynamics: while the open-loop reaches a steady-state with identical dynamics for every density, only dependent on HSL washout (see Figure A.2(b)), in close-loop circuit low densities correspond to higher rise time, comparable to the one of open-loop in the same conditions. Finally, to investigate the properties of the designed system, disturbance rejection capability was studied in response to an HSL pulse (i.e., injection at a given time point). After pulse application, the close-loop circuit enabled faster return to the steady-state than the open-loop configuration (see Figure A.2(c)) for different pulse amplitudes (data not shown). This dynamics is density-dependent, with faster response at higher densities (see Figure A.2(d)).

## A. Synthetic close-loop controller

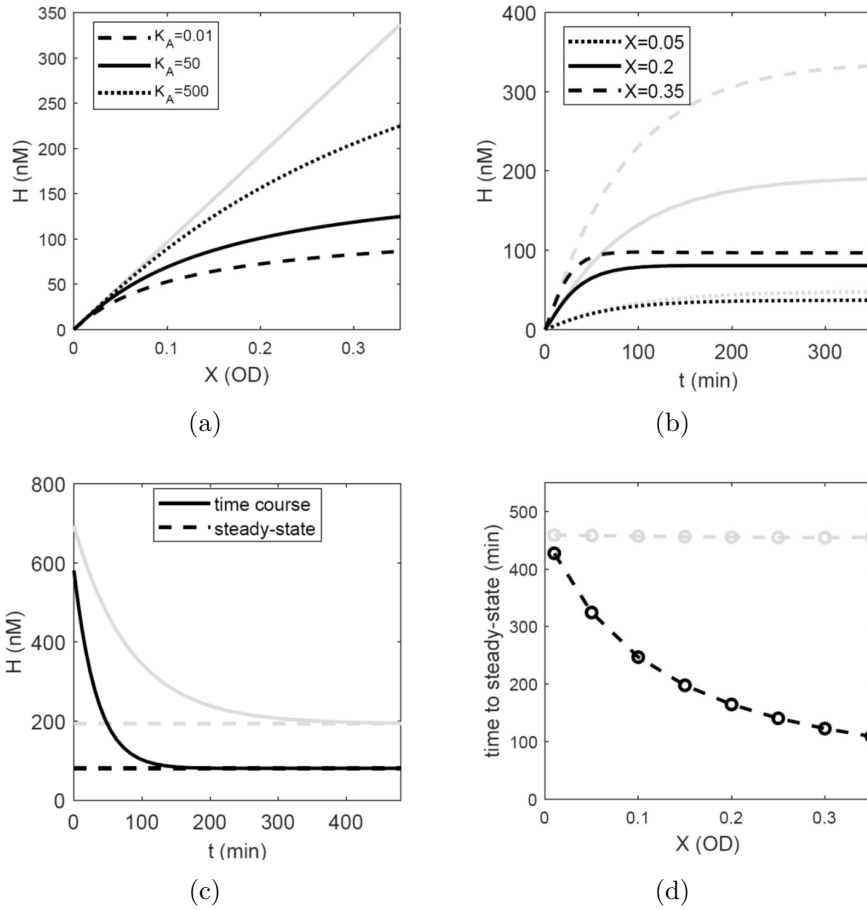


Figure A.2: **Preliminary simulations.** **a.** Simulated HSL ( $H$ ) at steady-state as a function of cell density ( $X$ ) maintained in the turbidostat, for the open- and close-loop (for different values of  $K_A$ ) systems. **b.** Simulated time courses of HSL in open- and close-loop circuits for different  $X$  values. **c.** Simulated time course of HSL for  $X = 0.2$  in response to an injection of HSL  $500\text{ nM}$  at  $t = 0$ , acting as a disturbance on the system output; dashed lines represent the steady-state levels and solid lines represent the response to the HSL pulse; **d.** the time to reach the steady-state (within a  $2\text{-nM}$  tolerance) after the application of the disturbance impulse (panel (c)) as a function of the maintained cell density (dots represent the simulated conditions and dashed lines are interpolation lines). In all the panels, grey and black curves indicate open- and close-loop circuits, respectively.

## A.3 Materials and Methods

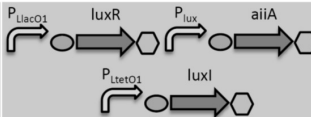




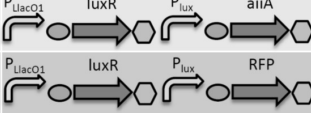
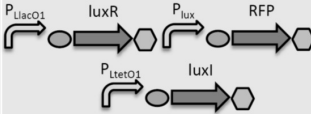
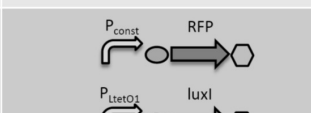
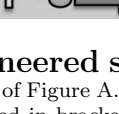
### Strains, plasmids and growth media

The MG1655-Z1 strain [146] was used as host in all the experiments shown in this work. It was transformed via heat shock to obtain recombinant strains. All the tested strains are listed in Figure A.3, explicitly describing the incorporated plasmids. The M9 medium supplemented with casamino acids, thiamine and glycerol was used in all the quantitative tests. Differently from its standard formulation [13], the pH of medium was adjusted to 6.0 to limit HSL degradation, occurring at higher pH values [88]. This modified medium will be referred to as M9pH6. DNA assembly was carried out as previously reported, via BioBrick<sup>TM</sup> Standard Assembly and standard molecular biology procedures using available parts in the MIT Registry Distributions 2008-2011 [147]. Plasmids are available in the Registry of Standard Biological Parts with codes indicated in Figure A.3. Details on their construction are present in the individual online Registry pages of the parts and in A.6.

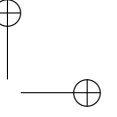
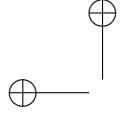
### Promoter characterization

Promoters were characterized using fluorescent reporter protein-based assays. Test constructs included the promoter under study and a red fluorescent protein (RFP) expression system assembled downstream. A characterization experiment was carried out for each used promoter-RBS-plasmid set, analyzing at least three biological replicates. Steady-state dose-response experiments and data analysis were carried out as previously described [79, 84] to obtain cell growth rate and RFP synthesis rate per cell ( $S_{cell}$ , in  $AU \text{ min}^{-1}$ ) in exponential phase, as a function of inducer concentration. The ConstRFP-MC (with the J107029 construct) strain was used as reference for  $S_{cell}$  nor-

## A. Synthetic close-loop controller

Strain name	Circuit description	BioBrick code(s)	Parameters estimated
CL1		J107072(4C5) co-transformed with K516210(3K3)	None
OL1		K516210(4C5)	$\beta_L, \kappa_L, \eta_L$
TetRFP-LC		K516230(4C5)	$\alpha_{tet}, \delta_{tet}, \kappa_{tet}, \eta_{tet}$
ConstRFP-MC		J1070[29, 61, 62, 63, 64, 31, 12](3K3)	$\beta_{Ar}, \kappa_{Ar}, \eta_{Ar}$
ConstAct-MC		J1070[65-71](3K3)	
SensRegAct-MC		J107072(3K3)	None
SensRegRFP-MC		J107053(3K3)	$\alpha_{lux}, \delta_{lux}, \kappa_{lux}, \eta_{lux}$
OL1-SensRegRFP		J107053(4C5) co-transformed with K516210(3K3)	
OL1-ConstRFP		J1070[29, 61, 62, 63, 64, 31, 12](3K3) co-transformed with J107072(4C5)	None

**Figure A.3: List of engineered strains tested in this study.** Circuit description used the same symbols of Figure A.1. BioBrick codes are relative to the DNA inserts, while vector backbones are reported in brackets. The **BBa** and **pSB** prefixes are omitted for inserts and vector backbones, respectively. Low- (pSB4C5) and medium- (pSB3K3) copy vectors were used.  $P_{const}$  represents a constitutive promoter: the (BBa\_)J23101, J23105, J23106, J23110, J23116, J23118, I14032 promoters were used; for this reason, ConstRFP-MC, ConstAct-MC and OL1-ConstRFP indicate a set of strains instead of a single one. Strains are grouped by model parameters estimated using these circuits (some of them were only used as test set).



### A.3. Materials and Methods

---

malization in all the experiments, to provide highly reproducible measurement units among the experiments and also in different studies. Detailed measurement protocols and analysis procedures are available in A.6. Assuming highly stable RFP protein (i.e.,  $\mu$  is much faster than its degradation rate), the following ODE model describes the dynamics of RFP production in exponential phase (Eqs.A.9-A.10).

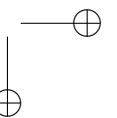
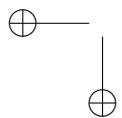
$$\frac{dI}{dt} = \alpha \cdot \left( \delta + \frac{1 - \delta}{1 + \left(\frac{K}{q}\right)^\eta} \right) - (m + \mu) \cdot I \quad (\text{A.10})$$

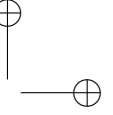
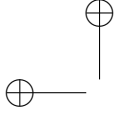
$$\frac{dR}{dt} = m \cdot I - \mu \cdot R \quad (\text{A.11})$$

where  $q$  is the inducer (ATc or HSL) concentration,  $I$  and  $R$  are the intracellular levels of immature (non-fluorescent) and mature (fluorescent) protein (in  $AU$ ),  $m$  is the fluorochrome first-order maturation rate ( $0.0167\text{min}^{-1}$  [43]), and the other symbols have the same meaning as in section A.2. Assuming the steady-state of intracellular species and defining, as previously described [29, 93], the average RFP synthesis rate per cell in exponential phase as  $\overline{S_{cell}} = m \cdot I$  (i.e., the synthesis term of Equation A.11), its expression can be written as (Equation A.12):

$$\overline{S_{cell}} = \frac{m}{m + \mu} \cdot \alpha \cdot \left( \delta + \frac{1 - \delta}{1 + \left(\frac{K}{q}\right)^\eta} \right) \quad (\text{A.12})$$

The  $\alpha$ ,  $\delta$ ,  $K$  and  $\eta$  parameters were estimated by fitting experimental ( $\overline{S_{cell}}$ ) data with Equation A.12. Such parameters describe the synthesis rate of the protein encoded by the gene downstream of the promoter, resulting from both transcription and translation processes.





## A. Synthetic close-loop controller

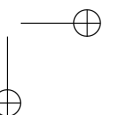
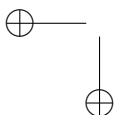
---

Assuming that: i) the Hill function is dependent on transcriptional regulation, ii) translation accounts only for a scale factor, and iii) biological parts are modular components, i.e., their behavior does not change upon interconnection in a different circuit, then the same Hill function is proportional to the synthesis rate of a different protein encoded by a gene downstream of the same promoter. According to these assumptions, RFP levels (in  $AU$ ) can be used to approximate the intracellular level of any protein produced via the same expression system.

Constitutive promoters were characterized analogously, with the exception that only the  $\alpha$  parameter was measured, while  $\delta$ ,  $K$  and  $\eta$  were fixed to zero.

### Enzyme characterization

HSL production (for LuxI) and degradation (for AiiA) assays were used to estimate enzyme activities. The two genes were assembled downstream of quantitatively characterized constitutive or inducible promoters, obtaining a set of test constructs with the same RBS-plasmid pair as the one used in the final system for luxI and aiiA. Colonies of recombinant strains bearing such constructs (see Figure A.2) were grown as for promoter characterization experiments to obtain a saturated culture; then it was 100-fold diluted into a 5 – ml culture in 15 – ml tube, and grown under the same conditions as above. When required, after 1 – h growth, enzyme expression was induced via different amounts of ATc. HSL measurements started after 1 additional hour ( $t = 0$ ), in which enzyme level could reach an intracellular steady-state. If not differently indicated, 100nM of HSL were added in AiiA assays at  $t = 0$  and HSL was measured via whole-cell biosensor at different time points, as reported in [88], with the exception that biosensor circuit was assembled into different plasmids, with the same antibiotic resistance(s) as the measured culture.





### A.3. Materials and Methods

The following models were used to describe LuxI- and AiiA-dependent HSL dynamics (Equation A.13 and Equation A.14, respectively):

$$\frac{dH}{dt} = \hat{k}_L \cdot X - \gamma_H \cdot H, \quad H(t_0) = H_0 \quad (\text{A.13})$$

$$\frac{dH}{dt} = -(\hat{k}_A \cdot X + \gamma_H) \cdot H, \quad H(t_0) = H_0 \quad (\text{A.14})$$

while cell growth was described by Equation A.1 with  $D = 0$ , since experiments were carried out in batch mode. The  $\hat{k}_L$  and  $\hat{k}_A$  parameters represent the per-cell HSL biosynthesis and degradation rates in the tested conditions. Cell density and HSL were simultaneously fitted with the equations above, estimating  $\hat{k}_L$  or  $\hat{k}_A$ ,  $\mu$ ,  $X_0$  and  $H_0$  (see section A.6 for examples of fitted data). The resulting  $\hat{k}_L$  and  $\hat{k}_A$  values were used to estimate the enzyme level-activity relations by fitting Eqs.A.15-A.16:

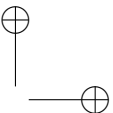
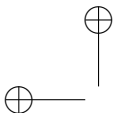
$$\hat{k}_L = \frac{\beta_L}{1 + \left(\frac{\kappa_L}{\bar{L}}\right)^{\eta_L}} \quad (\text{A.15})$$

$$\hat{k}_A = \frac{\beta_A}{1 + \left(\frac{\kappa_A}{\bar{A}}\right)^{\eta_A}} \quad (\text{A.16})$$

where  $\bar{L}$  and  $\bar{A}$  are the predicted steady-state levels of LuxI and AiiA when their expression is driven by  $P_{LtetO1}$  in the OL1 and ConstAct-MC for LuxI and AiiA, respectively.

### Simulation and fitting procedures

All the simulations and fittings were performed via MathWorks Matlab R2007b. The *ode23s* solver was used to compute ODE solutions. The weighted least squares algorithm, implemented via the



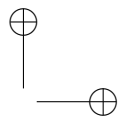
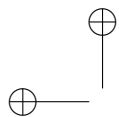
## A. Synthetic close-loop controller

---

*lsqnonlin* routine, was used for parameter estimation. Residual weights were set as in [79].

### Turbidostat experiments

Recombinant cells were inoculated and grown overnight as for promoter characterization experiments. Cells were then 100-fold diluted into a 20 to 40ml volume ( $V$ ) of medium in a 0.4-liter vessel, incubated with a Minifor (Lambda) bioreactor at high agitation speed (10 units). The reactor, routinely sterilized according to manufacturer instructions, was equipped with temperature/pH and oxygen probes. Temperature was maintained at 37°C. Growth was continued until the culture reached a target  $OD$ , then the continuous mode started: a PRECIFLOW (Lambda) peristaltic pump was activated to provide sterile medium from a reservoir to the culture, with input flux ( $F$ ) set to a suitable value for diluting cells at rate equal to their growth:  $F = D \cdot V$  and  $D = \mu$ . An overflow canula, connected with another peristaltic pump, maintained a constant volume in the vessel by eliminating the excess culture. The canula was placed at proper level to maintain a target culture volume, calibrated before each experiment. Samples were withdrawn to measure  $OD$  and HSL as described in section A.3. Unless differently stated, induction was carried out at the beginning of turbidostat mode by adding 100ng/ml of ATc into both vessel and reservoir. The choice of this ATc concentration is discussed below (section A.4). In the described setup, small changes of canula position could result in relevant  $V$  changes. For this reason, the adjustment of  $F$  (initially set by considering the initial  $V$ ) was carried out during experiments, according to  $X$  time course variations and assuming a constant  $\mu$ . Based on Equation A.1, the  $D$  parameter was eventually estimated as piecewise constant function in each time



## A.4. Results

---

interval between the  $i$ -th and  $(i + 1)$ -th samplings (Equation A.17).

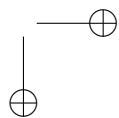
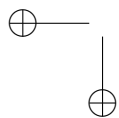
$$D(t_i) = \mu - \frac{X(t_{i+1}) - X(t_i)}{(t_{i+1} - t_i) \cdot X(t_i)} \quad (\text{A.17})$$

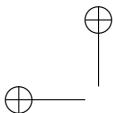
Disturbance rejection experiments were carried out analogously, except that cells were 50-fold diluted instead of 100-fold, and ATc was immediately added. HSL (i.e., the disturbance) was added to the culture during turbidostat mode.

## A.4 Results

### Individual parts characterization

As first step of a bottom-up design process, promoters and enzymes were individually characterized in ad-hoc constructed biological measurement systems, i.e., assemblies between promoter and reporter gene (for promoter characterization) and between previously characterized promoter and enzyme-encoding gene (for enzyme characterization). The fitted curves of steady-state inducer-activity function of the promoters included in the close-loop circuit (see Figures A.4(a)-A.4(b)) and the enzyme level-activity characteristics (see Figures A.4(c)-A.4(d)) are reported in Figure A.4. On the other hand, data of the promoters used as inputs for AiiA characterization (see Figure A.2) are reported in A.6. All the experimental data were well-fitted by the used models. Almost all parameters were estimated with low uncertainty and with values consistent with the ones available in literature, measured in different strains or conditions (see Table A.1). Promoters and enzymes activities can be successfully tuned from nearly null values to maximal ones by regulating the cognate inducer concentration and enzyme expression, respectively, thereby demonstrating that the input signals are all suitable to exhaustively





## A. Synthetic close-loop controller

---

characterize the parts. It is worth noting that, differently from the initial assumptions, the  $n$  values of the enzyme level-activity relations (Eqs.15-16) are much higher than 1 (although  $n_A$  estimate is affected by high uncertainty), thereby resulting in highly nonlinear, switch-like, activation functions (see Table A.1).

### Interconnected subcircuits characterization

Given a strain and a growth condition, it is known that parameters related with biological parts function can change upon interconnections or simply reusing the parts in different circuits [23, 79, 148]. This situation is often referred to as context-dependent variability, and may hamper the predictable and modular composition of synthetic circuits. For this reason, before proceeding with a predictability analysis using the values estimated in section A.3, the robustness of some crucial parameter estimates was tested in additional in vivo experiments. First, the HSL biosynthesis capability of LuxI was tested in the OL1-SensRegRFP strain, bearing two plasmids with HSL production module and biosensor with RFP, respectively. HSL production capability in batch and turbidostat experiments was consistent with OL1 model simulations (see Figure A.5(a)). These results demonstrated that the LuxI module function is maintained in a larger circuit similar to CL1. Second, the copy number (CN) of the medium-copy vector in co-transformed context was compared with the copy number maintained in the one-plasmid context. The CNs were found to be identical (see Figure A.5(b)). Taken together, the results of Figure A.5(a)-A.5(b) demonstrated that the two used vectors maintain a stable CN in all the tested conditions. Third, as in the first case, the  $P_{lux}$  activity was tested in a co-transformed context with the OL1-SensRegRFP strain, which includes a more complex circuit than SensRegRFP-MC, used in section A.3 to individually characterize  $P_{lux}$ . Results showed an about 2-fold lower  $P_{lux}$  activity in OL1-SensRegRFP than SensRegRFP-MC,

## A.4. Results

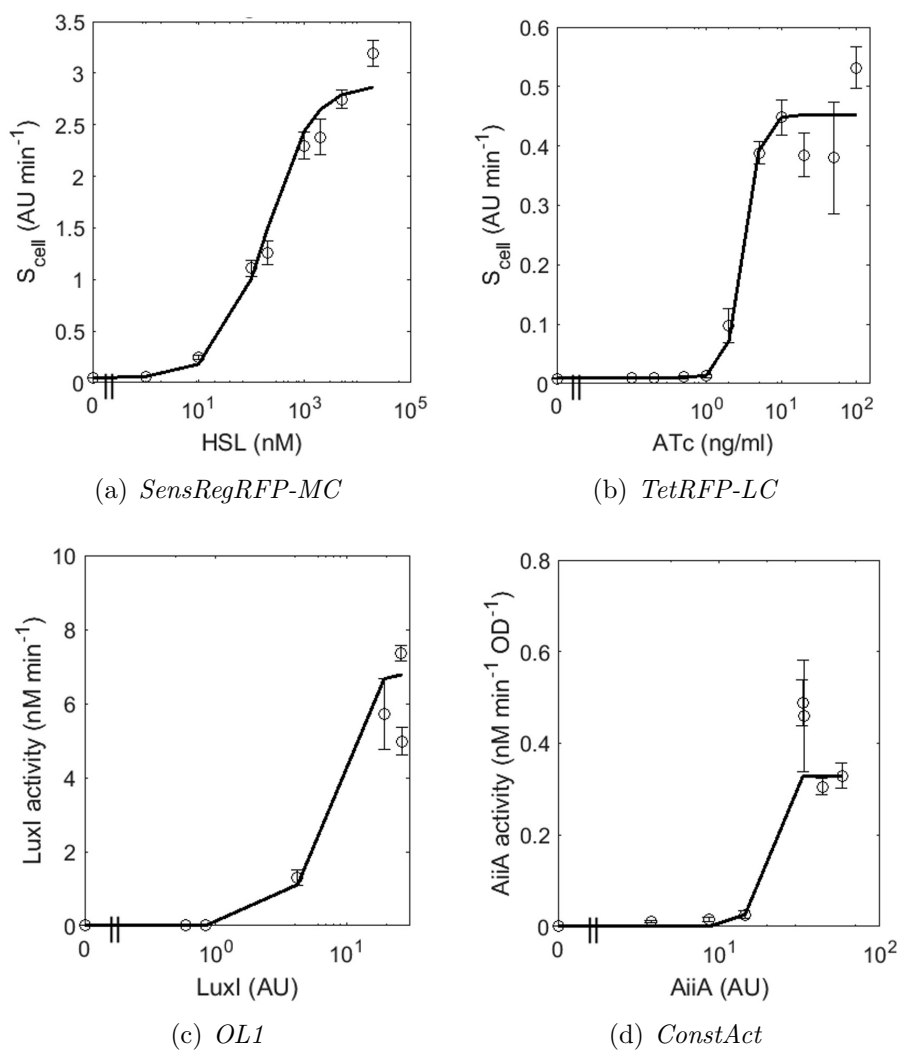
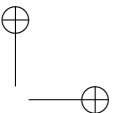
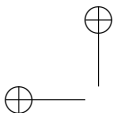


Figure A.4: **Fitting of steady-state transfer functions of the individual devices of close-loop circuit.** a.  $P_{lux}$  promoter. b.  $P_{LtetO1}$  promoter. c LuxI enzyme. d AiiA enzyme. In all the panels, circles represent experimental data, error bars represent standard errors of the mean and solid lines represent model fitting



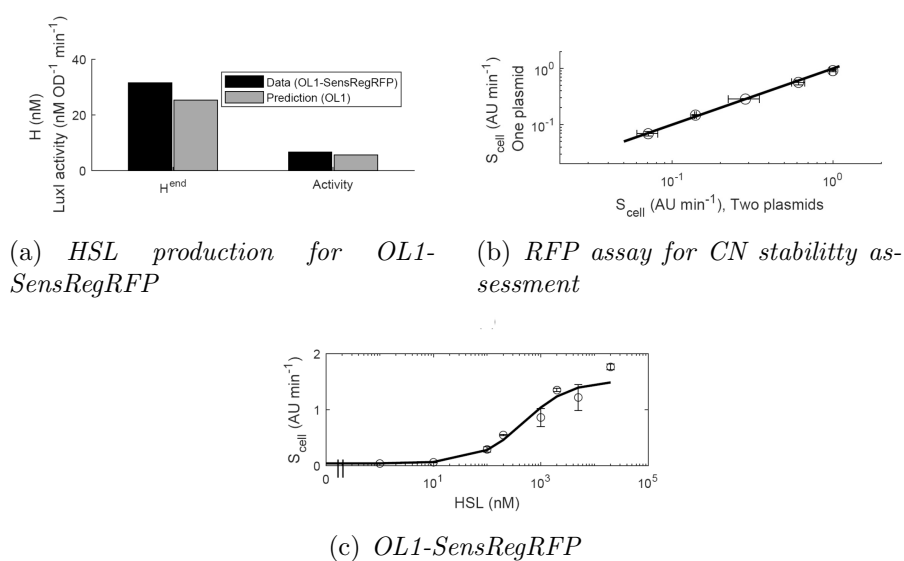
## A. Synthetic close-loop controller

---

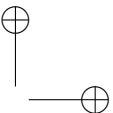
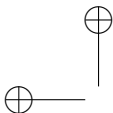
and a higher  $K_{lux}$  (see Table A.1 and Figure A.5(c)). Since the vector CNs were found to be unaffected in co-transformed context, such differences could be due to burden effects significantly affecting the resulting  $P_{lux}$  activation curve. By over-expressing LuxR via IPTG addition in both strains, the measured transfer function also confirmed the results above (see section A.6). The observed difference in parameter values between one- and two-plasmid contexts is consistent with context-dependent activity changes observed in other studies [13, 79, 84]. However, it highlights the need of characterizing individual components in more than one context, as similar as possible to the final one, to find robust parameter estimates to properly simulate the final system. To further confirm that the  $P_{lux}$  activity difference observed between the two tested contexts also persists using *aiiA* as downstream gene (i.e., the one used in the final system), two similar circuits were tested in batch mode: SensRegAct-MC and CL1 (in absence of ATc induction), both including a  $P_{lux}$ -regulated *AiiA* expression in one- and two-plasmid context, respectively, as before. Such strains are expected to implement HSL sensing, regulation and degradation (see Figure A.2 and upper part of Figure reffig:CL1a) without HSL biosynthesis. The simulated HSL degradation time courses, obtained using the specific  $P_{lux}$  parameters in one- and two-plasmid contexts (see Table A.1), were consistent with the measured data, depicting a systematically lower HSL degradation for CL1, compatible with the weaker  $P_{lux}$  activity in this context (see section A.6). For the described reasons, the  $\alpha_{lux}$ ,  $\delta_{lux}$ ,  $K_{lux}$  and  $\eta_{lux}$  values in co-transformed context were used as Hill function parameters of  $P_{lux}$ .

Finally, RFP synthesis rate time course in OL1-SensRegRFP upon ATc induction was also used to indirectly evaluate if HSL dynamics, including biosynthesis and self-induction of  $P_{lux}$  in the same culture, could be captured by the model. Results showed that measured fluorescence values were in reasonable agreement with the simulated profiles (see section A.6).

## A.4. Results



**Figure A.5: Analysis of interconnected subcircuits.** **a.** Characterization of HSL biosynthesis of *OL1-SensRegRFP*:  $H^{end}$  (i.e., HSL concentration at the end of a turbidostat experiment) and enzymatic activity in batch experiment were measured and compared with the prediction of the model, which was trained on *OL1*. **b.** Fluorescence assays on the *ConstRFP-MC* library (one-plasmid context) and on the *OL1-ConstRFP* library (two-plasmid context); similar RFP values between one- and two-plasmid contexts mean similar *CN*; circles represent data points, error bars represent standard errors of the mean and solid line represents the bisector. **c** Fitting of steady-state transfer function of *Plux* promoter in a two-plasmid context (*OL1-SensRegRFP*); circles represent experimental data, error bars represent standard errors of the mean and solid line represents model fitting.



## A. Synthetic close-loop controller

---

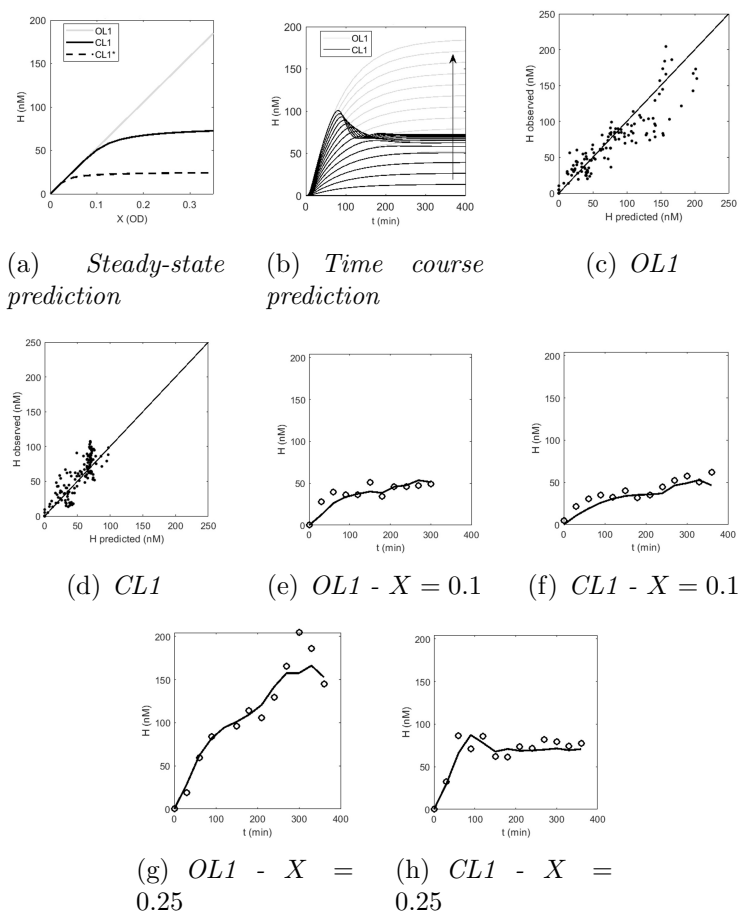
### Analysis of the final system in turbidostat

Simulations of the fully identified model of Eqs.A.1-A.4 showed that CL1 has a predicted cell density-HSL profile varying less than 1.5-fold for  $0.1 < X < 0.35$ , while OL1 showed a 3.5-fold change (see Figure A.6(a)). By replacing the  $P_{lux}$ -related parameter values with the values estimated previously in one-plasmid context (see section A.4), simulations showed a significantly different trend, with 3-fold lower steady-state level (see Figure A.6(a)), confirming the importance of reliable estimation of such parameters. The HSL steady-state and time course (Figs.A.6(a)-A.6(b)) confirm the conclusions drawn in section A.2 about cell density-dependent features of the close-loop circuit. In addition, the time-course simulations showed that damped oscillations arose with the estimated parameter set for  $X > 0.1$  (see Figure A.6(b)). Note that OL1 and CL1 start showing a different behavior for  $X > 0.1$  (see Figure reffig:CL6a). A univariate sensitivity analysis, reported in A.6 for small variations of the parameters involved in feedback control, highlights that the switch point of AiiA enzymatic activity has major impact on HSL output, that the AiiA maximum activity is never reached at steady-state, and that such parameter plays a minor role in the HSL steady-state regulation.

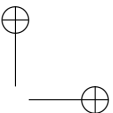
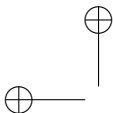
As the final step, the constructed system was tested and the data were used as validation set to assess the predictability of the identified model. A set of 12 turbidostat experiments, with different cell densities, were carried out for CL1, while OL1 was used as control in other 11 experiments. Average cell densities in the range  $0.05 - 0.35$  were maintained (see section A.6). Data and predictions of all the experiments are available in A.6. A global overview of the prediction capability of the model is shown in Figure reffig:CL6c-A.6(d), in which the predicted and measured HSL values are compared for the CL1 and OL1 circuit, respectively. In particular, predictions are



## A.4. Results



**Figure A.6: Prediction of open-loop and close-loop circuits behavior.** **a.** Simulation of steady-state HSL; the curves of OL1 and CL1 simulated using the parameters of SensRegRFP-MC values (herein indicated as CL1\*) are also shown. **b.** Simulation of OL1 and CL1 dynamics; the arrow indicates increasing cell density  $X$ , from 0.025 to 0.35 with step 0.025. **c-d** Overall model prediction performance for OL1 **c** and CL1 **d** in turbidostat experiments; dots represent experimental data and solid line represents the identity line. **e-h** HSL data (circles) and model predictions (solid line) for different experiments of OL1 and CL1 in turbidostat at different cell densities  $X$ ; the typical values of  $X$  are reported in the titles.

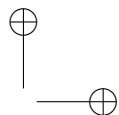
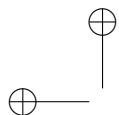


## A. Synthetic close-loop controller

---

characterized by high correlation coefficients (0.87 for CL1, 0.91 for OL1). Two representative experiments for each circuit (Experiment #3 and #8 for OL1; Experiment #4 and #10 for CL1, see Section A.6 for experiment numbering and description) are reported in Figures A.6(e)-A.6(h). The time course is well predicted by the mathematical model in both CL1 and OL1. As expected, CL1 and OL1 show comparable HSL control behavior at low cell density (about 0.1), in which HSL reaches a level of about  $50nM$  (see Figure A.6(e)-A.6(f)). At higher cell density (0.25), OL1 showed a final HSL concentration of  $153nM$  (see Figure A.6(g)), whereas CL1 showed robust control of HSL level, with a superelongation also predicted by the model not exceeding  $87nM$ , and a steady-state value of  $72nM$  (see Figure A.6(h)). Cell density-dependent fluctuations of HSL were observed in many experiments, due to target molecule washing at non-constant rates ( $D \neq \mu$ ) during  $D$  adjustment in turbidostat mode. In addition, oscillations not correlated with  $OD$  fluctuations were observed in CL1 experiments at high cell densities (e.g., see Figure A.6(h)) as predicted by the model.

Finally, the capability of the close-loop circuit to reject disturbances on HSL output was tested in additional turbidostat experiments at very similar cell densities ( $X = 0.23 \pm 0.03$ , see Figure A.7(a)). Disturbances were applied as HSL injections, and the experimental vs predicted time courses were compared for both CL1 and OL1 (see Figure A.7(b)). Model predictions were consistent with experimental measurements, thereby confirming the robust disturbance rejection properties of the close-loop design, as predicted by model simulations. It is worth noting that the implemented close-loop circuit has two means of tuning HSL level: ATc and IPTG can be used to regulate LuxI and LuxR synthesis rates, respectively, thereby changing HSL production process and sensor transfer function. The turbidostat tests carried out in this work only included a saturating amount of ATc ( $100ng/ml$ ) without IPTG. This condition represents an optimal sit-



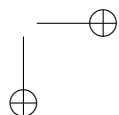
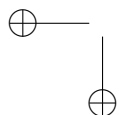
## A.5. Conclusion

---

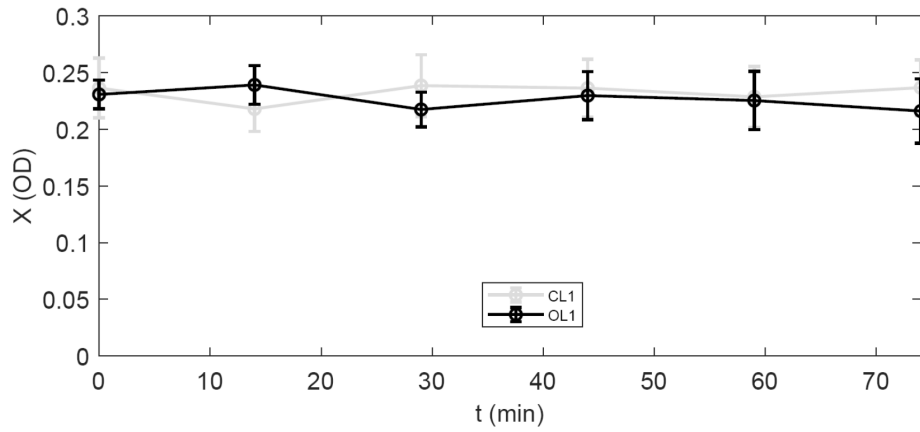
uation to yield high-quality experimental measurements of HSL, since LuxI synthesis rate is in a robust full-induction state, not affected by small ATc variations (see Figure A.4(b)), and HSL steady-state level is maximized (while lower ATc and higher IPTG concentrations are expected to give lower HSL levels). Circuit predictability, robustness against cell density variation and disturbance rejection features, on which this study is focused, have been successfully demonstrated in the tested condition, while a systematic analysis of CL1 with different inducer concentrations is beyond the scope of this work.

## A.5 Conclusion

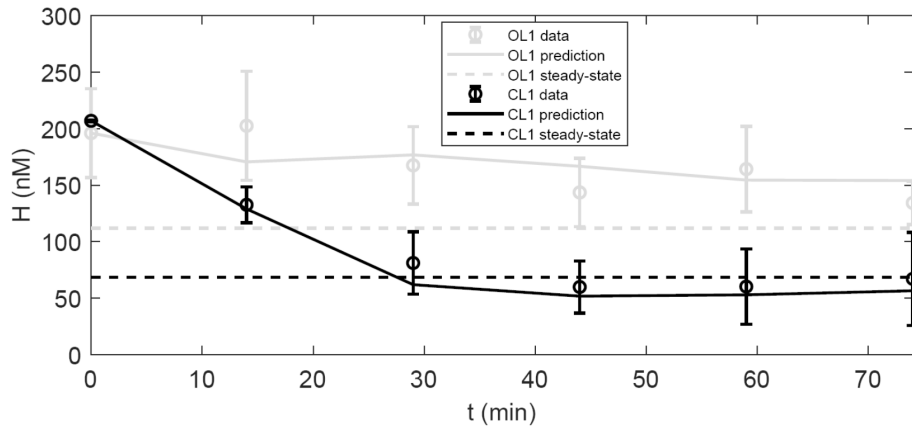
In this work, we showed the bottom-up design, from individual parts, of a close-loop control circuit (CL1). The comparison between simulations and in vivo data in different scenarios showed not only that the model-based bottom-up approach was successful in accurately predicting HSL output of CL1, but also that the designed close-loop scheme has the expected properties of robustness against cell density variation and disturbance rejection on the output. Nonetheless, a number of limitations, also faced in this work, still affect the execution of such task and are herein discussed. The composition of a predictable biological system is affected by context-dependent variability, as it was observed for the  $P_{lux}$ -based sensor/regulator module (i.e., the device showing major activity variation). Despite the other circuit modules did not show relevant activity variation when re-used in higher-order circuits, the accurate prediction of the final system behavior required the measurement of  $P_{lux}$  in more than one context, which is an important aspect for future bottom-up designed circuits. The cell load caused by heterologous gene networks was not explicitly modeled in this work; the extensive characterization of cell resource usage by individual parts may further improve their re-use in complex circuits,



## A. Synthetic close-loop controller

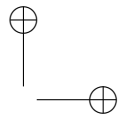
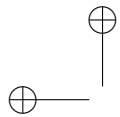


(a) Cellular concentration.



(b) HSL level.

**Figure A.7: Model prediction of disturbance rejection on HSL output.** **a** Measured  $X$  time course in CL1 and OL1. **b**. Data and prediction of HSL time course after HSL disturbance injection ( $t = 0$ ) when the system was at steady-state. Circles represent data points, dashed lines represent the predicted steady-state considering average  $X$  and  $D$  values, solid lines in panel **a**. represent linear interpolation and in panel **b**. represent model simulations. Error bars represent standard errors of the mean of two independent replicates.



## A.5. Conclusion

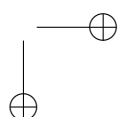
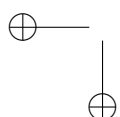
---

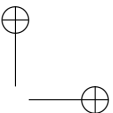
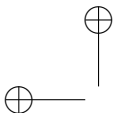
thereby avoiding multiple characterization steps. Apart from intrinsic function variation of individual parts, the design of close-loop circuits is also hampered by the significant differences between circuit modules in biology and control theory; in fact, despite preliminary efforts have been carried out to abstract sensor, regulator and actuator modules in living systems, their proper tuning, e.g., design of regulators with intended dynamic properties, still represents a challenge that will require further theoretical work. Taken together, the described limitations currently prevent the definition of detailed design guidelines for close-loop circuits from the bottom-up and their solution will represent a milestone in the fields of synthetic biology and control theory applied to biological engineering.

Finally, the feedback control scheme implemented in this work could be improved by engineering faster response dynamics and more energy-saving control strategies, as proposed in A.6.

Even if the implemented circuit relied on well-studied quorum sensing/quenching network elements, thereby enabling the study of a simplified model system, the demonstration of the bottom-up design of a robust control strategy for an extracellular molecule represents an important achievement in synthetic biology.

It is worth noting that different designed circuit may present different critical issues, e.g., non-modeled crosstalk among circuit components or excessive toxicity for the cell of a specific molecule. However, we believe that the experience gained in this study can be applied, with reasonable generalization, to decrease trial-and-error steps in close-loop systems design, by following the engineering-inspired strategy used in this work: i) define control theory blocks (process, sensor, regulator, actuator) using biological components; ii) characterize individual candidate parts, in proper biological measurement systems, in more than one context; iii) estimate the parameters of the mathematical models able to describe the function of each individual part; iv) use the parametrized models to predict the final interconnected





## A. Synthetic close-loop controller

---

system. The used strategy may benefit many industrial applications involving engineered cell factories implementing robust extracellular control of relevant molecules, such as fuels, drugs, cosmetics or key intermediates of the respective pathways.

## A.6 Supplementary Notes

### Supporting methods

#### Cloning

The *E. coli* TOP10 strain (Invitrogen) was used for the in vivo amplification of plasmids. TOP10 were transformed by heat-shock according to manufacturer's instructions and transformed strains were grown in L-broth (LB: sodium chloride 10g/l, tryptone 10g/l, yeast extract 5g/l) at 37°C. Long-term glycerol stocks, routinely stored at -80°C, were prepared by mixing 750µl of a saturated culture with 250µl of glycerol 80%. All the circuits used in this study were assembled from existing Registry parts according to the BioBrick<sup>TM</sup> Standard Assembly procedure and a number of standard molecular biology methods: plasmids were extracted from saturated 5 - ml cultures (grown in LB at 37°C, 220rpm) through the NucleoSpin Plasmid kit (Macherey-Nagel); DNA was digested as appropriate, with the EcoRI/XbaI/SpeI/PstI enzymes, and the fragments of interest were extracted from 1% agarose gel by NucleoSpin Extract II kit (Macherey-Nagel) before proceeding with ligation. As a result, each part used in this work is compliant to the BioBrick<sup>TM</sup> Standard. Consequently, every junction between assembled parts has the TACTAG sequence if the downstream part is a coding sequence, otherwise the sequence is TACTAGAG. All the DNA-modifying enzymes were purchased from Roche Diagnostics. The DNA of all the constructed parts was screened via diagnostic digest/electrophoresis, and was sequence-verified via the



## A.6. Supplementary Notes

---

BMR Genomics DNA analysis service (Padova, Italy). The vector backbones used in the studied circuits were pSB4C5 and pSB3K3 from the MIT Registry, enabling the maintenance of the assembled genetic circuits at low (3 to 7) and medium (15 to 40), respectively, with values in brackets depending on the used strain. Antibiotics were always used to maintain the incorporated plasmids, according to their selection marker: chloramphenicol ( $12.5\text{mg/l}$ ) and kanamycin ( $20\text{mg/l}$ ).

### Fluorescence and growth assays

For steady-state characterization of promoters, long-term glycerol stocks were streaked on LB agar plates supplemented with the proper antibiotic(s), to isolate single colonies, considered as biological replicates. Plates were incubated overnight at  $37^\circ\text{C}$ , then  $0.5\text{ ml}$  of selective M9pH6 (M9 salts - #M6030, Sigma Aldrich -  $11.28\text{g/l}$ , thiamine hydrochloride  $1\text{mM}$ ,  $\text{MgSO}_4$   $2\text{mM}$ ,  $\text{CaCl}_2$   $0.1\text{mM}$ , casamino acids  $0.2\%$ , glycerol  $0.4\%$ , pH adjusted to 6.0 via HCl addition) were inoculated with single colonies and incubated overnight in  $2 - \text{ml}$  tubes at  $37^\circ\text{C}$ ,  $220\text{rpm}$ . Cultures were 100-fold diluted in  $1\text{ml}$  of selective M9pH6 in  $15 - \text{ml}$  tubes, grown for  $1 - \text{h}$  and subsequently induced with a proper amount of ATc or HSL to reach the desired concentrations (when explicitly indicated, isopropyl- $\beta$ -D-1-thiogalactopyranoside - IPTG - was also added at the final concentration of  $500\mu\text{M}$ ). After  $1 - \text{h}$  growth in the same conditions as above,  $200\mu\text{l}$  were transferred into a 96-well microplate. Cultures were assayed via the Infinite F200 microplate reader (Tecan), programmed with the i-control (Tecan) software to perform a kinetic cycle as follows: linear shaking ( $3 - \text{mm}$  amplitude,  $15\text{s}$ ), wait ( $5\text{s}$ ), absorbance measurement ( $600\text{nm}$ ), red fluorescence measurement (excitation at  $535\text{nm}$ , emission at  $620\text{nm}$ , gain=50), repeat cycle every  $5\text{min}$ . In every microplate experiment,  $200\mu\text{l}$  of M9pH6 and a non-fluorescent MG1655-Z1 culture were included in triplicate to enable the estimation of absorbance and fluorescence backgrounds, re-

## A. Synthetic close-loop controller

---

spectively. Finally, the ConstRFP-MC strain with the J107029 insert, including a constitutive RFP expression cassette under the control of the J23101 constitutive promoter, herein named REF culture, was also included in triplicate to enable  $S_{cell}$  computation.

For dynamic experiments on OL1-SensRegRFP, a similar procedure was followed, with the following exceptions: the 00-fold dilution was directly carried out in  $200\mu l$  of M9pH6 in the 96-well microplate, the inducer ( $100ng/ml$  for aTc) was added to the microplate wells when the culture reached an  $OD$  of about 0.02, M9pH6 was supplemented with IPTG ( $500\mu M$ ).

HSL (#K3007, Sigma Aldrich) was dissolved in deionized water to prepare a  $2mM$  stock. IPTG (#I1284, Sigma Aldrich) and aTc (#631310, Clontech) were purchased as ready-made  $200mM$  and  $2mg/ml$  stocks, respectively. The three inducers were routinely stored at  $-20^{\circ}C$ .

Absorbance and fluorescence background signals were subtracted from raw absorbance and fluorescence over time ( $t$ ), to obtain cell density ( $X$ , expressed as optical density -  $OD$  - proportional to the per-well cell count) and RFP ( $R$ , expressed as arbitrary units of raw RFP -  $AUR$  - proportional to the per-well number of fluorescent proteins). A signal proportional to RFP synthesis rate per cell ( $S_{cell}^{raw}$ , expressed as  $AUR OD^{-1} min^{-1}$ ) was computed over time in the exponential growth phase (EGP, identified via visual inspection, typically  $0.02 < OD < 0.14$  in microplate experiments) for each culture as the numeric time derivative of fluorescence, divided by the mean cell density (Equation A.18):

$$S_{cell,i}^{raw} = \frac{R_i - R_{i-1}}{t_i - t_{i-1}} \cdot \frac{2}{X_i + X_{i-1}}, \quad \forall i \in EGP \quad (A.18)$$

The  $S_{cell}^{raw}$  time series of the REF strain replicates ( $S_{cell,REF}^{raw}$ ) were averaged over the EGP and the mean value ( $\overline{S_{cell,REF}^{raw}}$ ) among the replicates was computed.  $\overline{S_{cell,REF}^{raw}}$  was used to normalize the  $S_{cell,REF}^{raw}$  values for





## A.6. Supplementary Notes

---

all the other cultures, obtaining the  $S_{cell}$  time series, expressed as  $AU \text{ min}^{-1}$  (Equation A.19):

$$S_{cell,i} = \frac{S_{cell,i}^{raw}}{S_{cell,REF,i}^{raw}} \quad (\text{A.19})$$

In steady-state experiments, the computed  $S_{cell}$  time series were averaged over time in the EGP, as described above, to obtain  $\overline{S_{cell}}$ .

Growth rate ( $\mu$ ) was computed as the slope of the regression line of the  $\log(X(t))$  time series in the EGP. The microplate reader was also used to measure cell density at specific time points in cultures growing in 15-ml tubes or turbidostat, as described above without programming kinetic cycles. In this case, sterile M9pH6 was always included as absorbance background, which was subtracted from the raw measured values of the culture under study to obtain  $X$ .

### Enzyme assays

Data from representative experiments of HSL biosynthesis and degradation are shown in the figure below (see Figure A.8). Three to four samples ( $OD$  and HSL) were measured over time in the exponential phase.

## Supporting results

### RBS selection for *luxI* and *aiiA*

While host strain, promoters and enzyme coding sequences were selected to meet the functional requirements of the designed circuit in terms of regulation, as described in section A.2, ribosome binding sites (RBSs) still needed to be selected to enable efficient translation of *LuxI* and *AiiA*. Translation efficiency is known to be highly RBS- and coding sequence-dependent [116] and, despite computational tools are

## A. Synthetic close-loop controller

---

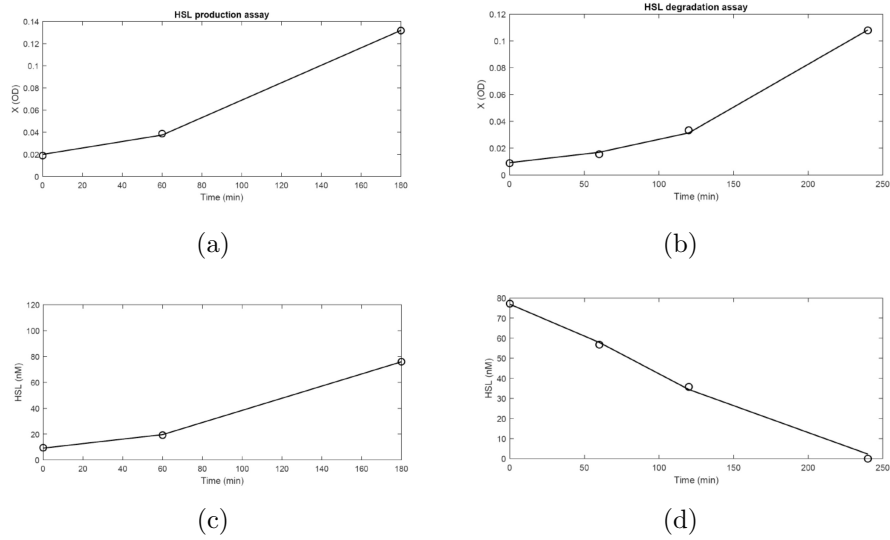


Figure A.8: **Example of data from HSL production (a,c) and degradation (b,d) assays.** Panels a. and b. show growth curves. Panels c and d show HSL time courses. These data come from an HSL production assay in which OL1 was induced with  $100\text{ng/ml}$  of ATc, while in the HSL degradation assay the ConstAct-MC construct with the J23101 constitutive promoter was used

## A.6. Supplementary Notes

Table A.2: **RBS selection for luxI and aiiA.** HSL concentration ( $nM$ ) after  $4-h$  from aTc ( $100ng/ml$ ) addition. Values are relative to single replicates assayed as described in section A.3.  $200nM$  of HSL were added at  $t = 0$  for experiments with aiiA.

RBS	BBa_B0030 -= <b>dimensionless</b>	BBa_B0031 (NBM, training set)	BBa_B0032 (BM, training set)	BBa_B0034 (BM, global fitting)
luxI	20	2	0	18
aiiA	171	22	169	0

available to support its prediction, their accuracy is currently limited [23]. For this reason, we chose the RBSs for LuxI and AiiA translation according to preliminary HSL production and degradation assays, respectively. Three popular RBSs from the Registry were assembled upstream of luxI and aiiA, under the regulation of  $P_{LtetO1}$  in a low- and medium-copy plasmid, respectively (see Table A.2). Some of the recombinant strains showed no HSL production for LuxI or HSL degradation for AiiA (see Table A.2), despite cell growth was comparable (data not shown). There was no consistency between the luxI and aiiA results in terms of RBS efficiency, e.g., BBa\_B0030 corresponded to the lowest activity for AiiA and the strongest for LuxI, and vice-versa for BBa\_B0031, confirming the relevant context-dependency of RBSs. The BBa\_B0030 and BBa\_B0034 RBSs were finally selected to enable efficient translation of LuxI and AiiA, respectively.

### Characterization of the promoters used in the measurement circuits

Experimental data are shown for all the promoter-RBS-plasmid sets used to drive the expression of aiiA that are not present in the final close-loop circuit. (see Figure A.9).

## A. Synthetic close-loop controller

---

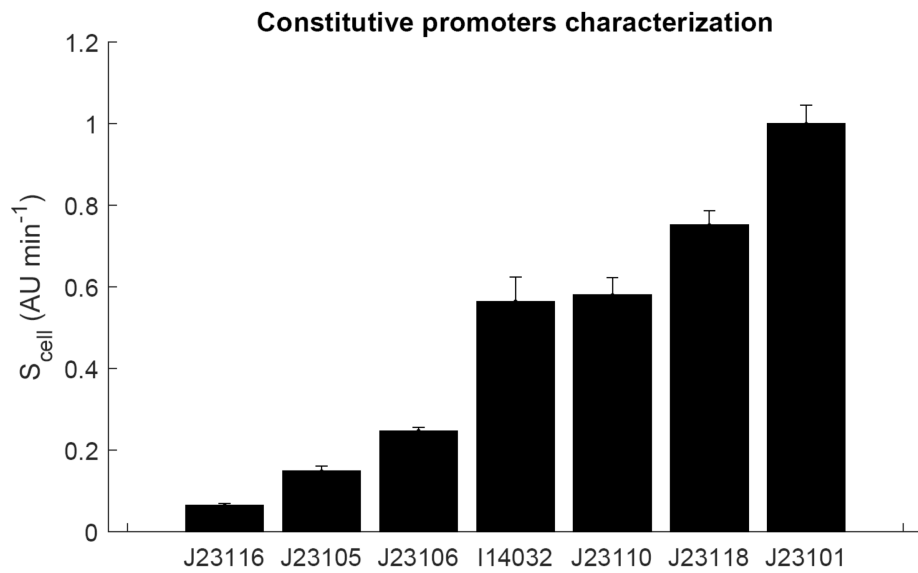


Figure A.9: **Characterization of the constitutive promoters used to drive *aiiA*.** Bars represent the average of at least three independent replicates and error bars represent standard errors of the mean.

**Additional data and modeling on interconnected subcircuits** **$P_{lux}$  transfer function in presence of over-expressed LuxR**

The  $P_{lux}$  promoter activity in the SensRegRFP-MC and OL1-SensRegRFP strains was also characterized in presence of IPTG ( $500\mu M$ ) to over-express the cognate LuxR transcriptional activator. Fluorescence assay results are shown in Figure A.10 with the fitted curves.

Results showed that the maximum activity is lower in co-transformed strain (as it was observed in absence of IPTG), but, in contrast with the experiments without IPTG, the half-maximum concentration value is comparable (see parameters in the Figure A.10 caption). It is known that cell load, caused by heterologous circuits (e.g., an additional expression plasmid in the cells) can decrease protein expression, consistent with the lower maximum activity in OL1-SensRegRFP compared with SensRegRFP-MC. Cell load globally affects the expression of all proteins, including LuxR, which is responsible of  $P_{lux}$  activation. When there is an excess of LuxR in the cells, the activation curve is expected to be stably maintained upon variations of LuxR; otherwise, small variations of LuxR may lead to changes in transfer function shape, namely, a decrease of LuxR can decrease the sensitivity of the switch by increasing the half-maximum concentration value and also by further decreasing the maximum activity [57]. Since without IPTG LuxR is expressed at very low levels, the change of  $\alpha_{lux}$  and  $K_{lux}$  values is fully compatible with cell burden, occurring on both LuxR and RFP in presence of an additional plasmid in OL1-SensRegRFP. Despite recent methodologies to study cell load via fluorimetric assays have been proposed, an extensive characterization is beyond the scope of this work.

**HSL degradation by  $P_{lux}$ -driven *aiiA* in one- and two-plasmid contexts** SensRegAct-MC and CL1 (in absence of ATc induction) were characterized to confirm that  $P_{lux}$  activity changes in presence of

## A. Synthetic close-loop controller

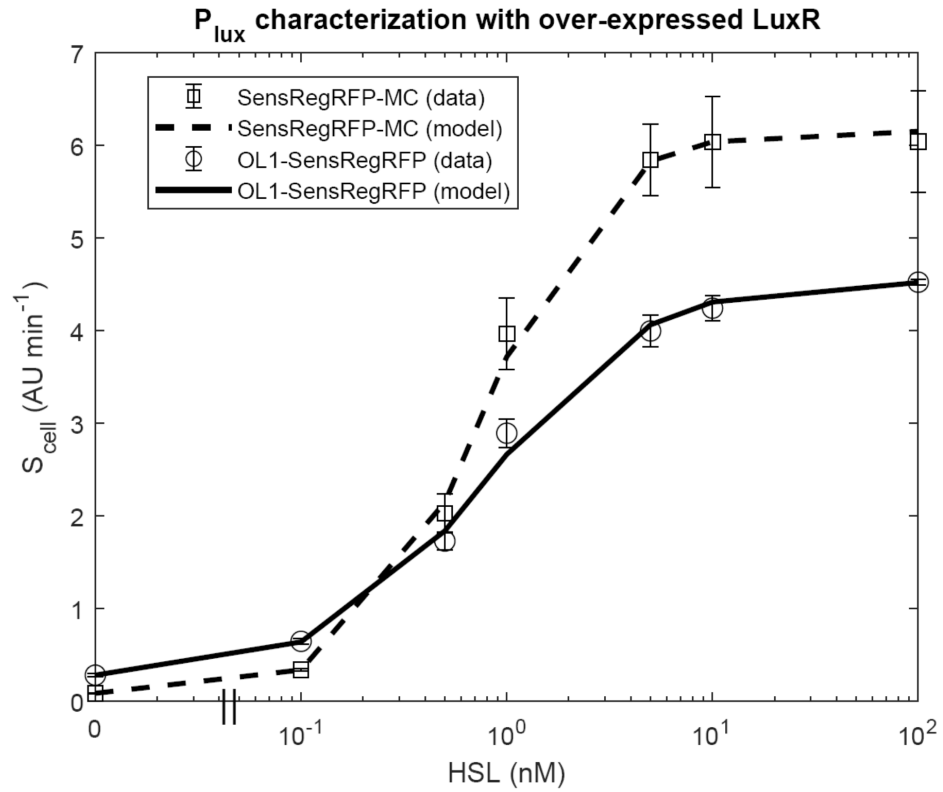


Figure A.10: **Fitting of SensRegRFP-MC and OL1-SensRegRFP in presence of IPTG to over-express LuxR.** Data points represent the average of at least three independent replicates and error bars represent standard errors of the mean. Estimated parameters were:  $\alpha_{lux} = 10.65(5\%)$ ,  $\delta_{lux} = 0.014(12\%)$ ,  $K_{lux} = 0.77(5\%)$ ,  $\eta_{lux} = 1.54(5\%)$  for SensRegRFP-MC;  $\alpha_{lux} = 7.47(1\%)$ ,  $\delta_{lux} = 0.067(8\%)$ ,  $K_{lux} = 0.81(6\%)$ ,  $\eta_{lux} = 1.14(6\%)$  for OL1SensRegRFP. Coefficient of variation is reported in brackets.

## A.6. Supplementary Notes

an additional plasmid, also when the inducible gene is *aiiA*. Different experiments were carried out in batch mode by adding HSL at  $t = 0$  and then monitoring HSL degradation over time (see Figs.A.11-A.11 for the results of individual tests). The equations below (Eqs.A.20-A.21) with parameters in Table A.1 were used to simulate the experimentally measured curves (see Figures A.11-A.12), considering that SensRegAct-MC and CL1 had a different parameter set ( $\alpha_{lux}$ ,  $\delta_{lux}$ ,  $K_{lux}$ ,  $\eta_{lux}$ , see Table A.1) to describe  $P_{lux}$  activity.

$$\frac{dA}{dt} = \alpha_{lux} \cdot \left( \delta_{lux} + \frac{1 - \delta_{lux}}{1 + \left(\frac{K_{lux}}{H}\right)^{\eta_{lux}}} \right) - (\mu + \gamma) \cdot A, \quad A(t_0) = \frac{\alpha_{lux} \cdot \delta_{lux}}{\mu + \gamma} \quad (\text{A.20})$$

$$\frac{dh}{dt} = - \left( \frac{\beta_A}{1 + \left(\frac{\kappa_A}{A}\right)^{n_A}} \cdot X + D + \gamma_H \right) \cdot H, \quad H(t_0) = H_0 \quad (\text{A.21})$$

where symbols have the same meaning as sections above. For each experiment,  $H_0$  was fixed to the measured value and the cell concentration time course was fitted via exponential model (Equation A.1 for  $D = 0$ ).

Prediction results (also summarized in Figure A.13) showed that the identified models successfully describe the experimental time series, and CL1 (in absence of ATc induction) had an overall higher HSL concentration at the end of experiment than SensRegAct-MC, as expected, since the former has a lower  $P_{lux}$  activity that can lead to lower HSL degradation.

**Model simulations in presence of additional dynamic processes** To assess the assumptions about kinetic processes involved in the designed system, i.e., that per-cell protein dynamics are gov-

## A. Synthetic close-loop controller

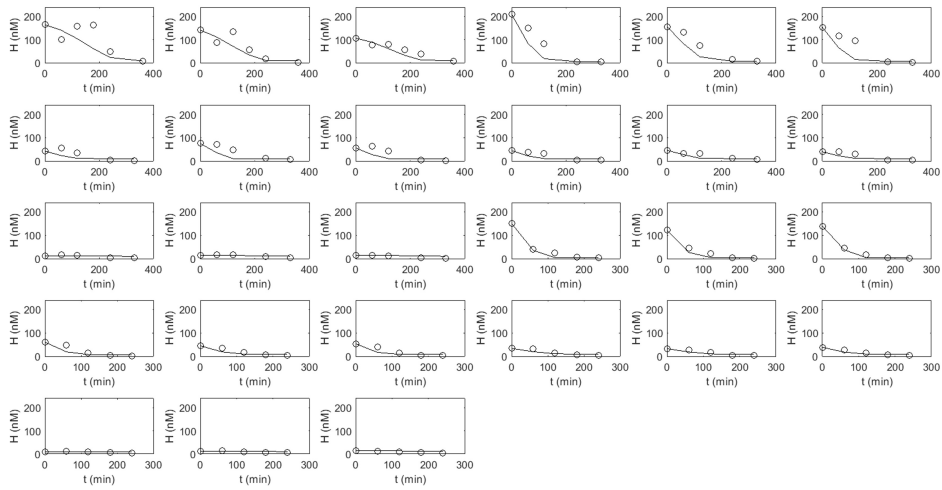


Figure A.11: **Time course experiments for SensRegAct-MC.** Circles represent the measured HSL concentration and solid lines represent model predictions.

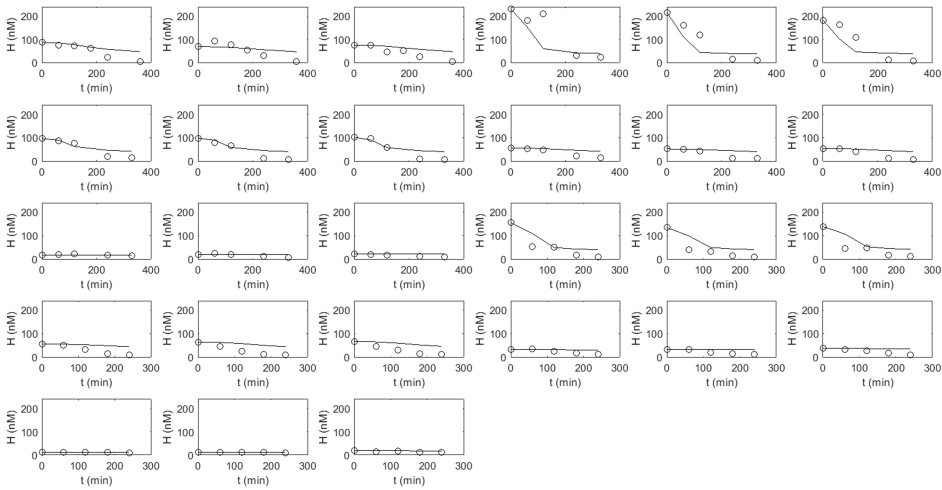


Figure A.12: **Time course experiments for CL1 without ATc.** Circles represent the measured HSL concentration and solid lines represent model predictions.



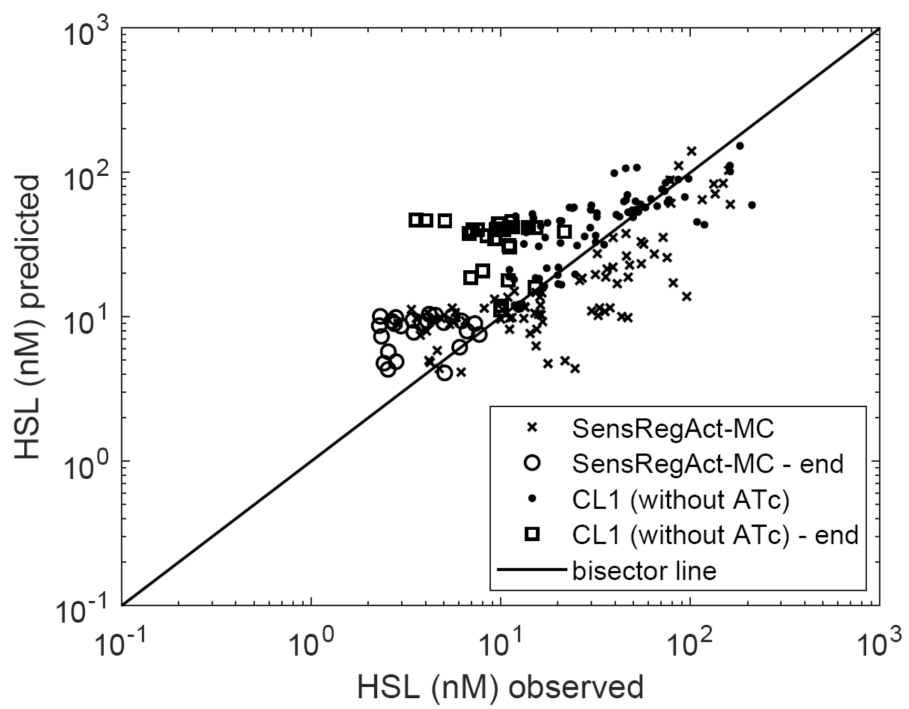
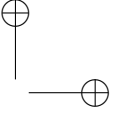
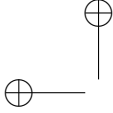


Figure A.13: Measured vs predicted HSL concentration in time course experiments of SensRegAct-MC and CL1 without ATc.



## A. Synthetic close-loop controller

---

erned by their inducer-dependent synthesis rate and their degradation/dilution rate with no other dynamic process, the OL1-SensRegRFP sub-circuit was adopted. In particular, a time course experiment was carried out by measuring RFP after the induction of HSL production module via ATc, and in presence of  $500\mu M$  IPTG. In this experiment, every dynamic process occurring in the cells, from ATc addition to RFP maturation, plays a role in determining RFP synthesis rate per cell time course. This test enables to understand if the modeled processes (i.e., enzymes degradation, HSL production,  $P_{lux}$  induction by HSL and RFP maturation) are sufficient to accurately describe the experimental data, or the inclusion of additional kinetic processes, e.g., in protein synthesis rate activation, is necessary. Simulation of the identified model in Eqs.A.22-A.25 (derived from Eqs.A.1-A.4 and Eqs.A.10-A.12) confirmed that experimental time course is captured with sufficient accuracy by considering only the dynamic processes included in this model (see Figure A.14). Despite not all the RFP profile was perfectly captured (e.g., the observed RFP synthesis rate superelongation is not described by the model), simulations were considered to be reasonably accurate in capturing RFP activation, without adding equations and parameters to the model.

## A.6. Supplementary Notes

---

$$\frac{dX}{dt} = \mu \cdot X, \quad X(t_0) = X_0 \quad (\text{A.22})$$

$$\frac{dL}{dt} = \alpha_{tet} \cdot \left( \delta_{tet} + \frac{1 - \delta_{tet}}{1 + \left( \frac{K_{tet}}{T} \right)^{\eta_{tet}}} \right) - (\mu + \gamma) \cdot L,$$

$$L(t_0) = \frac{\alpha_{tet} \cdot \delta_{tet}}{(\mu + \gamma)} \quad (\text{A.23})$$

$$\frac{dH}{dt} = \frac{\beta_L}{1 + \left( \frac{\kappa_L}{L} \right)^{n_L}} \cdot X - \gamma_H \cdot H, \quad H(t_0) = 0 \quad (\text{A.24})$$

$$\frac{dS_{cell}}{dt} = m \cdot \alpha_{lux} \cdot \left( \delta_{lux} + \frac{1 - \delta_{lux}}{1 + \left( \frac{K_{lux}}{H} \right)^{\eta_{lux}}} \right) - (\mu + m) \cdot S_{cell},$$

$$S_{cell}(t_0) = \frac{m \cdot \alpha_{lux} \cdot \delta_{lux}}{\mu + m} \quad (\text{A.25})$$

In Eqs.A.22-A.25,  $X_0$  was set to the experimentally measured cell density at  $t = 0$ .

Moreover, to confirm that additional dynamic processes are not crucial to be included, the final close-loop system was also simulated by considering a dynamics in the enzyme synthesis rate per cell (Eqs.A.26-

## A. Synthetic close-loop controller

---

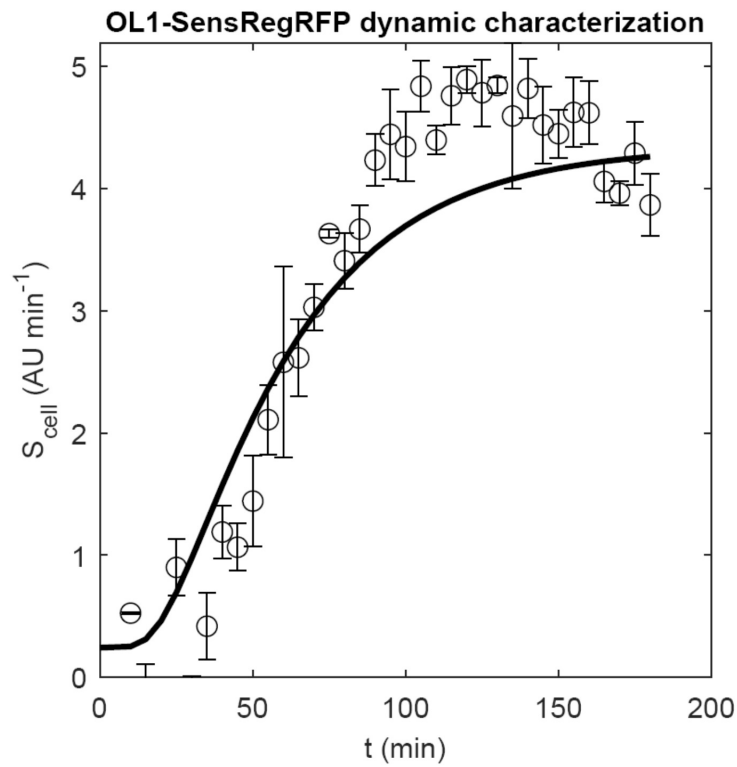


Figure A.14: **Time course experiment of OL1-SensRegRFP in response to ATc addition ( $100\text{ng/ml}$  at  $t = 0$ ).** Circles represent the average of three independent measurements, error bars represent standard errors of the mean and solid line represents model simulation.

A.31).

$$\frac{dX}{dt} = \mu \cdot X - D \cdot X, \quad X(t_0) = X_0 \quad (\text{A.26})$$

$$\frac{dS_L}{dt} = r_{tet} \cdot \alpha_{tet} \cdot \left( \delta_{tet} + \frac{1 - \delta_{tet}}{1 + \left( \frac{K_{tet}}{T} \right)^{\eta_{tet}}} \right) - r_{tet} \cdot S_L,$$

$$S_L(t_0) = \alpha_{tet} \cdot \delta_{tet} \quad (\text{A.27})$$

$$\frac{dS_A}{dt} = r_{lux} \cdot \alpha_{lux} \cdot \left( \delta_{lux} + \frac{1 - \delta_{lux}}{1 + \left( \frac{K_{lux}}{H} \right)^{\eta_{lux}}} \right) - r_{lux} \cdot S_A,$$

$$S_L(t_0) = \alpha_{lux} \cdot \delta_{lux} \quad (\text{A.28})$$

$$\frac{dL}{dt} = S_L - (\mu + \gamma) \cdot L, \quad L(t_0) = \frac{\alpha_{tet} \cdot \delta_{tet}}{(\mu + \gamma)} \quad (\text{A.29})$$

$$\frac{dA}{dt} = S_A - (\mu + \gamma) \cdot A, \quad L(t_0) = \frac{\alpha_{lux} \cdot \delta_{lux}}{(\mu + \gamma)} \quad (\text{A.30})$$

$$\frac{dH}{dt} = \frac{\beta_L}{1 + \left( \frac{\kappa_L}{L} \right)^{n_L}} \cdot X - \left( \frac{\beta_A}{1 + \left( \frac{\kappa_A}{A} \right)^{n_A}} \cdot X + D + \gamma_H \right) \cdot H,$$

$$H(t_0) = 0 \quad (\text{A.31})$$

In this model, S represent the protein synthesis rates per cell ( $AU \text{ min}^{-1}$ ) of LuxI ( $S_L$ ) and AiiA ( $S_A$ );  $\frac{1}{r}$  represent the time constants of protein synthesis rates dynamics. The  $r$  parameters were fixed to  $0.1 \text{ min}^{-1}$  (i.e.,  $7 \text{ min}$  of half-induction time) to run the model with a worst-case time constant value [93]. The other symbols have the same meaning as in the main text. Results showed that the simulation of the Eqs.1-4 (main text) and Eqs.S5-S8 models did not lead to relevant differences in experimental time course predictions (see

## A. Synthetic close-loop controller

Figure A.15). Finally, prediction performances, in terms of correlation (see section A.4), of CL1 and OL1 did not improve by using the model in Eqs.A.26-A.31 for different values of  $r$  (data not shown). Taken together, the described results suggested that the behavior of the designed circuit did not change by including an additional dynamic process on the protein synthesis rates per cell.

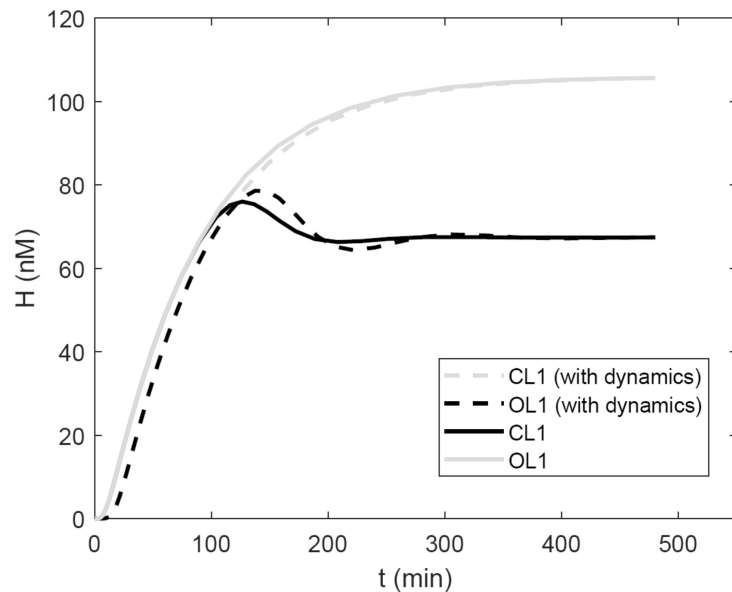


Figure A.15: **Predicted time course of HSL in turbidostat at  $X = 0.2$  for OL1 and CL1.** Simulated using the model described in the main text and the model including protein synthesis rate dynamics with  $r_{tet} = r_{lux} = 0.1 \text{ min}^{-1}$ .

### Sensitivity analysis of the identified model

A sensitivity analysis was carried out on the 7 parameters involved in close-loop regulation, i.e., the Hill function parameters describing



## A.6. Supplementary Notes

---

the behavior of sensor, regulator and actuator. Univariate analysis was performed by investigating the effect on steady-state HSL output level ( $H^{SS}$ ) for small variations of the parameter (10%), given the values of all the other parameters. In particular, for each  $i$ -th parameter  $p_i$  under study, sensitivity indexes  $S_i$  were computed as described in Equation A.32 to yield a dimensionless value.

$$S_i = \frac{\Delta H^{SS}/H^{SS}}{\Delta p_i/p_i} \quad (\text{A.32})$$

Results of sensitivity analysis are reported in Figure A.16 for  $X = 0.2$ . At this cell density value, the sensor/regulator/actuator set is actively involved in HSL control, while for  $X$  values lower than 0.1 their contribution is small and the system works as an open-loop circuit (see section A.2). For this reason, the impact of the studied parameters for  $X < 0.1$  is negligible and results in sensitivity indexes close to zero (data not shown). The indexes computed in Figure A.16 show that many parameters have relevant impact on output level: among the parameter set describing the actuator transfer function (i.e.,  $\beta_A$ ,  $\kappa_A$ ,  $n_A$ ), variation of  $\kappa_A$  has the largest impact. These effects are explained by the shape of the transfer function of AiiA (see Figure A.4(d)), in which the estimated value of the Hill coefficient is very large and small variations do not lead to large  $H^{SS}$  changes; similarly,  $\beta_A$  variations do not significantly alter  $H^{SS}$  since the AiiA never reaches intracellular levels higher than  $\kappa_A$  (see Figure A.17), thereby motivating the low importance of  $\beta_A$ . It is worth noting that, despite only a small part of the AiiA transfer function is spanned at the steady-state, the full AiiA activity range is spanned transiently before reaching the  $H^{SS}$  equilibrium value due to H and A oscillations, and during H disturbance rejection in which AiiA becomes over-expressed in response to HSL levels above  $H^{SS}$ . As expected, even small changes of  $\kappa_A$  alter the  $H^{SS}$  value, since a change in the half-maximum activity of the

## A. Synthetic close-loop controller

AiiA switch can tune the intracellular level at which enzymatic activity starts increasing. Concerning the parameter set describing the sensor/regulator joint transfer function (i.e.,  $\alpha_{lux}$ ,  $\delta_{lux}$ ,  $K_{lux}$ ,  $\eta_{lux}$ ), all of them except  $\delta_{lux}$  have sensitivity index above 0.5, thereby demonstrating their important impact on  $H^{SS}$  upon small variations. Such impact is indirectly explained by the large impact of  $\kappa_A$ , illustrated above: a change in the  $P_{lux}$  transfer function shape results in a variation of AiiA level, thereby affecting the HSL-dependent AiiA activity which is eventually responsible of HSL degradation.

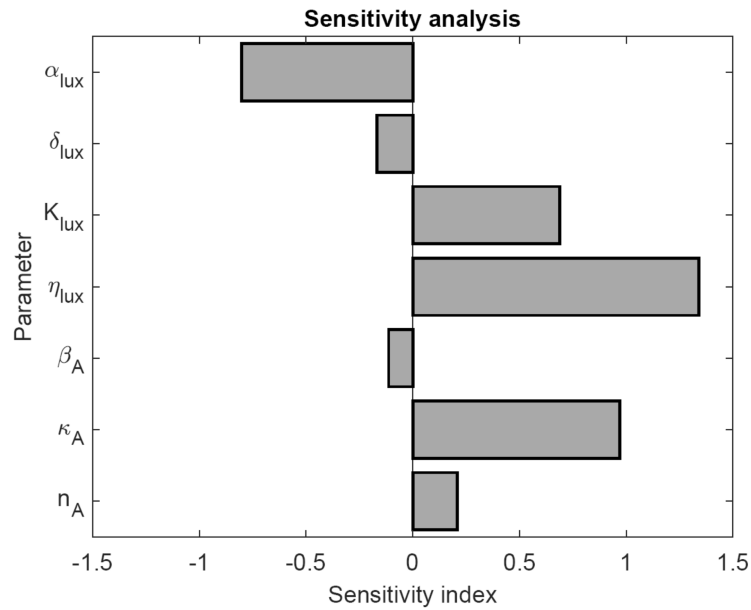


Figure A.16: Sensitivity indexes for the parameters involved in close-loop control for  $X = 0.2$ .



## A.6. Supplementary Notes

---

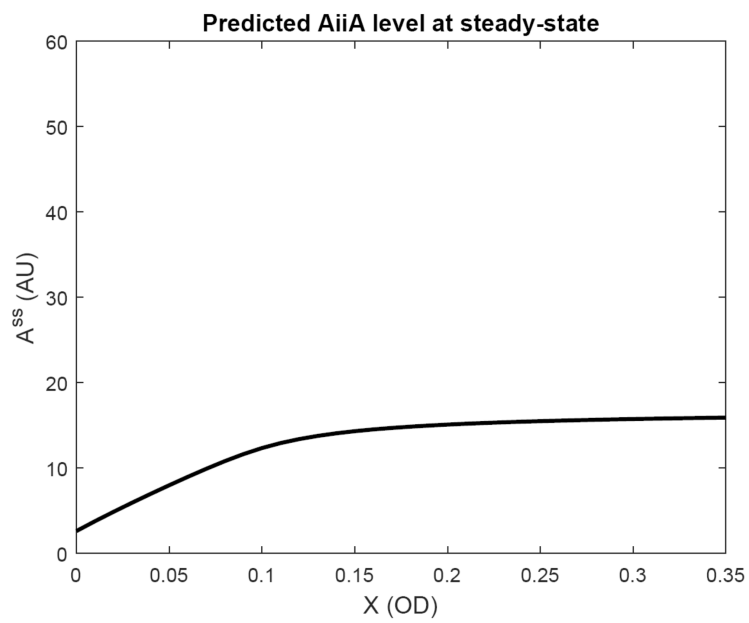


Figure A.17: Simulated steady-state level of intracellular AiiA ( $A^{SS}$ ) for different values of  $X$ .

## A. Synthetic close-loop controller

### HSL time courses in turbidostat

The full experimental dataset of chemostat experiments for CL1 and OL1 is herein shown, in terms of cell density and HSL measurements, as well as the predicted HSL profiles (see Figures A.18-A.22).

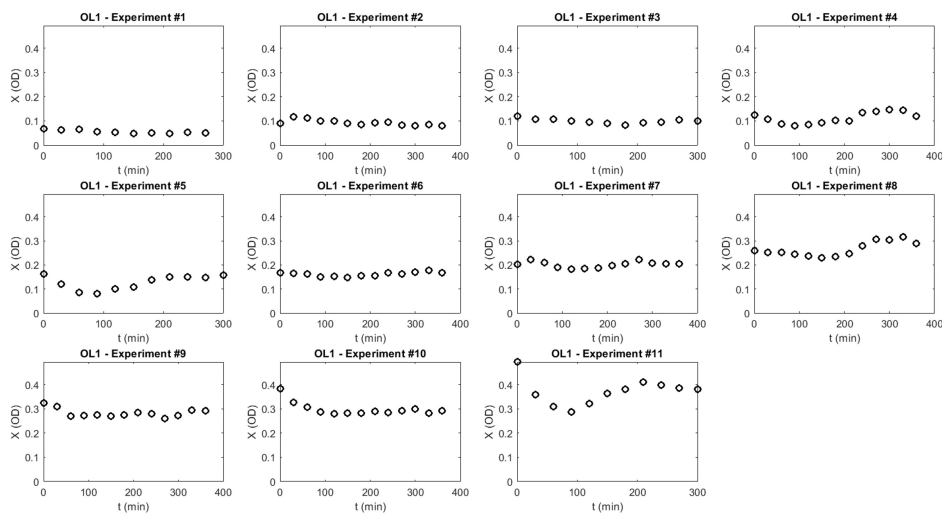


Figure A.18: Cell density time courses in turbidostat experiments of OL1. Circles represent measured  $X$  values.

### Supporting conclusions

#### Alternative designs of close-loop circuits for HSL control

Although the in-depth analysis of alternative close-loop design schemes is beyond the scope of this work, it is worth mentioning that HSL regulation could be achieved via other strategies. In an effort to construct a circuit for protein production at predictable mean and variance, Vignoni et al. [115] proposed a different design in which luxI

## A.6. Supplementary Notes

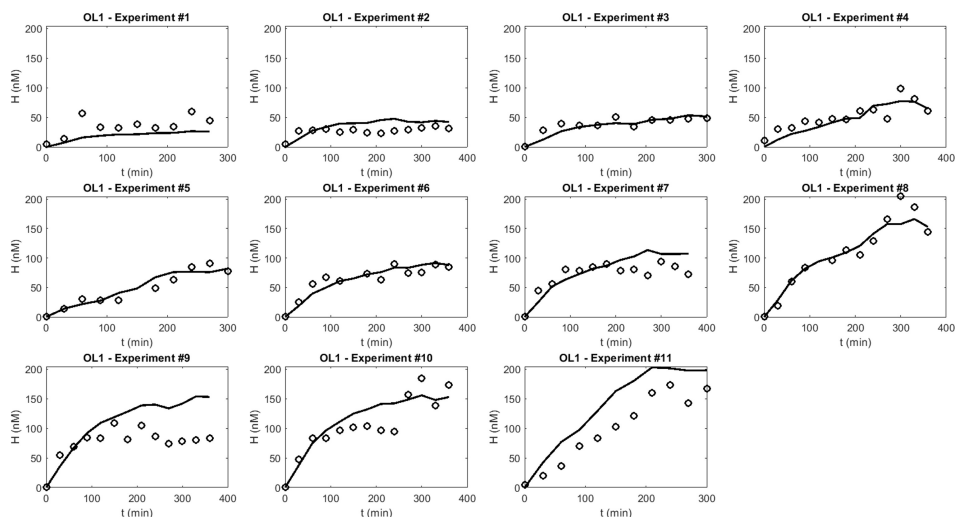


Figure A.19: **HSL time courses in turbidostat experiments of OL1.** Circles represent measured HSL values and solid line represents model prediction.

was expressed by an HSL-repressible promoter, thereby closing the loop at transcriptional level. Intuitively, disturbance rejection upon HSL injections could not be efficiently exerted in such design, since no HSL degradation module was present, and the regulation affected HSL production only. A more complex design mixing the one proposed in [115] and the one of our work could be possible, with the advantage of producing HSL only when necessary, thereby exerting a double control (transcriptional and enzymatic) on the HSL biosynthesis module, thereby potentially implementing a faster dynamics in reaching the steady-state than open-loop design, and eventually decreasing the cellular energy demand. However, the fine tuning of additional parts would be necessary and simulations revealed that such tuning is crucial to avoid dead-zones in HSL control (data not shown).

## A. Synthetic close-loop controller

---

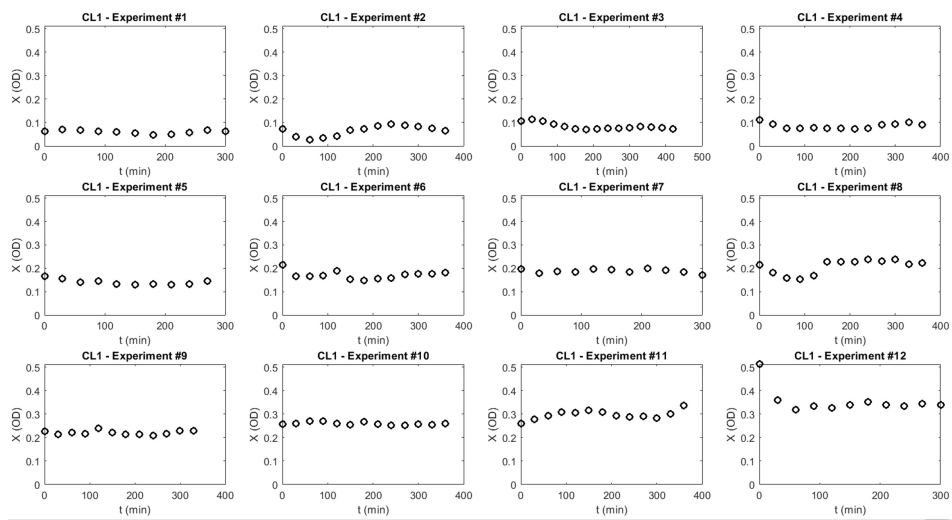


Figure A.20: Cell density time courses in turbidostat experiments of CL1. Circles represent measured  $X$  values.

## A.6. Supplementary Notes

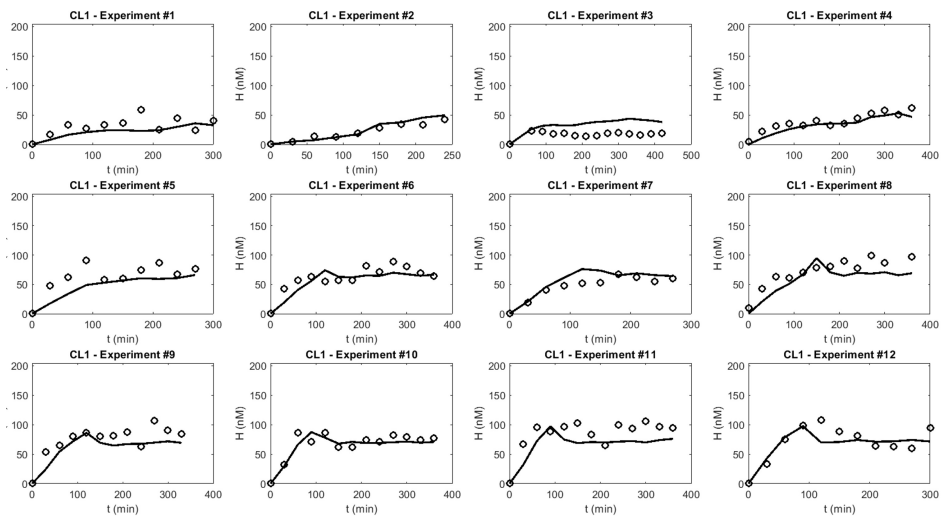


Figure A.21: HSL time courses in turbidostat experiments of CL1. Circles represent measured HSL values and solid line represents model prediction.

## A. Synthetic close-loop controller

---

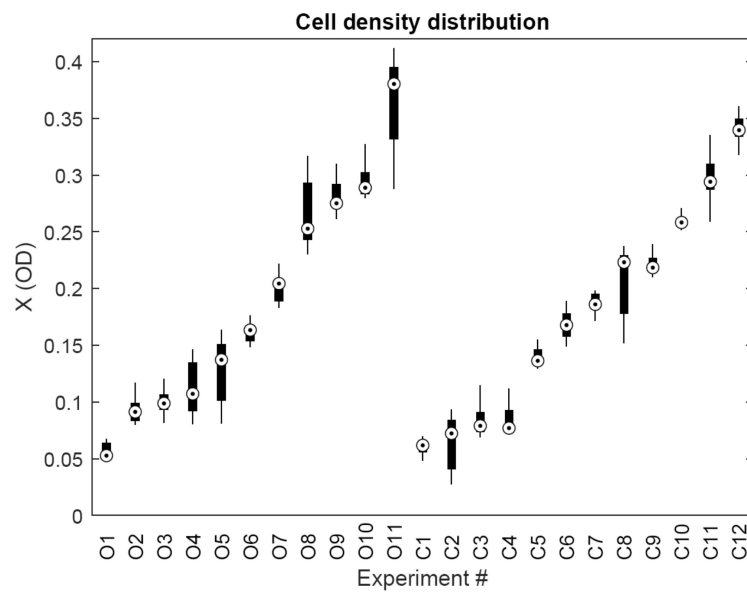
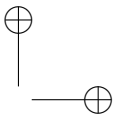
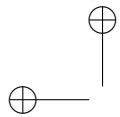


Figure A.22: **Summary of cell density values distribution in all the turbidostat experiments of OL1 and CL1.** Boxplots are sorted by mean value of cell density. The experiments are labeled with the *O* and *C* prefixes for OL1 and CL1, respectively.



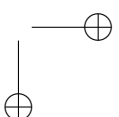
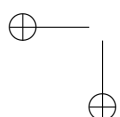
# Appendix **B**

## Supplementary information for Chapter 2

### B.1 Strains, reagents and cloning

The *E. coli* TOP10 (Invitrogen) strain was used as a host for cloning and quantitative assays. The strain was transformed by heat shock at 42 °C, according to manufacturer's instructions. LB medium was used during plasmid propagation. Antibiotics were always added to maintain plasmids in recombinant strains: ampicillin (100mg/l), kanamycin (50mg/l) or chloramphenicol (12.5mg/l). Long-term bacterial stocks were prepared for all the engineered strains by mixing 750µl of a saturated culture with 250µl of 80% glycerol, and stored at -80°C.

All the plasmids used in this study were constructed through BioBrick™ Standard Assembly [147] and conventional molecular biology techniques. As a result, standard DNA junctions (TACTAG upstream of coding sequences, TACTAGAG otherwise) are present between assembled parts. The BioBrick™ basic or composite parts used for





## B. Supplementary information for Chapter 2

---

DNA assembly were retrieved from the MIT Registry 2008-2011 DNA Distribution [102], except  $P_{luxRep}$ , which was constructed in a previous study [84], and the weak BBa\_B0033 RBS that was placed upstream of tetR gene via mutagenic PCR, replacing BBa\_B0031.

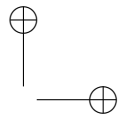
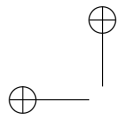
DNA purification kits (Macherey-Nagel), restriction enzymes and T4 DNA ligase (Roche), Phusion Hot Start II PCR kit and T4 polynucleotide kinase (Thermo Scientific) were used according to manufacturer's instructions. Plasmids were sequenced via the BMR Genomics DNA analysis service (Padova, Italy). Oligonucleotides for mutagenesis (REV\_LUXWT: 5'-tttattcgactataacaaccattttcttgcg-3', REV\_LUXREP: 5'-gctagcattatacctgtacgatcctacaggtg-3', FWD\_Tet33: 5'-tactagagtcacacaggactactagatgtccagattagataaaaagtaaag-3') were obtained from Metabion International AG.

M9 supplemented medium (11.28g/l M9 salts, 1mM thiamine hydrochloride, 2mM  $MgSO_4$ , 0.1mM  $CaCl_2$ , 0.2% casamino acids and 0.4% glycerol) was used in quantitative experiments. HSL (#K3007, Sigma Aldrich) was dissolved in deionized water to prepare a 2mM stock, stored at  $-20^\circ\text{C}$ .

## B.2 Data overview

The maximum output value and switch point of the input devices with  $P_{lux}$  and  $P_{luxRep}$  ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$  and  $X_{rep}$ ) is consistent with previous studies where similar circuits were characterized [13, 84, 87, 91, 92, 93]. Among these studies, a circuit analogous to  $X_{3r}$ , in the same plasmid, gave a maximum activity value of 8 RPU [87] and the half-maximum value was at  $\sim 2nM$  HSL [87, 93]. A different study with a circuit identical to  $X_{2r}$ , however, reported a lower maximum activity (4 RPU) and a much higher switch point ( $> 700nM$ ) than observed here [13]. All the mentioned studies were carried out in different E. coli strains, which might explain the observed quantitative



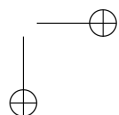
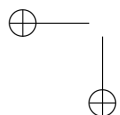


## B.2. Data overview

---

differences. In particular, in the latter case [13], the strain constitutively over-expressed LacI, resulting in a low LuxR production which can affect the static transfer function by lowering its maximum value and increasing the switch point [91]. The characterization of  $X_{1r}$  was also available from a previous study [93] in TOP10, but in a high-copy vector and in non-standard units, yielding a switch point ( $\sim 1nM$  of HSL) comparable to the one in the present work. Analogously,  $P_{luxRep}$  showed a consistent behavior with the previously reported characterization [84].

Among the 2–block cascades, the TetR-based NOT gate characterization showed a higher maximum activity than previously reported in a different strain, but the input level, in terms of RPU, that causes a half-maximum response was highly consistent [13]. The LacI-based NOT gate was not characterized previously in a circuit with this RBS. As expected, the maximum value of  $P_{LlacO1}$  was much lower than previously reported in the same strain ( $\sim 6$  RPU) or a different one ( $\sim 2.5$  RPU) [13], since it was always characterized in presence of the gratuitous inducer isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), which binds the endogenously expressed LacI repressor of  $P_{LlacO1}$ . In the strain used in this study, LacI is constitutively expressed at low level, resulting in a relatively low output for  $P_{LlacO1}$  ( $\sim 0.5$  RPU).



## B. Supplementary information for Chapter 2

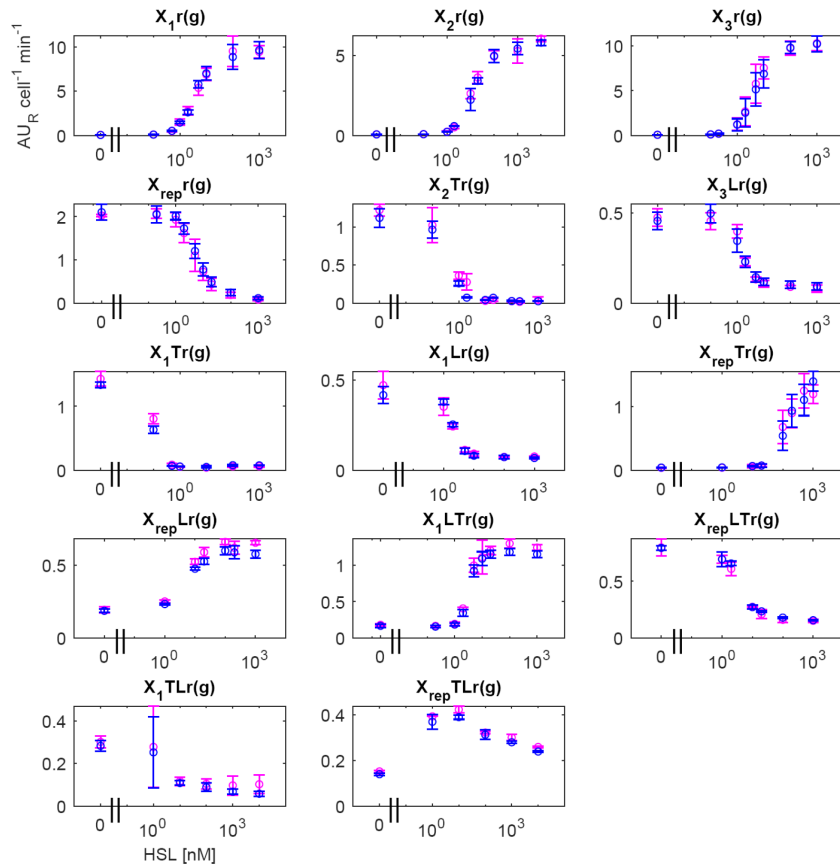


Figure B.1: **RFP output data for the cascade circuits tested in this work (training and test set) as a function of HSL concentration.** Circles represent the average measured values, while error bars represent the 95% confidence intervals of the mean. Magenta and blue colours correspond to the RFP output of circuits without (r suffix) and with (rg suffix) Monitor cassette, respectively.

## B.2. Data overview

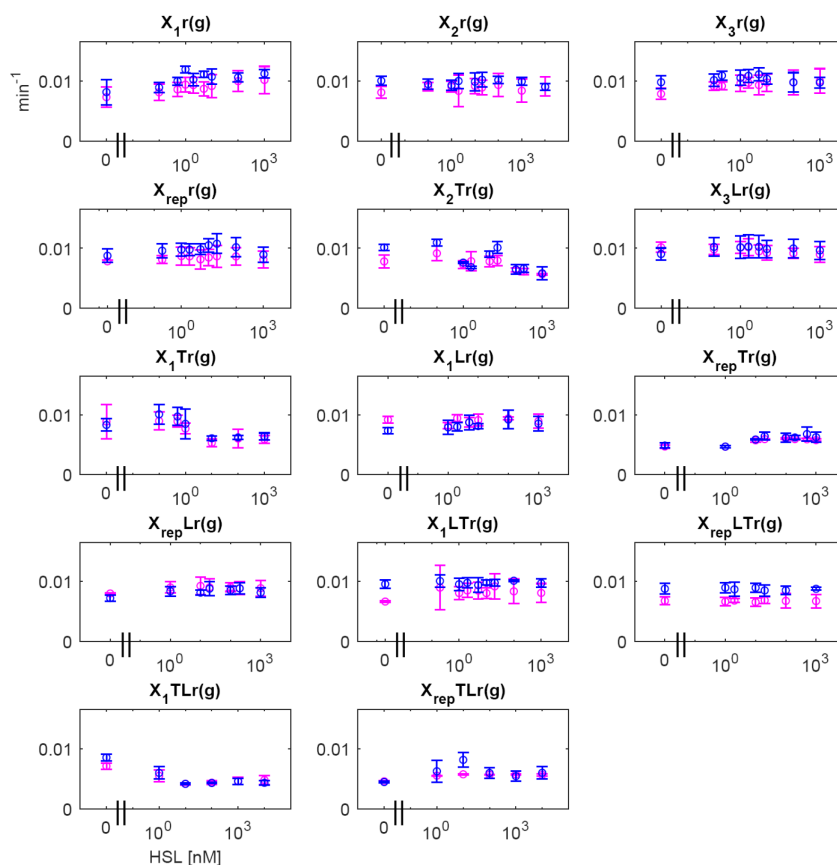


Figure B.2: **Growth rate data for the cascade circuits tested in this work (training and test set) as a function of HSL concentration.** Circles represent the average measured values, while error bars represent the 95% confidence intervals of the mean. Magenta and blue colours correspond to the growth rate of circuits without (r suffix) and with (rg suffix) Monitor cassette, respectively.

## B. Supplementary information for Chapter 2

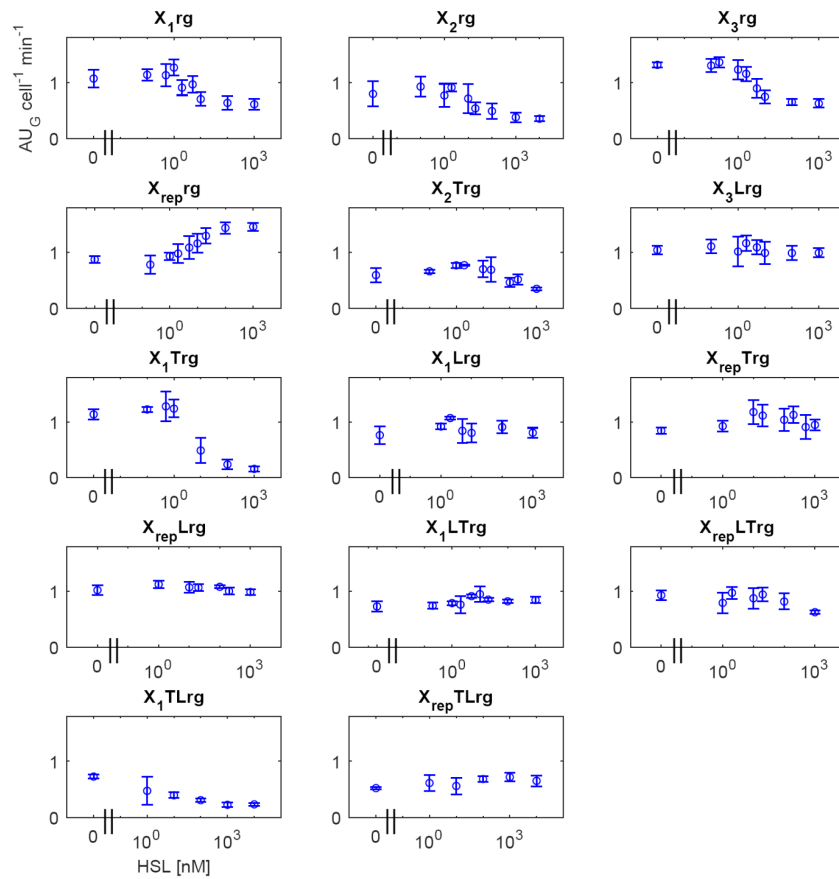


Figure B.3: **GFP output data for the cascade circuits with Monitor cassette tested in this work (training and test set) as a function of HSL concentration.** Circles represent the average measured values, while error bars represent the 95% confidence intervals of the mean.

## B.2. Data overview

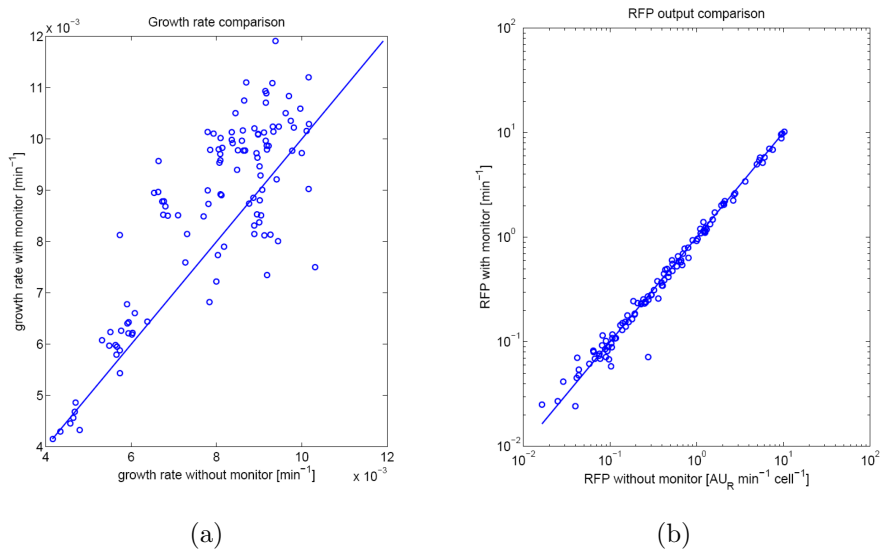


Figure B.4: **Comparison plots between strains without and with Monitor cassette.** Growth rate **a)** and RFP **b)** at all the tested HSL concentrations for all the circuits (training and test set). Circles represent the average values, while solid line represents the bisector line.

## B. Supplementary information for Chapter 2

---

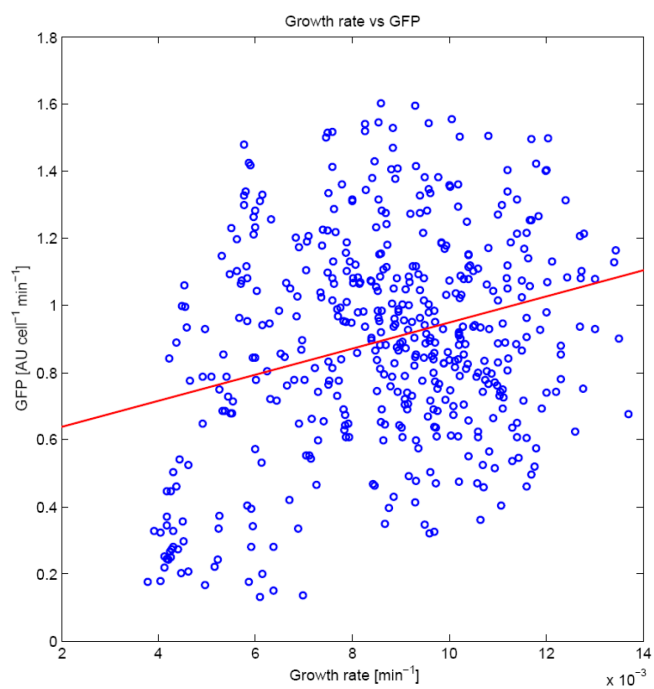


Figure B.5: **Correlation between GFP and growth rate in all the strains (training and test set) with the Monitor cassette at all the tested HSL concentrations.** Circles represent the individual values in all the biological replicates, while solid line represents the regression line.

## B.2. Data overview

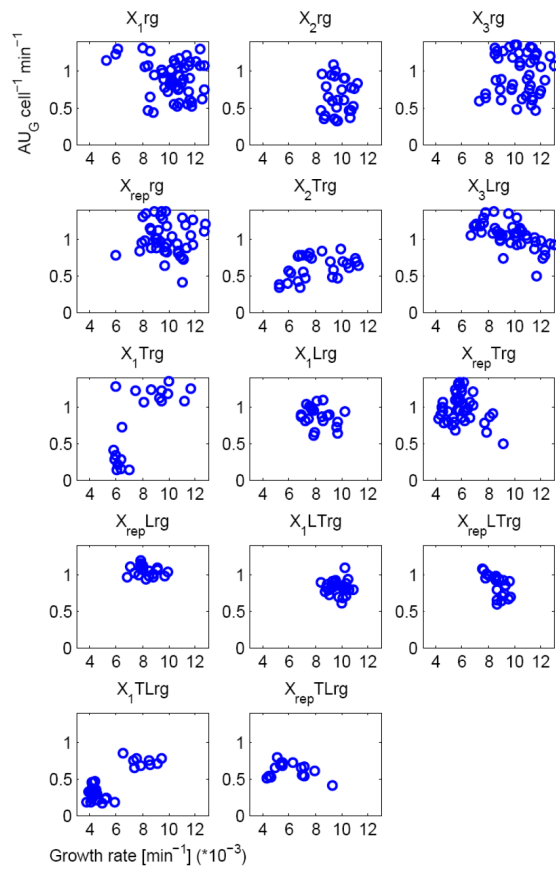
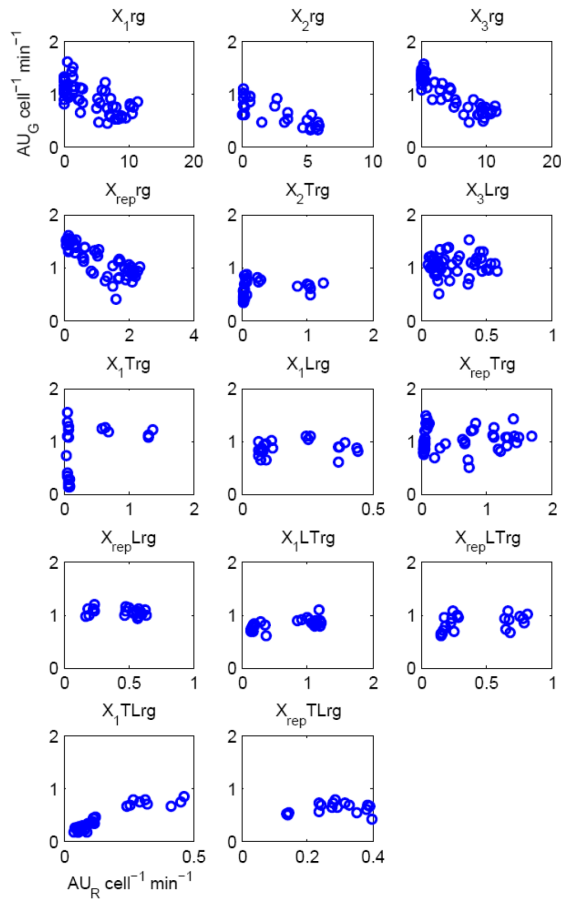


Figure B.6: **Correlation between GFP and growth rate for each strain with the Monitor cassette at all the tested HSL concentrations.** Circles represent the individually measured values in all the biological replicates.

## B. Supplementary information for Chapter 2



**Figure B.7: Correlation between GFP and RFP for each strain with the Monitor cassette at all the tested HSL concentrations.**

Correlation between GFP and RFP for each strain with the Monitor cassette at all the tested HSL concentrations. Circles represent the individually measured values in all the biological replicates.



## B.2. Data overview

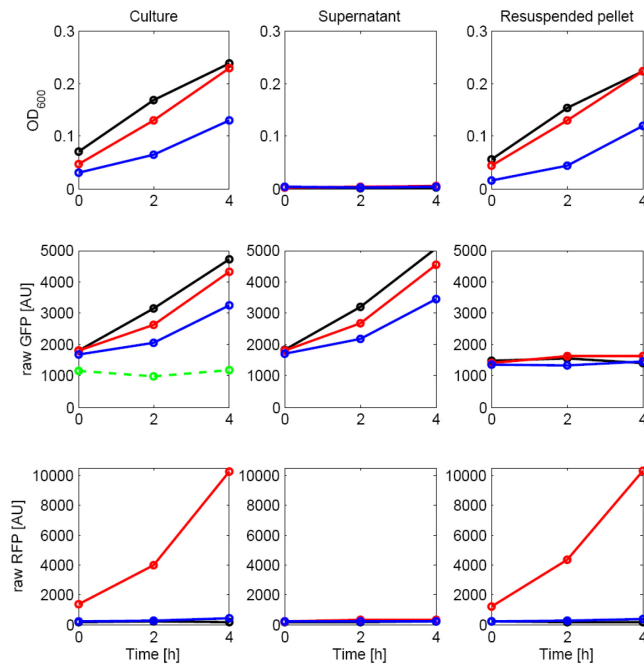
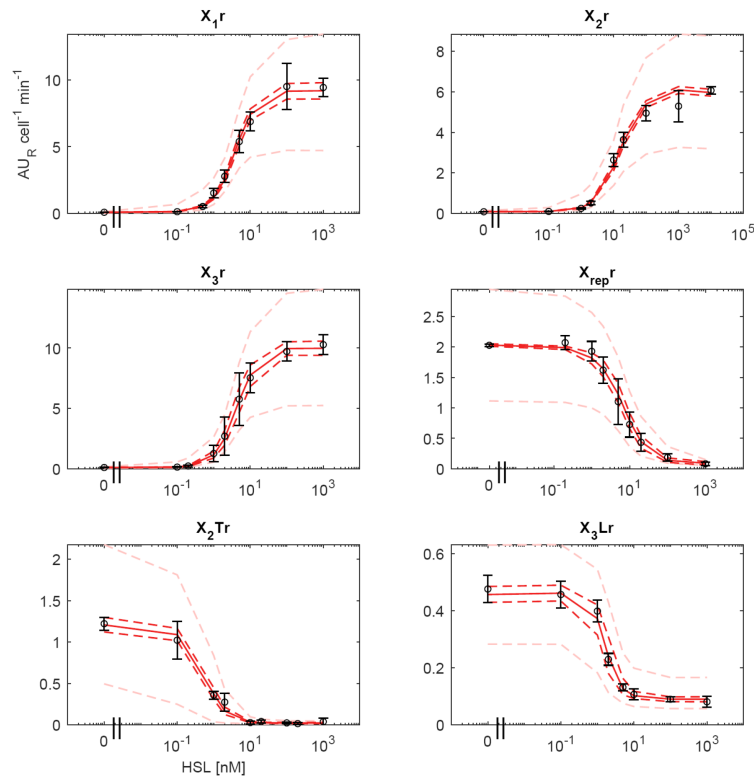


Figure B.8: **OD<sub>600</sub>, raw GFP and raw RFP values measured in culture, supernatant and pellet of three strains.** TOP10 (black; non-fluorescent), X<sub>2</sub>Tr (red; expressing RFP) and X<sub>rep</sub>Tr (blue; containing RFP, not expressed) were inoculated as described in the Section B.1; the 100-fold dilution was carried out in a final volume of 6 ml in 50 ml tubes and the cultures were incubated in the same conditions as before until they reached an OD<sub>600</sub> of about 0.05. Then, they were sampled every 2 hours. At each sampling time, absorbance, GFP and RFP were measured for 200 μl of culture; then, 1 ml was withdrawn, transferred into a 1.5 – ml tube and centrifuged (13,000 rpm, 2 min). Absorbance, GFP and RFP were measured in the supernatant (200 μl). Finally, supernatant was discarded, pellet was resuspended with 1 ml of fresh medium and absorbance, GFP and RFP were measured (200 μl). Green data points and dotted line represent the raw GFP of the medium without cells. All the measurements were carried out with the Infinite F200 reader (Tecan), as described in the Section 2.1. The reported data show that the raw GFP autofluorescence is due to the supernatant (see raw GFP in culture and supernatant), not to the cell pellet (see raw GFP in the resuspended pellet), although it increases during cell growth. On the other hand, red fluorescence is due to cell RFP expressing cells in the pellet.

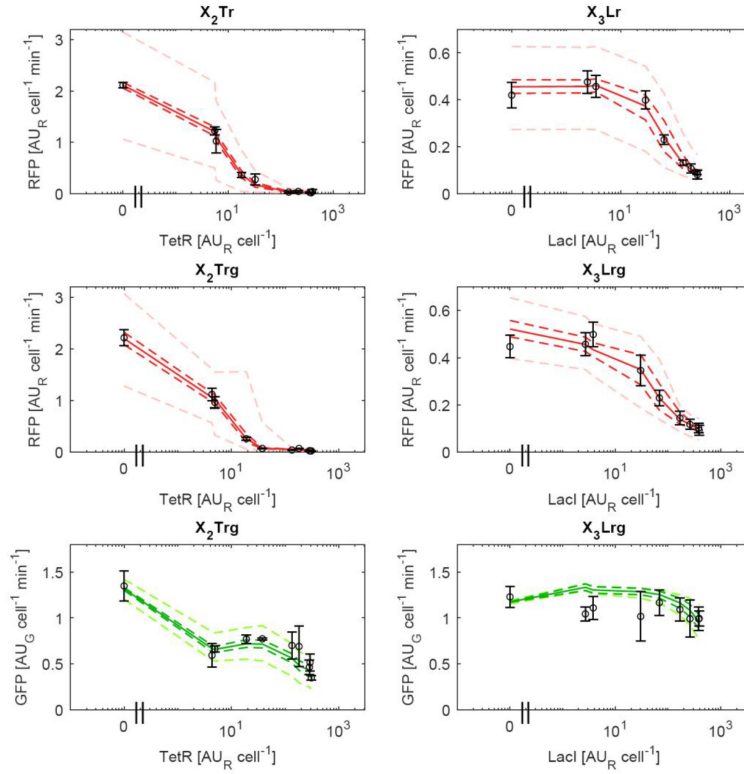
### B.3 Additional fittings and sensitivity analysis



**Figure B.9: NBM fitting of the measured HSL-dependent output in all the training set circuits without Monitor cassette.**

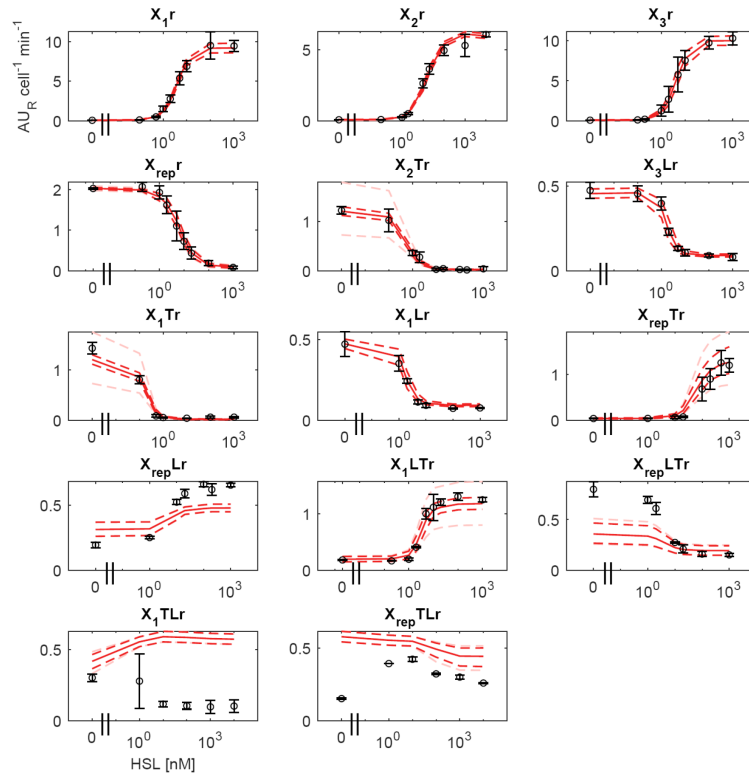
Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.

### B.3. Additional fittings and sensitivity analysis



**Figure B.10: Comparison of the two NOT gate characteristics as a function of the predicted per cell concentration of the TetR or LacI repressor in presence of Monitor module.** Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis. The TetR and LacI values in the x-axes were computed as the nominal (i.e., without Monte Carlo approach) values predicted by the model. The data points showed for TetR or LacI=0 were obtained by measuring the output of constructs similar to  $X_2\text{Tr}(g)$  and  $X_3\text{Lr}(g)$  but without their input block, and were used in the fitting procedure. This genetic context enables to measure the activity of  $P_{LtetO1}$  and  $P_{LlacO1}$  in absence of their cognate repressor, which cannot be removed in the  $X_2\text{Tr}(g)$  and  $X_3\text{Lr}(g)$  circuits due to the basic activity of promoters in the  $X_2$  and  $X_3$  input blocks.

## B. Supplementary information for Chapter 2



**Figure B.11: Univariate sensitivity analysis of the NBM by applying a variation on the  $\delta$  parameter of the Hill functions.**

Panels show fitting ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$ ,  $X_{rep r}$ ,  $X_{2Tr}$  and  $X_{3Lr}$  circuits) and predictions (remaining circuits) of the measured HSL-dependent output in all the training and test set circuits without Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after univariate sensitivity analysis.

### B.3. Additional fittings and sensitivity analysis

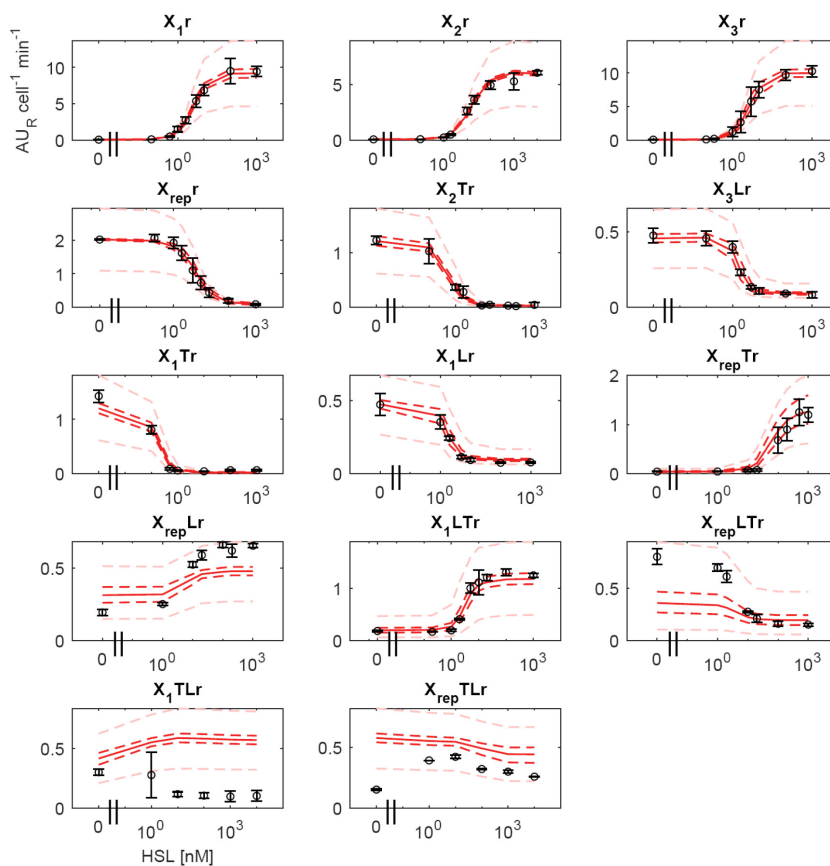
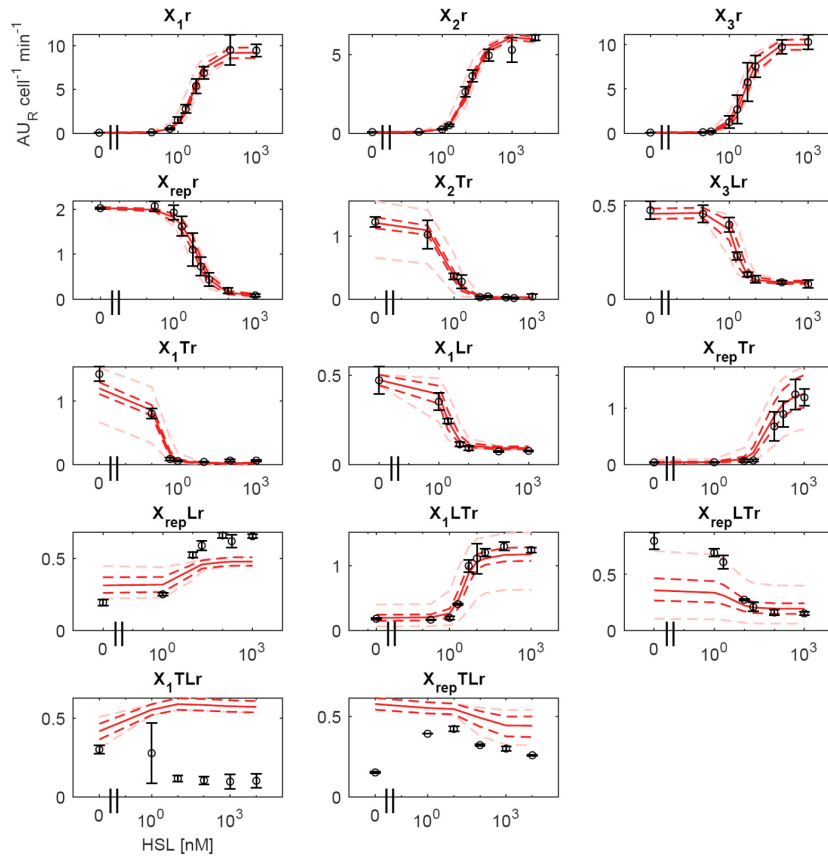


Figure B.12: **Univariate sensitivity analysis of the NBM by applying a variation on the  $\alpha$  parameter of the Hill functions.** Panels show fitting ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$ ,  $X_{rep r}$ ,  $X_{2Tr}$  and  $X_{3Lr}$  circuits) and predictions (remaining circuits) of the measured HSL-dependent output in all the training and test set circuits without Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after univariate sensitivity analysis.

## B. Supplementary information for Chapter 2



**Figure B.13: Univariate sensitivity analysis of the NBM by applying a variation on the  $K$  parameter of the Hill functions.**

Panels show fitting ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$ ,  $X_{rep r}$ ,  $X_{2Tr}$  and  $X_{3Lr}$  circuits) and predictions (remaining circuits) of the measured HSL-dependent output in all the training and test set circuits without Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after univariate sensitivity analysis.

### B.3. Additional fittings and sensitivity analysis

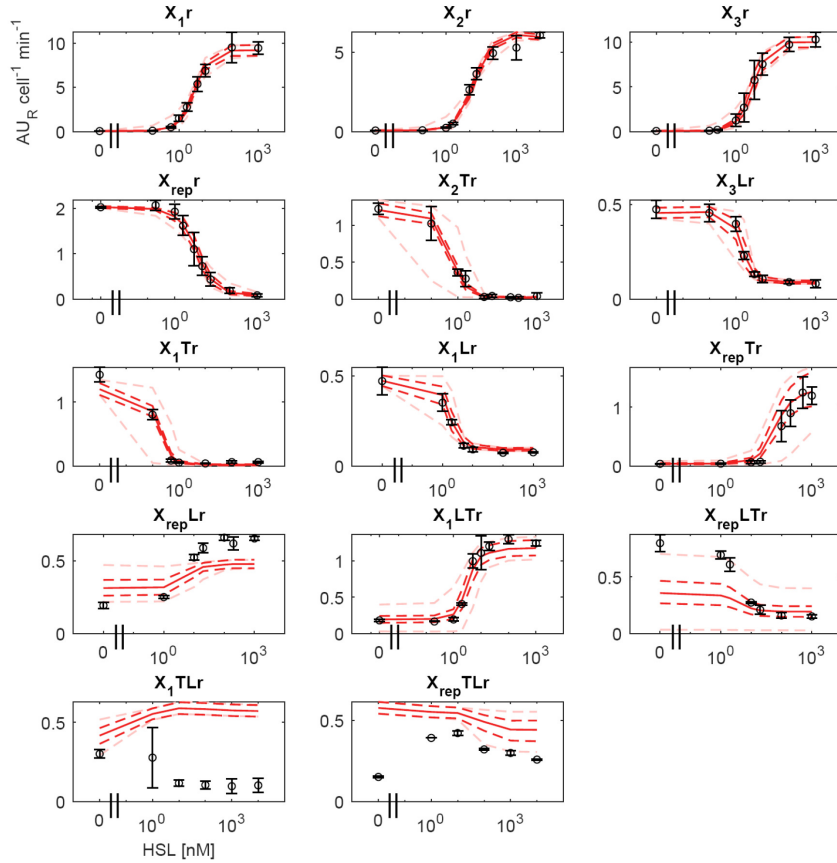


Figure B.14: **Univariate sensitivity analysis of the NBM by applying a variation on the  $\eta$  parameter of the Hill functions.**

Panels show fitting ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$ ,  $X_{rep r}$ ,  $X_{2Tr}$  and  $X_{3Lr}$  circuits) and predictions (remaining circuits) of the measured HSL-dependent output in all the training and test set circuits without Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after univariate sensitivity analysis.

## B. Supplementary information for Chapter 2

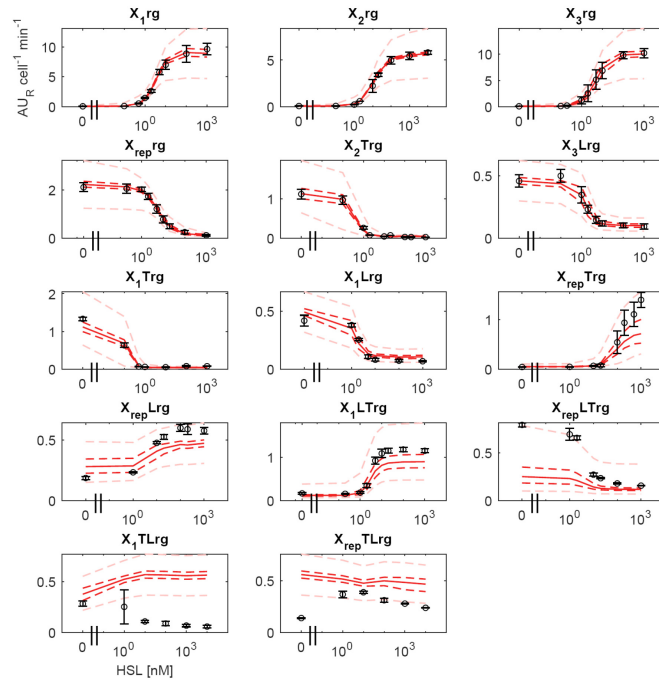
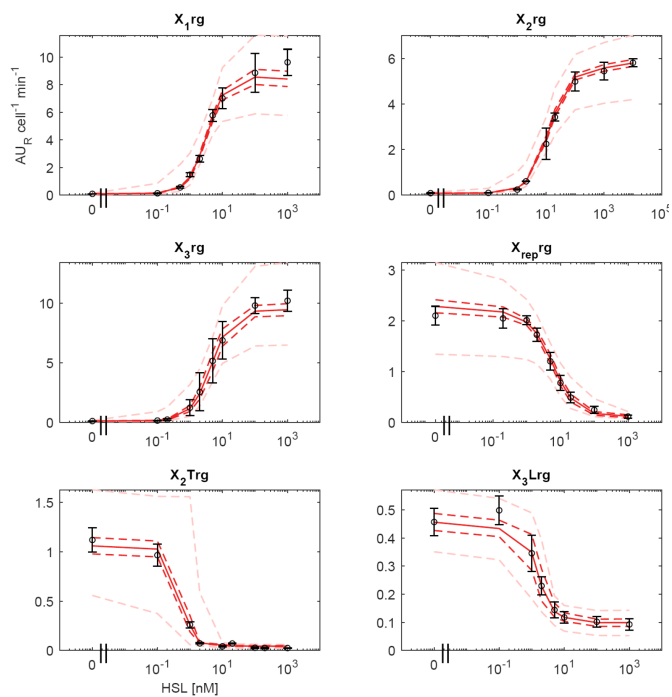


Figure B.15: **Fitting and prediction results for the NBM learned and simulated against RFP data of the circuits with the Monitor cassette.** Panels show fitting ( $X_{1r}$ ,  $X_{2r}$ ,  $X_{3r}$ ,  $X_{rep r}$ ,  $X_{2Tr}$  and  $X_{3Lr}$  circuits) and predictions (remaining ones) of the measured HSL-dependent output in all the training and test set circuits. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.



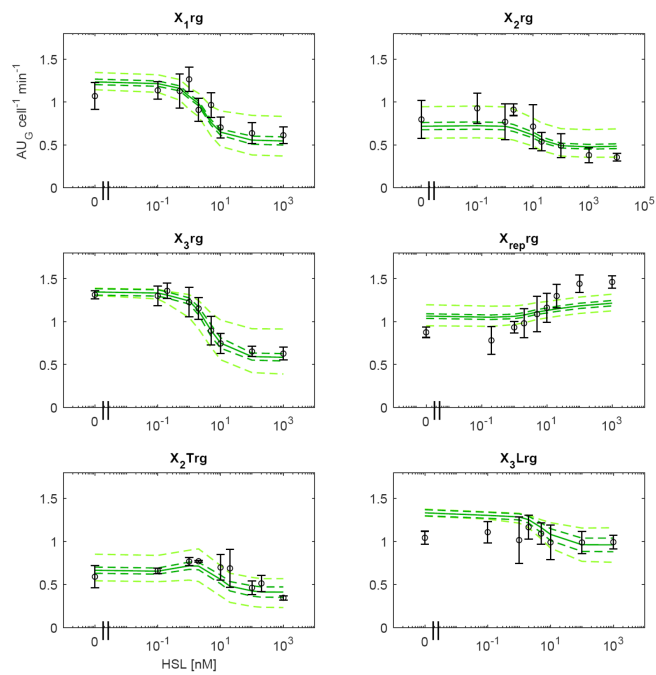
### B.3. Additional fittings and sensitivity analysis



**Figure B.16: BM fitting of the measured HSL-dependent RFP output in all the training set circuits with Monitor cassette.**

Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark red lines are the 95% confidence bands of the output distribution. Dashed light red lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.

## B. Supplementary information for Chapter 2



**Figure B.17: BM fitting of the measured HSL-dependent GFP output in all the training set circuits with Monitor cassette.**

Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid line represents the median predicted output of the model calculated via Monte Carlo simulations for each HSL concentration tested. Dashed dark green lines are the 95% confidence bands of the output distribution. Dashed light green lines are the 95% confidence bands of the output distribution calculated after multivariate sensitivity analysis.

### B.3. Additional fittings and sensitivity analysis

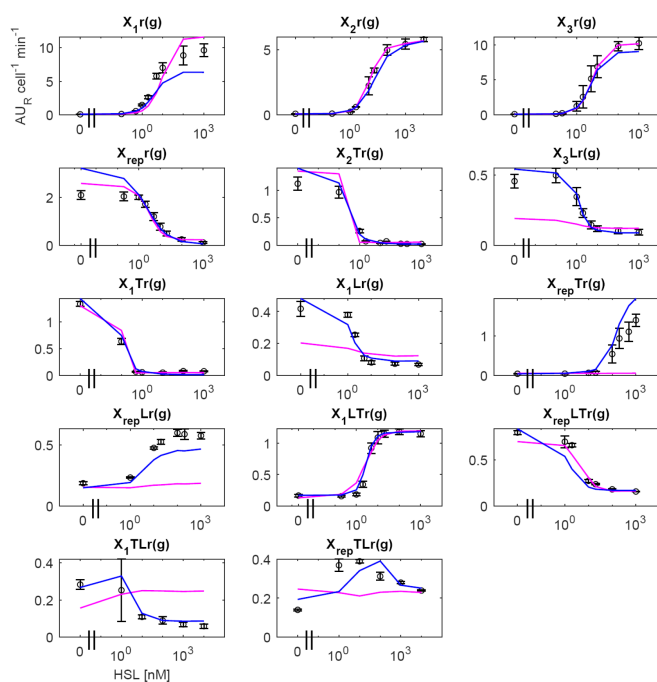
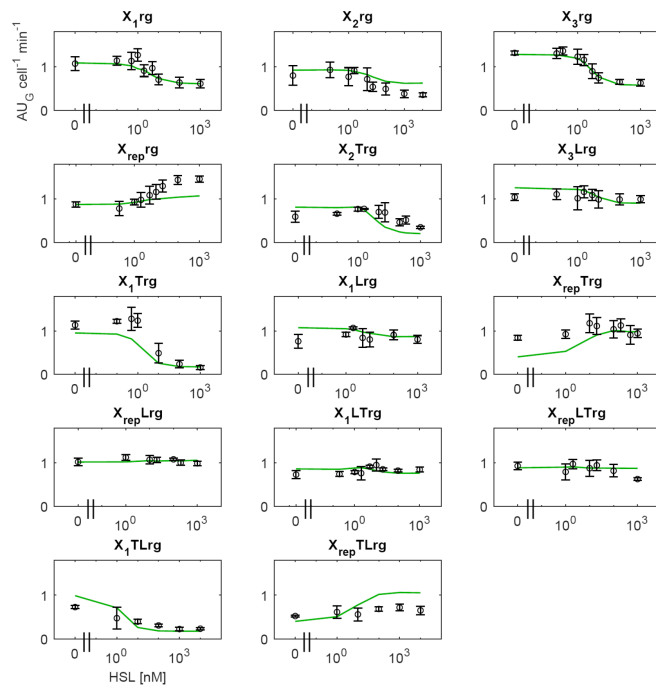


Figure B.18: **Results of fitting using all the available data (training and test set) using NBM and BM: RFP data.** Fitting of the measured HSL-dependent RFP output in all the circuits with Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid lines represent the median predicted output of the NBM (magenta) and BM (blue) calculated via Monte Carlo simulations for each HSL concentration tested.

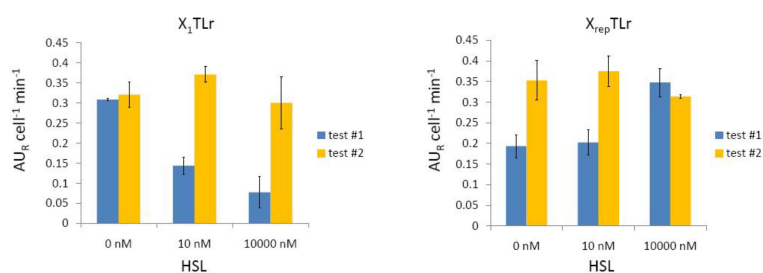
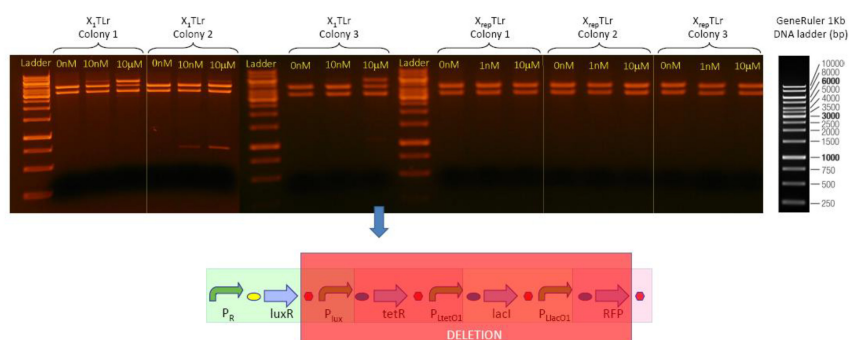
## B. Supplementary information for Chapter 2



**Figure B.19: Results of fitting using all the available data (training and test set) using BM: GFP data.**

Fitting of the measured HSL-dependent GFP output in all the circuits with Monitor cassette. Circles represent the average measured value and error bars represent the 95% confidence intervals of the mean. Solid lines represent the median predicted output of the BM calculated via Monte Carlo simulations for each HSL concentration tested.

## B.4 Evolutionary stability analysis

(a) *RFP restoring test.*(b) *Electrophoresis.*

### Figure B.20: Evolutionary stability of the $X_1\text{TLr}$ and $X_{rep}\text{TLr}$ circuits.

**a)** Phenotypic stability of strains with  $X_1\text{TLr}$  and  $X_{rep}\text{TLr}$ . Strains were tested at three different HSL concentrations (test#1), reported in the x-axis, and then re-inoculated and tested in a growth medium without HSL ( $X_1\text{TLr}$ ) or with 10,000 nM of HSL ( $X_{rep}\text{TLr}$ ) (test#2). In this experiment, we evaluated if the strains could restore the RFP output observed at zero ( $X_1\text{TLr}$ ) or full induction ( $X_{rep}\text{TLr}$ ), corresponding to conditions in which the expression of TetR is repressed, after an experiment carried out at different HSL concentrations. Data points represent the mean of three biological replicates and error bars represent the 95% confidence intervals of the mean. **b)** Genetic stability of the two strains. Electrophoresis results (ethidium bromide staining) are shown for all the tested strains and HSL concentrations after the experiment above. A description of the GeneRuler 1Kb DNA ladder (Thermo Scientific) is also provided (adapted from the user guide of product #SM0312, Thermo Scientific). The mutation found in the second replicate of  $X_1\text{TLr}$  (previously tested with 10,000 nM of HSL) is reported.



## B. Supplementary information for Chapter 2

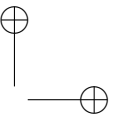
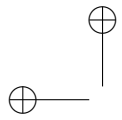
---

**Protocol for Panel a.** Cultures were tested in microplate reader as previously described, with HSL concentrations of 0, 10 and 10,000  $nM$  ( $X_1$ TLr) or 0, 1 and 10,000  $nM$  ( $X_{rep}$ TLr). At the end of the test (18-h growth in microplate reader), all the  $X_1$ TLr cultures were centrifuged, the supernatant was removed and the pellet was resuspended with 200  $\mu l$  of fresh selective medium without HSL. This washing step was performed to remove HSL from the induced  $X_1$ TLr cultures. Five hundred  $\mu l$  of M9 were inoculated with 5  $\mu l$  of the  $X_1$ TLr (washed) or  $X_{rep}$ TLr cultures in 2 –  $ml$  tubes. HSL (final concentration of 10,000  $nM$ ) was added to the  $X_{rep}$ TLr cultures. All the cultures were incubated overnight at 37°C, 220 rpm, and then they were tested again in the microplate reader in absence of HSL ( $X_1$ TLr) or with 10,000  $nM$  of HSL ( $X_{rep}$ TLr).

**Results from Panel a.** All the strains showed stable behaviour, since the RFP output in test#2 is comparable to the RFP output in test#1 for HSL= 0 ( $X_1$ TLr) or 10,000  $nM$  ( $X_{rep}$ TLr), suggesting that HSL-dependent RFP changes in test#1 were not due to stability mutants.

**Protocol for Panel b.** In parallel with the inoculation of the 500 –  $\mu l$  cultures, 2  $\mu l$  of the  $X_1$ TLr (washed) or  $X_{rep}$ TLr cultures were used to inoculate 10  $ml$  of selective medium in 50 –  $ml$  tubes. Cultures were grown overnight at 37°C, 220 rpm. Plasmid DNA was purified, digested with EcoRI-PstI and ran on 1% agarose gel. The second and first biological replicates of  $X_1$ TLr and  $X_{rep}$ TLr, respectively, were sequenced with primers C0062VF (5'-GAATGTTTAGCGTGGGCATG-3') and VR (5'-ATTACCGCCTTTGAGTGAGC-3').

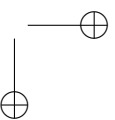
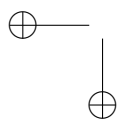
**Results from Panel b.** From electrophoresis screening, all the constructs showed the correct bands, corresponding to vector back-



---

#### B.4. Evolutionary stability analysis

bone ( $3.2Kbp$ ) and insert ( $4.1Kbp$ ); however,  $X_1TLr$  cultures grown in test#1 with  $HSL = 10,000$  and  $10nM$  also showed bands of unexpected size. From sequencing results, only the  $X_1TLr$  culture grown in test#1 with  $10,000nM$  of HSL showed DNA alterations, in a small portion of the population (according to the chromatogram), while the other sequenced plasmids did not show detectable mutations with the used primers. The observed mutation was a deletion of all the circuit after the *luxR* gene and before the transcriptional terminator of the RFP gene.



## B.5 Degradation tag sensitivity analysis

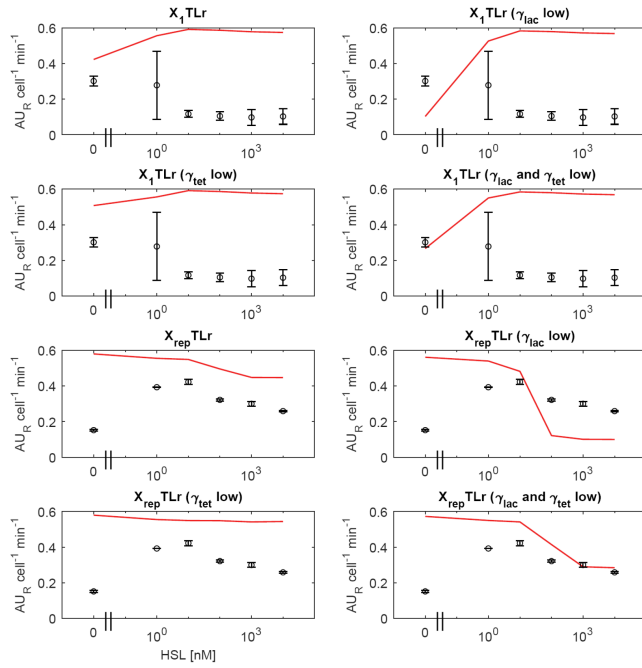
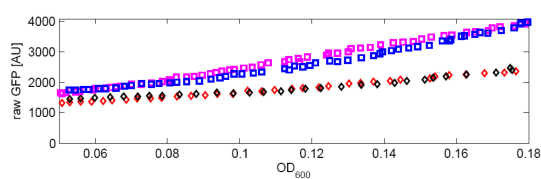


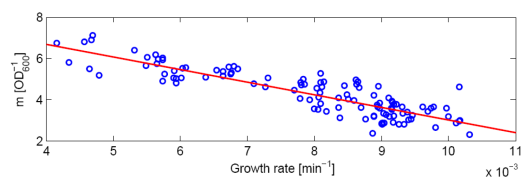
Figure B.21: **Simulation of  $X_1\text{TLr}$  and  $X_{rep}\text{TLr}$  with the NBM for different values of  $\lambda_{tet}$  and  $\lambda_{lac}$  parameters.** Data are reported (circles and error bars represent the 95% confidence intervals of the mean) and the simulated RFP output is shown (solid line). The parameters of the NBM reported in Table B.1 were used for the simulations. The  $\lambda_{tet}$  and  $\lambda_{lac}$  parameter values were set at the nominal ones (see Section 2.1) or they were decreased by 100-fold (marked as *low*) to qualitatively evaluate the effect of enzymatic queuing, which might cause slower TetR and LacI degradation, since they share the same LVA tag.



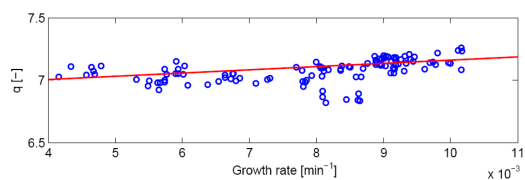
## B.6 GFP autofluorescence analysis



(a)



(b)



(c)

**Figure B.22: Raw GFP autofluorescence dependence on  $OD_{600}$  and cell growth rate.** **a)**  $OD_{600}$  vs raw GFP characteristic of two strains showing fast (relative to the whole circuit collection) growth ( $X_2Tr$  and  $X_{rep}r$ ; black and red diamonds, respectively) and two strains showing slow growth ( $X_{rep}Tr$  and  $X_{rep}TLr$ ; magenta and blue squares, respectively) without HSL. They are reported, as single biological replicate, as an example to highlight distinct characteristics, dependently on growth rate. Curves like the ones shown here were fitted (with exponential regression), estimating  $m$  and  $q$  parameters (see Section 2.1) to obtain the autofluorescence background of the circuits. **b-c)** The  $m$  and  $q$  parameters for all the circuits in all the conditions are plotted against growth rate values. The  $m$  parameter (in **b**) shows a significant growth rate-dependent trend, as expected from Panel **a**, while the growth rate-dependent trend of the  $q$  parameter **c** is not statistically significant (confidence intervals of the slope include zero).

## B.7 Tables of parts and parameters data

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

<b>Name</b>	<b>BioBrick code</b>
$P_{lux}$	BBa_R0062
$P_{luxRep}$	BBa_J107100
$P_R$	BBa_R0051
$P_{LlacO1}$	BBa_R0011
$P_{LtetO1}$	BBa_R0040
Constitutive promoter of the Monitor cassette	BBa_J23100
Reference constitutive promoter	BBa_J23101
Strong RBS	BBa_B0030
Weak RBS	BBa_B0031
Medium RBS used in the Monitor cassette	BBa_B0032
Weak RBS used to decrease tetR translation	BBa_B0033
Strong RBS	BBa_B0034
Double transcriptional terminator	BBa_B0015
Synthetic transcriptional terminator	BBa_B1006
LuxR coding sequence	BBa_C0062
LacI coding sequence	BBa_C0012
TetR coding sequence	BBa_C0040
mRFP1 coding sequence	BBa_E1010
GFPmut3b coding sequence	BBa_E0040

## B.7. Tables of parts and parameters data

---

**Table B.2: Statistics on circuits bearing the burden monitor.**  
 Statistically significant correlation coefficients are reported in bold. Correlation coefficients with opposite sign than expected are reported as  $< 0$  and  $> 0$  for growth rate vs GFP and RFP vs GFP, respectively.

<b>Circuit</b>	<b>Growth rate vs GFP</b>	<b>RFP vs GFP</b>
$X_1rg$	$< 0$	$-0.68$
$X_2rg$	$0.02$	$-0.8$
$X_3rg$	$< 0$	$-0.9$
$X_{rep}rg$	$< 0$	$-0.8$
$X_2Trg$	$0.41$	$> 0$
$X_3Lrg$	$< 0$	$> 0$
$X_1Lrg$	$< 0$	$> 0$
$X_1Trg$	$0.72$	$> 0$
$X_{rep}Lrg$	$< 0$	$-0.23$
$X_{rep}Trg$	$< 0$	$> 0$
$X_1LTrg$	$< 0$	$> 0$
$X_1TLrg$	$0.84$	$> 0$
$X_{rep}LTrg$	$< 0$	$> 0$
$X_{rep}TLrg$	$< 0$	$> 0$

## B.8 Evidence of metabolic burden using a different monitor cassette

To confirm that the decrease of monitor output was not due to crosstalk or unwanted gene regulations, we tested a further control circuit with a different monitor system, in which both promoter and reporter were varied, compared with the J23100-GFP monitor. The test construct was composed by two modules born on the same low copy plasmid and tested in the same conditions as the other circuits:

- An  $X_1$  HSL-inducible expression cassette (Figure 2.1) driving the expression of TetR protein (*loading module*);
- A  $PLlacO1$  promoter used to constitutively express RFP (*monitor module*)

A decrease in the RFP signal for increasing values of HSL induction - proportional to TetR protein synthesis - was still observable (see Figure B.23).

Hence, the metabolic burden due to over-expression of a high-demanding protein synthesis module arose again despite the use of a different monitor system, reducing the probability of biases in the study given from further hidden interactions between the parts used.

## B.8. Evidence of metabolic burden using a different monitor cassette

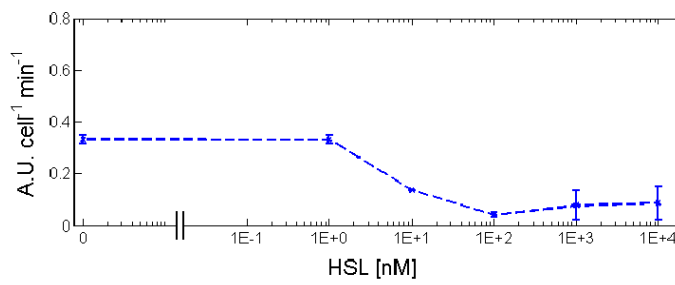
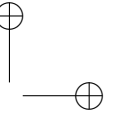
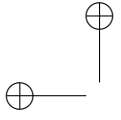
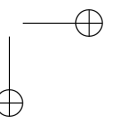
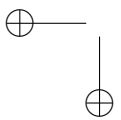


Figure B.23: **Effect of metabolic burden with a different monitor system.** Variation in constitutive RFP synthesis rate (on y axes) for increasing TetR expression levels proportional to HSL concentration (on x axes). Stars represent the average measured value and error bars represent the 95% confidence intervals of the mean.



B. Supplementary information for Chapter 2

---



# Appendix **C**

## Supplementary information for Chapter 3

### C.1 List of constructs and estimated parameters

Table C.1: **Table of constructs.** The table shows the BioBrick<sup>TM</sup> codes for the constructs built in the study following the proper order of parts. Those in bold were conceived, standardized via PCR amplification and deposited in the Registry of Standard Biological Parts by ourselves during this work.

Name	Construct	Purpose
dCas9	<b>BBa_J107200</b>	Nuclease null Sp.Cas9 protein from Addgene plasmid #44249
tracr	<b>BBa_J107201</b>	tracr RNA with tetraloop for Sp.dCas9 from Addgene plasmid #44251
A37	BBa_J23100 + BBa_B0032 + Bba_E0040 + BBa_B0015 + pSB4C5	GFP-based metabolic burden monitor

*Continue in the next page ...*

## C. Supplementary information for Chapter 3

... continue from the previous page

Name	Construct	Purpose
AEdCas9	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_R0062 + BBa_J107200 + pSB4C5	LuxR-driven dCas9 inducible cassette for toxicity measurement
M12_Inv	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_R0062 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB4C5	LuxR-driven RFP inducible cassette for toxicity comparison with monitor
AEd116gPtet	AEdCas9 + BBa_J23116 + gPtet + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on tetR target promoter - low sgRNA
AEd100gPtet	AEdCas9 + BBa_J23100 + gPtet + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on tetR target promoter - medium sgRNA
AEd119gPtet	AEdCas9 + BBa_J23119 + gPtet + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on tetR target promoter - high sgRNA
AEd116gPlac	AEdCas9 + BBa_J23116 + gPlac + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on lacI target promoter - low sgRNA
AEd100gPlac	AEdCas9 + BBa_J23100 + gPlac + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on lacI target promoter - medium sgRNA
AEd119gPlac	AEdCas9 + BBa_J23119 + gPlac + BBa_J107201 + pSB4C5	dCas9 efficiency characterization on lacI target promoter - high sgRNA
E62	BBa_R0040 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	tetR target promoter in medium copy
I13521	BBa_R0040 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	tetR target promoter in high copy
E52	BBa_R0011 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	LacI target promoter in medium copy
A33	BBa_R0011 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	LacI target promoter in high copy
J116dCas	BBa_J23116 + BBa_J107200 + pSB3K3	Constitutive dCas9 expression cassette in medium copy
E20	BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_R0062 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB4C5	Template for optimized Lux-inducible guide expression cassette
AETracr	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_J107202 + BBa_B0034 + BBa_E1010 + BBa_B0015 + BBa_J107201 + pSB4C5	Template for guide mutagenesis in LuxR expression cassette

Continue in the next page ...



## C.1. List of constructs and estimated parameters

*... continue from the previous page*

Name	Construct	Purpose
AE-3A	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + <b>BBa_J107202</b> + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB4C5	Optimized LuxR-inducible expression cassette characterization
AYA	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + BBa_R0011 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB4C5	LacI-inducible expression cassette characterization
AYAtracr	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + BBa_R0011 + BBa_B0034 + BBa_E1010 + BBa_B0015 + BBa_J107201 + pSB4C5	Template for guide mutagenesis in LacI-inducible expression cassette
AY-A	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + <b>BBa_J107203</b> + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB4C5	Mutagenized LacI-inducible guide expression cassette characterization
AEgPtet	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_J107202 + gPtet + BBa_J107201 + pSB4C5	LuxR-inducible guide cassette for repression characterization
AEgPlac	A37 + BBa_R0051 + BBa_B0030 + BBa_C0062 + BBa_B1006 + BBa_J107202 + gPlac + BBa_J107201 + pSB4C5	LuxR-inducible guide cassette for repression characterization
AYgPtet	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + BBa_R0011 + gPtet + BBa_J107201 + pSB4C5	LacI-inducible guide cassette for repression characterization
AY-AgPtet	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + <b>BBa_J107202</b> + gPtet + BBa_J107201 + pSB4C5	Mutagenized LacI-inducible tet guide for repression characterization
AYgPluxH	A37 + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + BBa_R0011 + gPluxH + BBa_J107201 + pSB4C5	LacI-inducible guide cassette for repression characterization
dCasE62	BBa_J23116 + BBa_J107200 + BBa_R0040 + BBa_B0034 + BBa_E1010 + BBa_B0015 + Psb3k3	Constitutive dCas9 and tet target promoter in medium copy
dCasE52	BBa_J23116 + BBa_J107200 + BBa_R0011 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	Constitutive dCas9 and lac target promoter in medium copy
dCasJ119H	BBa_J23116 + BBa_J107200 + BBa_J107100 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	Constitutive dCas9 and $P_{luxRep}$ target promoter <i>J119H</i> or $P_{luxRep}$ in medium copy
dCasP122	BBa_J23116 + BBa_J107200 + BBa_J107111 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	Constitutive dCas9 and $P_{luxRep}$ target promoter #122 in medium copy

*Continue in the next page ...*

### C. Supplementary information for Chapter 3

... continue from the previous page

Name	Construct	Purpose
dCasP2	BBa_J23116 + BBa_J107200 + BBa_J107101 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	Constitutive dCas9 and $P_{luxRep}$ target promoter #2 in medium copy
dCasP44	BBa_J23116 + BBa_J107200 + BBa_J107105 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB3K3	Constitutive dCas9 and $P_{luxRep}$ target promoter #44 in medium copy
J119H	BBa_J107100 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	$P_{luxRep}$ target promoter <i>J119H</i> in high copy
P122	BBa_J107111 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	$P_{luxRep}$ target promoter #122 in high copy
P2	BBa_J107101 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	$P_{luxRep}$ target promoter #2 in high copy
P44	BBa_J107105 + BBa_B0034 + BBa_E1010 + BBa_B0015 + pSB1A2	$P_{luxRep}$ target promoter #44 in high copy
AYgPluxH22	AYgPluxH + Bba_R0051 + Bba_B0030 + Bba_C0062 + Bba_B0015 + pSB4C5	LuxR competition experiments
AEgPlacY35	AEgPlac + BBa_J23118 + BBa_B0034 + BBa_C0012 + BBa_B0015 + pSB4C5	LacI competition experiments

... end of the table.

**Table C.2: CRISPRi model parameters.** **G(N)B** stands for Global fitting using (No) Burden model, **F(N)B** for two step fitting using (No) Burden model and **P(N)B** for two step fitting with predictions using (No) Burden model. In **blue**, constants or analytically derived parameters; when covariance matrices presented singularities or parameters were estimated as very low values, Not computed (N.c.) was reported.

Name	GNB	(CV%)	GB	(CV%)	FNB	(CV%)	FB	(CV%)	PNB	(CV%)	PB	(CV%)
<i><math>\alpha_{GFP}</math></i>	0.046	(N.a.)	0.046	(N.a.)	0.046	(N.a.)	0.046	(N.a.)	0.046	(N.a.)	0.046	(N.a.)
<i><math>\alpha_{RFP}</math></i>	0.017	(N.a.)	0.017	(N.a.)	0.017	(N.a.)	0.017	(N.a.)	0.017	(N.a.)	0.017	(N.a.)
$\alpha_{A33}$	23.13	(30.8 %)	89.15	(6.2 %)	23.71	(N.c.)	89.15	(10.2 %)	23.71	(N.c.)	89.15	(10.1 %)
$\alpha_{E52}$	19.29	(3.2 %)	43.82	(5.0 %)	19.28	(N.c.)	43.82	(6.8 %)	19.28	(N.c.)	43.82	(6.9 %)
$\alpha_{E62}$	13.25	(0.9 %)	19.85	(1.8 %)	13.27	(N.c.)	19.85	(3.4 %)	14.01	(N.c.)	19.73	(7.3 %)
$\alpha_{I13521}$	21.55	(3.0 %)	50.03	(3.8 %)	21.88	(N.c.)	50.03	(5.5 %)	23.12	(N.c.)	46.97	(8.5 %)
$\alpha_{PLlacO1}$	5.88	(3.5 %)	5.86	(4.6 %)	5.67	(3.2 %)	5.86	(4.6 %)	5.67	(3.2 %)	5.86	(4.6 %)
$\alpha_{P_{lux}-3a}$	3.67	(1.9 %)	4.44	(2.6 %)	3.61	(1.7 %)	4.44	(4.0 %)	3.61	(1.7 %)	4.44	(4.0 %)
$\delta_{A33}$	1.09	(656.3 %)	0.79	(110.8 %)	0.48	(N.c.)	0.79	(876.9 %)	0.48	(N.c.)	0.79	(882.3 %)
$\delta_{E52}$	0.19	(107.5 %)	0.23	(87.0 %)	0.19	(N.c.)	0.23	(244.3 %)	0.19	(N.c.)	0.23	(289.0 %)
$\delta_{E62}$	2E-02	(178.5 %)	8E-03	(115.8 %)	8E-03	(N.c.)	8E-03	(2614.5 %)	2E-14	(N.c.)	2E-14	(N.a.)
$\delta_{I13521}$	2E-14	(N.a.)	4E-12	(N.a.)	2E-13	(N.c.)	4E-12	(N.a.)	8E-13	(N.c.)	6E-12	(N.a.)
$\delta_{PLlacO1}$	9E-04	(40.3 %)	7E-03	(30.1 %)	8E-03	(22.6 %)	7E-03	(30.6 %)	8E-03	(22.6 %)	7E-03	(30.6 %)
$\delta_{P_{lux}-3a}$	5E-03	(39.4 %)	3E-02	(29.8 %)	3E-02	(9.8 %)	3E-02	(11.3 %)	3E-02	(9.8 %)	3E-02	(11.3 %)
$\eta_{PLlacO1}$	4.10	(8.6 %)	3.84	(7.2 %)	3.72	(14.5 %)	3.84	(19.6 %)	3.72	(14.5 %)	3.84	(19.6 %)
$\eta_{P_{lux}-3a}$	1.32	(3.3 %)	1.52	(3.7 %)	1.52	(3.6 %)	1.52	(4.3 %)	1.52	(3.6 %)	1.52	(4.3 %)
$k_{A33}$	0.10	(119.4 %)	0.12	(38.2 %)	0.17	(N.c.)	0.12	(60.8 %)	0.17	(N.c.)	0.12	(61.0 %)
$k_{E52}$	5E-03	(63.8 %)	3E-03	(56.3 %)	5E-03	(N.c.)	3E-03	(226.2 %)	5E-03	(N.c.)	3E-03	(226.1 %)
$k_{E62}$	2E-03	(41.9 %)	3E-03	(29.4 %)	3E-03	(N.c.)	3E-03	(44.5 %)	8E-03	(N.c.)	5E-03	(74.4 %)
$k_{I13521}$	2E-02	(42.9 %)	1E-02	(29.9 %)	3E-02	(N.c.)	1E-02	(17.3 %)	2E-02	(N.c.)	1E-02	(35.5 %)
$k_{iPLlacO1}$	2.85	(4.7 %)	2.69	(5.9 %)	2.73	(4.9 %)	2.69	(6.1 %)	2.73	(4.9 %)	2.69	(6.1 %)
$k_{iP_{lux}-3a}$	1.53	(4.4 %)	1.58	(5.2 %)	1.50	(3.6 %)	1.58	(4.3 %)	1.50	(3.6 %)	1.58	(4.3 %)
$S_x$	-	(N.a.)	1.70	(N.a.)	-	(N.a.)	1.70	(N.a.)	-	(N.a.)	1.70	(N.a.)
$S_m$	-	(N.a.)	1.47	(N.a.)	-	(N.a.)	1.47	(N.a.)	-	(N.a.)	1.47	(N.a.)
$\Sigma_{LacI}$	-	(N.a.)	0.00	(N.a.)	-	(N.a.)	9E-14	(N.a. %)	-	(N.a.)	9E-14	(N.a.)
$\Sigma_{LacR}$	-	(N.a.)	0.14	(3.4 %)	-	(N.a.)	0.14	(12.1 %)	-	(N.a.)	0.14	(12.1 %)
$\Sigma_{HC}$	-	(N.a.)	0.12	(N.a.)	-	(N.a.)	0.12	(N.a.)	-	(N.a.)	0.12	(N.a.)
$J_{RFP}$	-	(N.a.)	0.02	(N.a.)	-	(N.a.)	0.02	(N.a.)	-	(N.a.)	0.02	(N.a.)

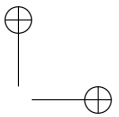
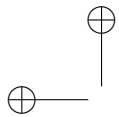
## C.2 Materials and reagents

### Inducers

- ***N*-(3-Oxohehexanoyl)-L-homoserine lactone (HSL)**: purchased from Sigma Aldrich (K3007), it binds and activates LuxR protein thus affecting Lux promoter transcriptional activity. HSL is dissolved in deionized water and conserved at  $-20^{\circ}\text{C}$  at  $200\text{mM}$ . Its small size allows it to freely diffuse across the plasma membrane;
- ***Isopropyl*  $\beta$ -D-thiogalactopyranoside (IPTG)**: purchased from Sigma Aldrich (I1284), ready made solution of concentration  $2\text{mM}$  conserved at  $-20^{\circ}\text{C}$ . It is a non-metabolizable analog of galactose, able to bind LacI, thus relieving repression on the lac promoter. It is different from lactose in inducing expression of  $P_{LlacO1}$  because it is also capable of diffusing freely across the lipid bilayer at a basal level that permits a graded response from a bacterial population.

### Antibiotics

- **Chloramphenicol (Cm)**: Resistance to this antibiotic was used as a marker for selection of all bacteria transformed with circuits implemented on the low copy plasmid vector pSB4C5 (LC). Conserved at  $-20^{\circ}\text{C}$   $34\text{mg/ml}$  and used at a final concentration of  $12.5\mu\text{g/ml}$ ;
- **Kanamycin (Kan)**: Resistance to this antibiotic was used as a selection marker for constructs in the medium copy plasmid vector pSB3K3 (MC). Conserved at  $-20^{\circ}\text{C}$   $50\text{mg/ml}$  and used at a final concentration of  $25\mu\text{g/ml}$ ;



---

## C.3. Cloning

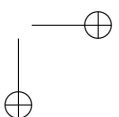
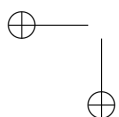
- **Ampicillin (Amp):** Resistance to this antibiotic was used as a selection marker for constructs in the high copy plasmid vector pSB1A2 (HC). Conserved at  $-20^{\circ}\text{C}$   $100\text{mg/ml}$  and used at concentration  $100\mu\text{g/ml}$ .

## C.3 Cloning

### Mutagenesis

Mutagenesis with divergent primers was adopted to customize gRNAs and to delete nucleotides after the transcription start sites of the used promoters, when indicated. For gRNA sequence insertion, tailed 40 nucleotide primers were used such that 20 nucleotides composed the gRNA sequence and the other 20 nucleotides annealed the 5' end of non annealing part of the sgRNAs. Deletion mutations were obtained via amplification of the desired part of the plasmid, excluding the nucleotides that needed to be deleted. The experimental protocol was as follows:

- Template plasmid DNA was purified as in B.1; the Phusion Hot Start Flex II was used according to manufacturer protocol; primer pairs used in the study are listed in Table C.3.
- The PCR cycle was run and followed by Dpn1 (Roche) digestion of the methylated template DNA; primer annealing temperature was calculated on the free online tool offered by New England Biolabs with parameters set as the standard for Phusion polymerases;
- PCR products were separated in a 1% agarose gel and extracted and purified as in B.1;



### C. Supplementary information for Chapter 3

- Blunt-end DNA fragments were phosphorylated by Polynucleotide Kinase (PNK - New England Biolabs) and ligated by T4 ligase; a  $20\mu L$  reaction mix was composed of:
  - A maximum of  $50ng$  of DNA brought to a volume of  $17\mu L$  with deionized water;
  - $2\mu L$  of ligase buffer;
  - $1\mu L$  of PNK;

The reaction was allowed to proceed at  $37^\circ C$  for 20 minutes, then  $1\mu L$  of ligase was added to the reaction mix and incubated for  $16h$  at  $16^\circ C$ ;

Finally, PNK and T4 Ligase were deactivated at  $75^\circ C$  for 10 minutes, and the same protocol for transformation and stock preparation described in B.1 was employed.

Table C.3: **Table of primers.** Primers used in the study are listed and their purpose is stated.

Template	Primer Pair	Sequence (5' - 3')	Aim	Product
Addgene	FW_ribcaterm_XbaI	TCGCGGCCGCTTCTA GAGAAAGAGGAGAAA GGATCTATGG	Conversion of dCas9 and its RBS into the BioBrick <sup>TM</sup> format	BBa_J107200
#44249	RV_dCas9_SP	CGGTTTCTTCCTGCA GCGGCCGCTACTAGT ATATAAACFCAGAAA GGCCCA		
Addgene	FW_tracr_XbaI	TCGCGGCCGCTTCTA GAGGTTTTAGAGCTA GAAATAGCAAG	Conversion of tracrRNA and its terminator into the BioBrick <sup>TM</sup> format	BBa_J107201
#44251	RV_tracr_suffix	GTTTCTTCCTGCAGC GGCCGCTACTAGTAA GTTACCCGACAAACA AC		
E20	FW_RFP-34	TACTAGAGAAAGAGG AGAAATACTAGATGG CTTCCTCCGAAG	Remove three Adenine nucleotides located after the TSS	E20-3A

*Continue in the next page ...*

### C.3. Cloning

... continue from the previous page

Template	Primer Pair	Sequence (5' - 3')	Aim	Product
	RV_Plux-3A	ATTCGACTATAACAA ACCATTTTCTTGCGT AAACCTGTAC		
AEtracr	FW_gPtet_tracr	TGTCAATCTCTATCA CTGATGTTTTAGAGC TAGAAATAGC	Insert Tet guide RNA and remove RFP	AEgPtet
	RV_Plux-3A	ATTCGACTATAACAA ACCATTTTCTTGCGT AAACCTGTAC		
AEtracr	FW_gPlac_tracr	ATAACAATTGACATT GTGAGGTTTTAGAGC TAGAAATAGC	Insert Lac guide RNA and remove RFP	aEgPlac
	RV_gPlux-3A	ATTCGACTATAACAA ACCATTTTCTTGCGT AAACCTGTAC		
AYAtracr	FW_gPtet_tracr	TGTCAATCTCTATCA CTGATGTTTTAGAGC TAGAAATAGC	Insert Tet guide RNA and remove RFP	AYgPtet
	RV_Plac	TGTGCTCAGTATCTT GTTATCCGCTC		
AYAtracr	FW_gPluxH_tracr	TGACACCTGTAGGAT CGTACGTTTTAGAGC TAGAAATAGC	Insert LuxH guide RNA and remove RFP	AYgPluxH
	RV_Plac	TGTGCTCAGTATCT TGTATCCGCTC		
AYA	FW_RFP-34	TACTAGAGAAAGAGG AGAAATACTAGATGG CTTCTCCGAAG	Remove one Adenine nucleotide located after the TSS	AY-A
	RV_Plac-A	GTGCTCAGTATCTT TTATCCGCTCA		
16S RNA	FW_16S	GAATGCCACGGTGAA TACGTT	Amplification of refer- ence gene for RT-qPCR	NA
	RV_16S	CACAAAGTGGTAAGC GCCCT		
RFP	FW_qRFP	GAAAGACGGTGGTCA CTACG	Amplification of target gene for RT-qPCR	NA
	RV_qRFP	TTGTGGGAGGTGATG TCCA		
Vector	VF2	TGCCACCTGACGTCT AAGAA	Sequencing of inserts from extremities in- wards	NA
	VR	ATTACCGCCTTTGAG TGAGC		

Continue in the next page ...

## C. Supplementary information for Chapter 3

*... continue from the previous page*

Template	Primer Pair	Sequence (5' - 3')	Aim	Product
LuxR	VF_C0062	GAATGTTTAGCGTGG GCATG	Sequencing downstream HSL-inducible cassette	NA
LacI	VF_seq_LacI	GCTTGCTGCAACTCT CTCAG	Sequencing downstream IPTG-inducible cassette	NA

*... end of the table.*

### Amplification and standardization of CRISPRi elements

To facilitate all the assemblies carried out in this work and to support the re-use of constructed parts in future works, the two main elements of the CRISPRi system (i.e. dCas9 and tracrRNA) were PCR amplified from Addgene plasmids #44249 and #44251 with the convergent primer pairs FW\_ribcas\_XbaI, RV\_dCas9\_SP and FW\_tracr\_XbaI, RV\_tracr\_suffix, respectively, to convert them into the BioBrick<sup>TM</sup> format. Resulting sequences were digested and ligated in the BioBrick-compliant pSB3K3 plasmid, with the protocol described in section B.1. The two resulting standardized sequences were submitted to the Registry of Standard Biological Parts as BBa\_J107200 and BBa\_J107201.

### sgRNA Design

All guide RNAs were designed on the on-line platform Benchling, using the *S. Pyogenes* PAM sequence NGG. Promoter regions of reporter genes were targeted via CRISPR:dCas9 complex as a means to block transcription initiation and thus reporter gene expression. The variable region of sgRNAs consisted exclusively of 20 base pairs complementary to the promoter region, specifically annealing to at least one nucleotide of the -35 box, while the constant region comprised the gRNA linker and tracrRNA with its own terminator. The list of designed sgRNAs is reported in Table C.4.



Table C.4: **List of gRNAs and relative targets.** The sequences of gRNAs conceived for the repression of target promoters are listed, as well as their respective binding locations on the target promoter. All gRNAs were designed to block the  $-35$  box of the promoter to inhibit polymerase binding. Red nucleotides represent gRNA target, red bold  $-35$ Box, black bold PAM sequence, green  $-10$ Box

Target Promoter	gRNA Sequence	Promoter Sequence
$P_{lac}$	ATAACAATTGACATTGTGAG	+ strand; AATTGTGAGCGGATA <b>ACAA</b> <b>TTGACATTGTGAGCGG</b> ATA ACAAG <b>ATACT</b> GAGCACA
$P_{tet}$	TGTCAATCTCTATCACTGAT	- strand; GTGCTC <b>AGTAT</b> CTCTATCAC TGATAGGGAT <b>TGTC</b> AATCTCT <b>ATCACTGATAGG</b> GA
$P_{luxRep}$	TGACACCTGTAGGATCGTAC	J119H; + strand; <b>TGACACCTGTAGGATCGTAC</b> <b>AGGTATAATGCTAGC</b>
		Promoter #122; + strand; <b>TGACACCTGTAGGATCGTAC</b> <b>AGGTACTATGCTAGC</b>
		Promoter #2 ; + strand; <b>TGACACCTGTAGGATCGTAC</b> <b>AGGTACAGTGCTAGC</b>
		Promoter #44 ; + strand; <b>TGACACCTGTAGGATCGTAC</b> <b>AGGTACTGTGCTAGC</b>



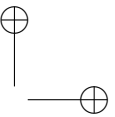
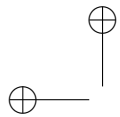
## C. Supplementary information for Chapter 3

---

### C.4 Microscopic Images

The Leica DMLS Type 020-518.500 bright field microscope was used to take images of bacterial cells for morphological analysis. Bacteria were magnified with the 100x/1.25 oil immersion objective, and static pictures were taken using the Nikon COOLPIX 4500 digital camera. For sample preparation, the following protocol was employed:

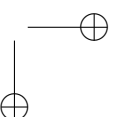
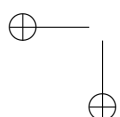
- Long term bacterial stocks were streaked on selective LB agar plates;
- colonies were used to inoculate 500 $\mu$ L of M9 medium supplemented with the appropriate antibiotic(s) - when indicated, HSL at the 100nM concentration was added - and incubated for 16hrs at 220rpm, 37°C;
- 20 $\mu$ L of culture were fixed on a glass slide by heating over a bunsen burner;
- Fixed cells were stained for 90s with a 0.5% Safranin solution diluted in deionized water;
- The staining solution was washed away with an adequate amount of running tap water and left 10 minutes to air dry under a fume hood;
- The slide was mounted with a cover slip with 60 $\mu$ L of EUKITT mounting medium and left to solidify under a fume hood;



## C.5 Flow Cytometry

Single cell analysis of transformed strains was performed on the Partec CyFlow Space flow-cytometer equipped with an argon ion laser. Excitation was carried out via a  $50mW$   $488nm$  laser and fluorescence emission was collected in FL1. At least 100,000 events were collected and stored for each sample. A non-fluorescent TOP10 culture was always included to measure the background fluorescence. Data obtained were analyzed through FloMax software. For sample preparation, the following protocol was followed:

- Glycerol stocks were streaked on LB agar plates supplemented with the appropriate antibiotic(s) and incubated at  $37^{\circ}C$  for  $16hrs$ ;
- Single colonies were picked and inoculated  $1mL$  of selective M9 medium with the IPTG inductions of 0, 1 or  $50\mu M$ , corresponding to the fully repressed, half-activated and fully activated configurations, respectively, and incubated at  $37^{\circ}C$ ,  $220rpm$  for  $16hrs$ ;
- Grown cultures were diluted 1 : 100 in  $2mL$  of selective M9 medium with the same IPTG concentration and allowed to grow for 2 – 3hrs;
- Finally cells were diluted 1 : 5 in sterile *PBS1X* and used for flow-cytometer analysis;



## C.6 Blanking of the auto-fluorescence in the GFP signal

Since auto-fluorescence in the emission spectra of GFP is strongly dependent from  $OD_{600}$  in bacterial cultures growing in M9 media, an instrument-dependent procedure to properly blank the signal has been developed, given the growth rate of a culture  $\mu$  and its  $OD_{600}$  time course.

Previous studies showed a monotonic increasing behavior of the GFP respect to the optical density that can be described via an exponential curve [89]; however, the shape of those curves (i.e. the parameters describing the exponential growth phase) depends also on the growth rate of the culture; in chapter 2, to overcome this dependency, each genetic circuit designed were implemented and tested in 2 different versions, with and without GFP monitor, assuming that the growth rates of the pair was the same. This approach, despite its robustness, is particularly time consuming due to the double amount of efforts in cloning and testing each condition.

Given the general exponential Equation 2.1 describing the auto fluorescence, a relation between the two parameters  $q$  and  $m$  - describing the linear relation between  $\ln(OD_{600})$  and GFP auto-fluorescence level - and the growth rate  $\mu$  of the culture - the most immediate parameter that differs between two genetic circuits implemented in the same host and tested in the same experimental condition - was sought. Towards this, all the non-GFP-expressing circuits tested in chapter 2 were considered, fitting auto-fluorescence signals with the equation C.3 and computing a linear regression between each parameter of the function - $q$  and  $m$  - and the growth rates of the culture (see Fig C.1). A linear dependence between the parameters and growth rate was determined,

## C.6. Blanking of the auto-fluorescence in the GFP signal

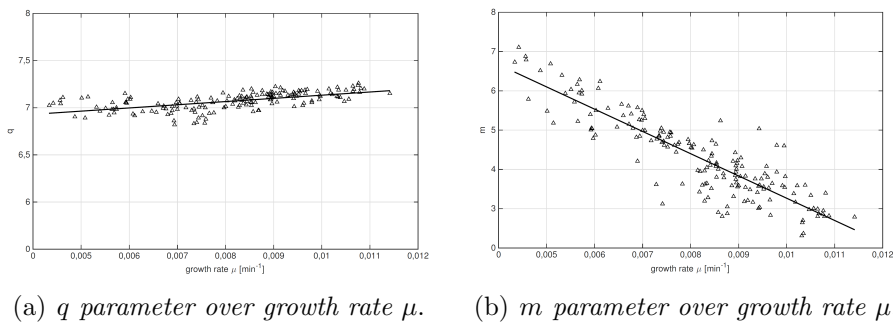


Figure C.1: **Linear regression of  $q(\mu)$  and  $m(\mu)$ .** Dependence of the  $m$  and  $q$  parameters used in equation 2.1, function of the growth rate  $\mu$

described by the following equations:

$$q(\mu) = q_q + m_q \cdot \mu \quad (\text{C.1})$$

$$m(\mu) = q_m + m_m \cdot \mu \quad (\text{C.2})$$

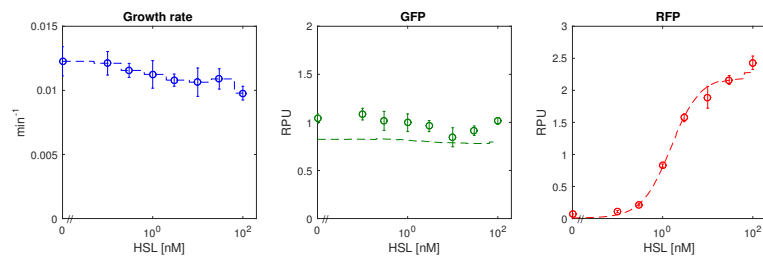
From here on, under the assumption of adopting same strain, instrument and testing conditions, it is possible to determine the auto-fluorescence signal of a liquid culture at a certain  $OD_{600}$  time point just by evaluating the overall growth rate, as follows:

$$GFP_{auto}(t) = e^{(q_q + m_q \cdot \mu) + (q_m + m_m \cdot \mu) \cdot OD_{600}(t)} \quad (\text{C.3})$$

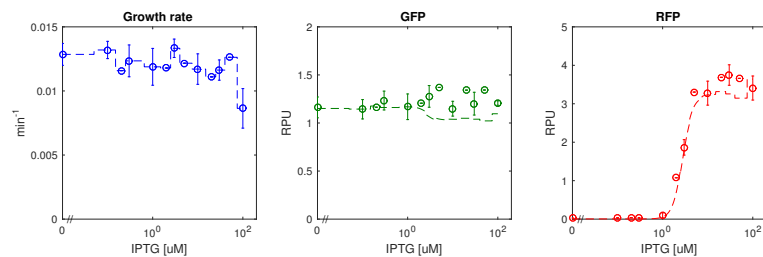
## C.7 Fitting and prediction from Section 3.3.4

### One step Global Fitting

#### Burden Model



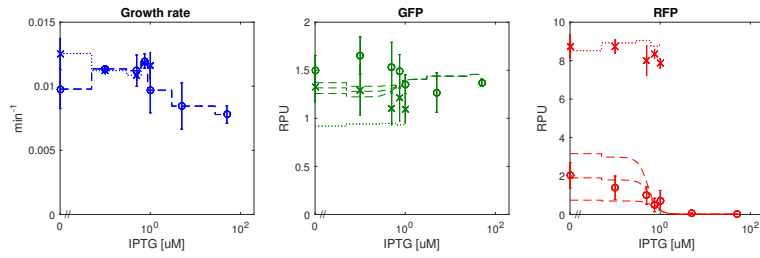
(a) *HSL-inducible sgRNA expression cassette (AE).*



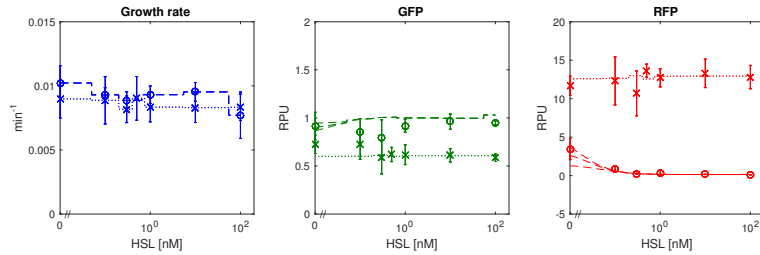
(b) *IPTG-inducible sgRNA expression cassette (AY).*

**Figure C.2: Burden model: Global fitting of sgRNA expression cassettes.** Interpolation of Growth rate and fitting of fluorescent protein signals for **a.** optimized HSL-inducible (AE) and **b.** IPTG-inducible (AY) expression cassettes. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

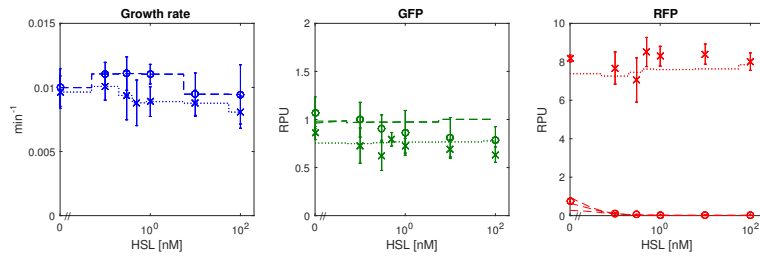
## C.7. Fitting and prediction from Section 3.3.4



(a) *IPTG-driven gPtet on MC target (E52)*.



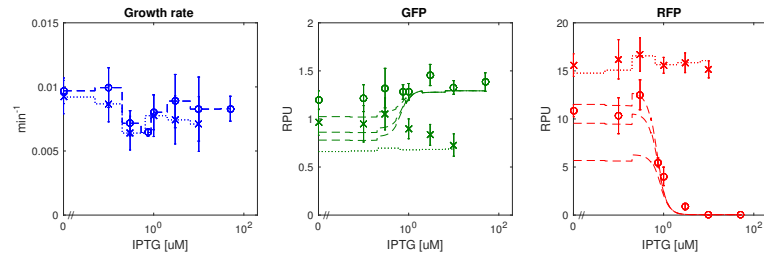
(b) *HSL-driven gPlac on MC target (E62)*.



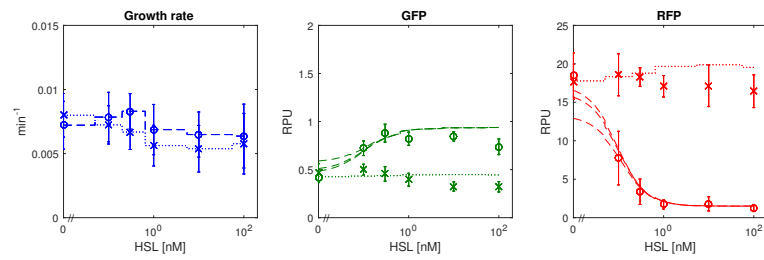
(c) *HSL-driven gPtet on MC target (E52)*.

**Figure C.3: Burden model: Global fitting of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals of sgRNAs repression exerted on MC targets. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

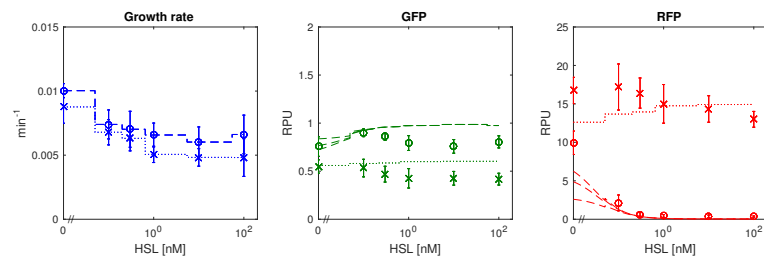
### C. Supplementary information for Chapter 3



(a) *IPTG-driven gPtet on HC target (I13521).*



(b) *HSL-driven gPlac on HC target (A33).*



(c) *HSL-driven gPtet on HC target (I13521).*

**Figure C.4: Burden model: Global fitting of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.



## C.7. Fitting and prediction from Section 3.3.4

### No Burden Model

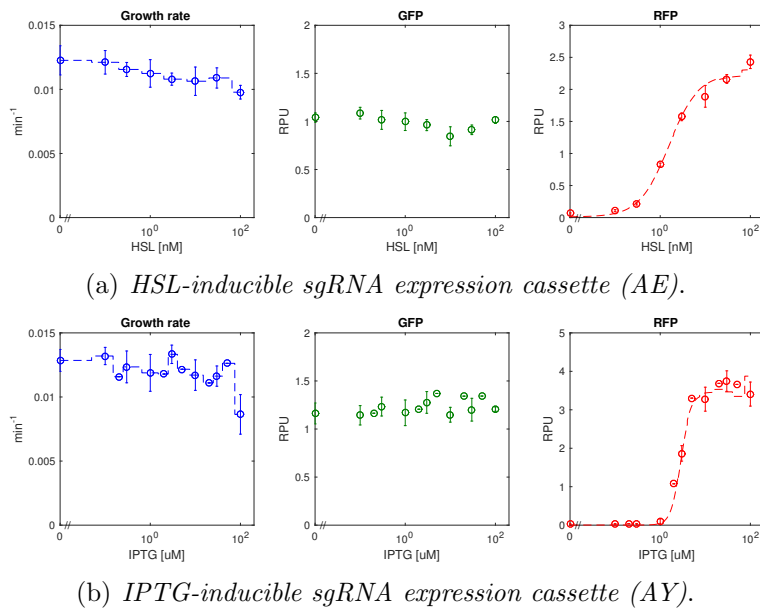
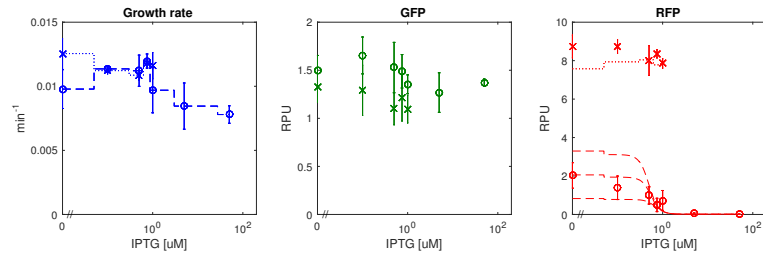
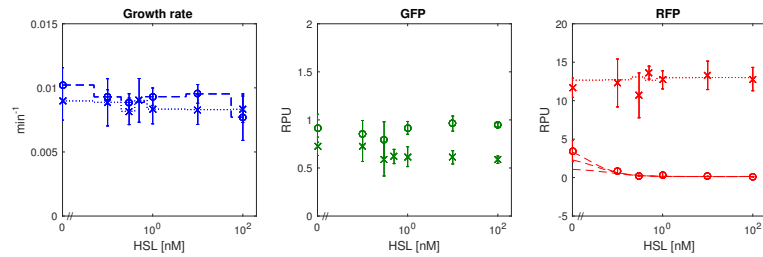


Figure C.5: **No Burden model: Global fitting of sgRNA expression cassettes.** Interpolation of Growth rate and fitting of fluorescent protein signals for **a.** optimized HSL-inducible (AE) and **b.** IPTG-inducible (AY) expression cassettes. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

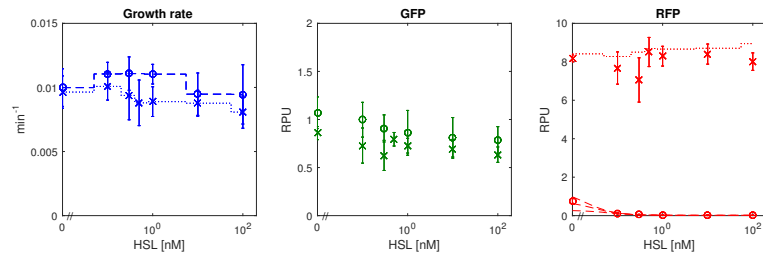
### C. Supplementary information for Chapter 3



(a) IPTG-driven *gPtet* on MC target (E52).



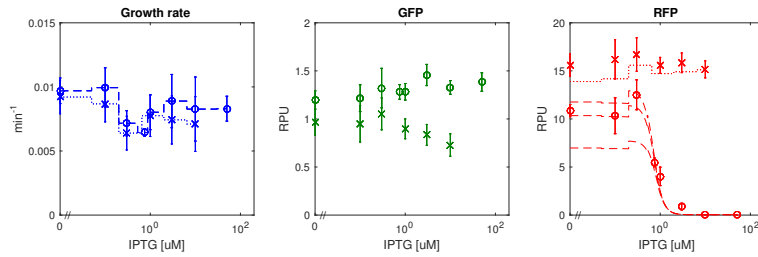
(b) HSL-driven *gPlac* on MC target (E62).



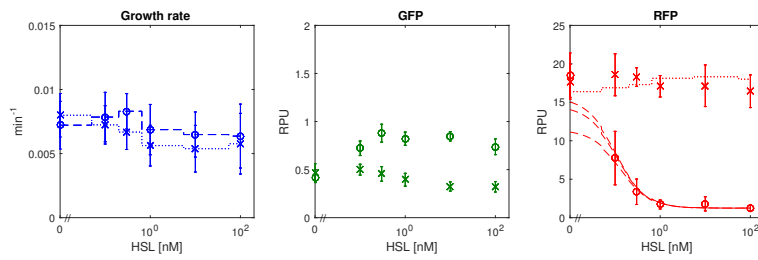
(c) HSL-driven *gPtet* on MC target (E52).

**Figure C.6: No Burden model: Global fitting of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals of sgRNAs repression exerted on MC targets. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

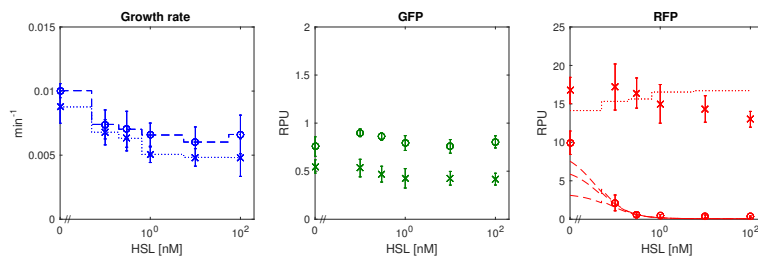
## C.7. Fitting and prediction from Section 3.3.4



(a) *IPTG-driven gPtet on HC target (I13521).*



(b) *HSL-driven gPlac on HC target (A33).*



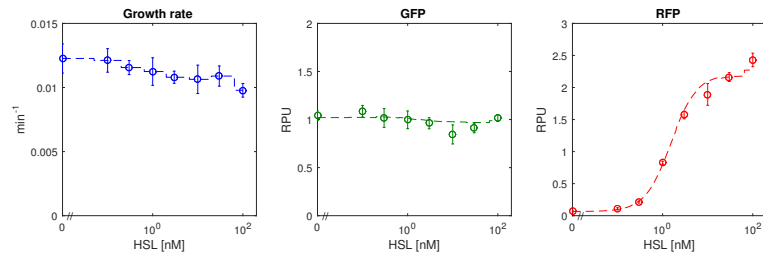
(c) *HSL-driven gPtet on HC target (I13521).*

**Figure C.7: No Burden model: Global fitting of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

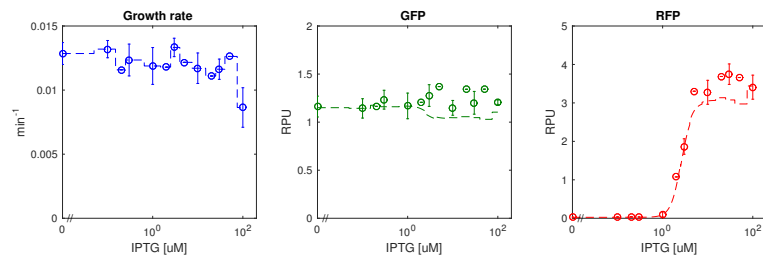
## C. Supplementary information for Chapter 3

### Two-step Global Fitting

#### Burden Model



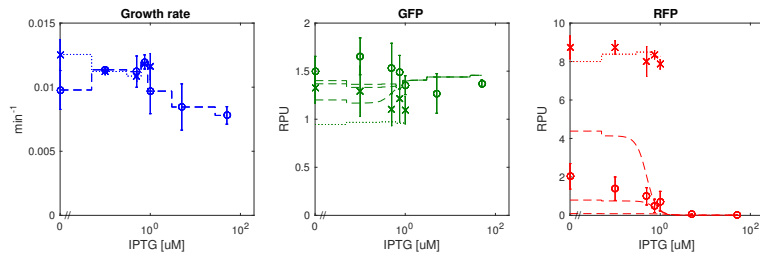
(a) *HSL-inducible sgRNA expression cassette (AE)*.



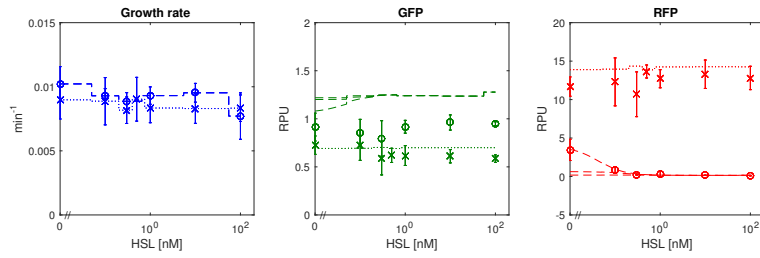
(b) *IPTG-inducible sgRNA expression cassette (AY)*.

**Figure C.8: Burden model: fitting of sgRNA expression cassettes.** Interpolation of Growth rate and fitting of fluorescent protein signals for **a.** optimized HSL-inducible (AE) and **b.** IPTG-inducible (AY) expression cassettes. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

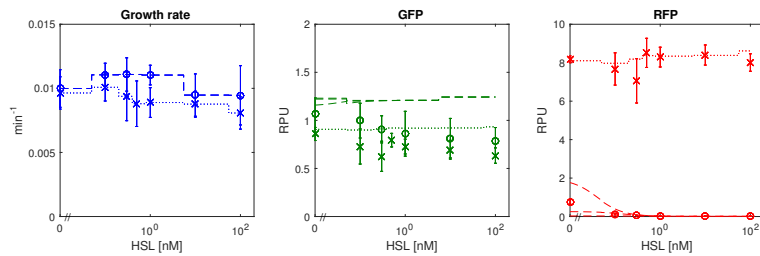
## C.7. Fitting and prediction from Section 3.3.4



(a) *IPTG-driven gPtet on MC target (E52)*.



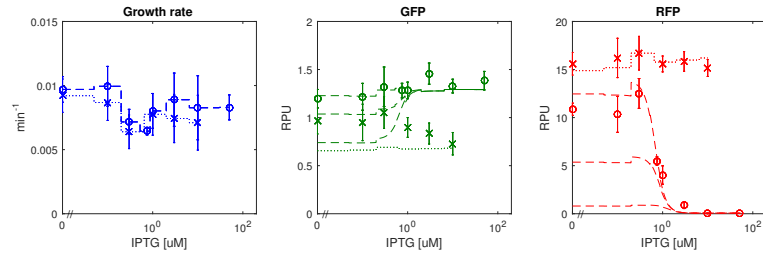
(b) *HSL-driven gPlac on MC target (E62)*.



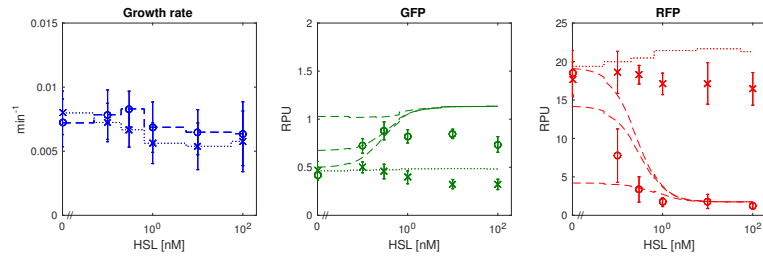
(c) *HSL-driven gPtet on MC target (E52)*.

**Figure C.9: Burden model: fitting w/o predictions of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

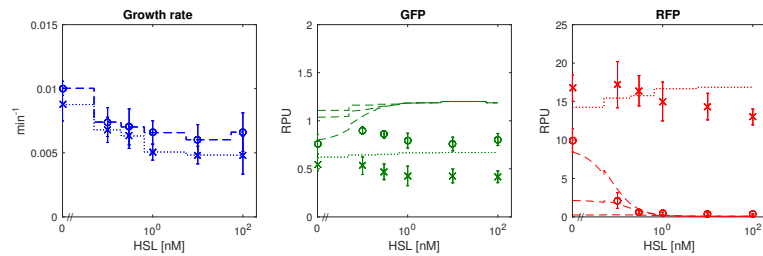
### C. Supplementary information for Chapter 3



(a) *IPTG-driven gPtet on HC target (I13521).*



(b) *HSL-driven gPlac on HC target (A33).*

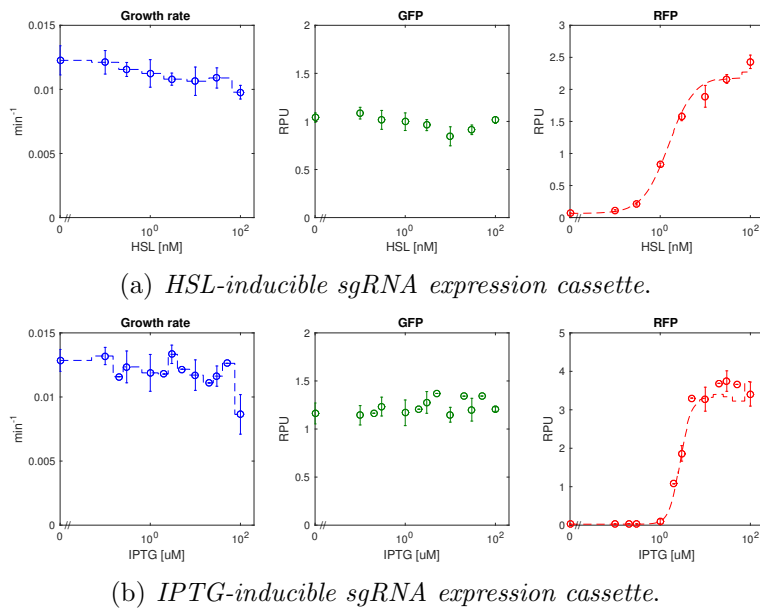


(c) *HSL-driven gPtet on HC target (I13521).*

**Figure C.10: Burden model: fitting w/o predictions of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

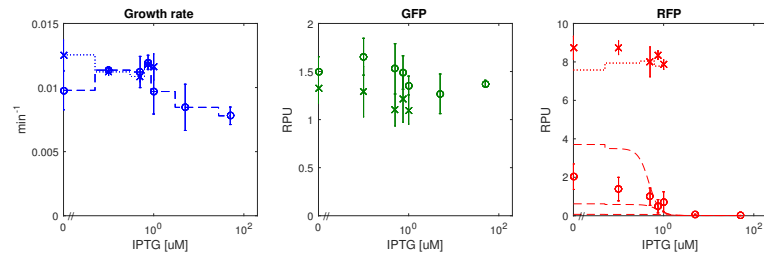
## C.7. Fitting and prediction from Section 3.3.4

### No Burden Model

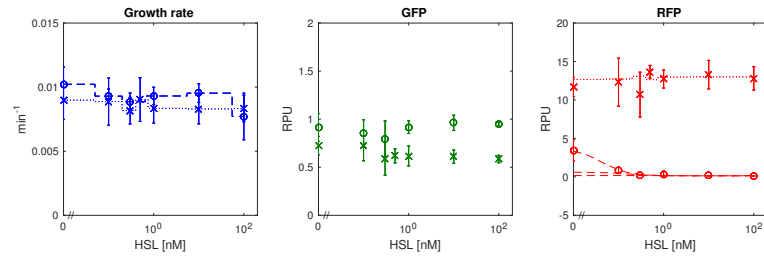


**Figure C.11: No Burden model: fitting of sgRNA expression cassettes.** Interpolation of Growth rate and fitting of fluorescent protein signals for **a.** optimized HSL-inducible (AE) and **b.** IPTG-inducible (AY) expression cassettes. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

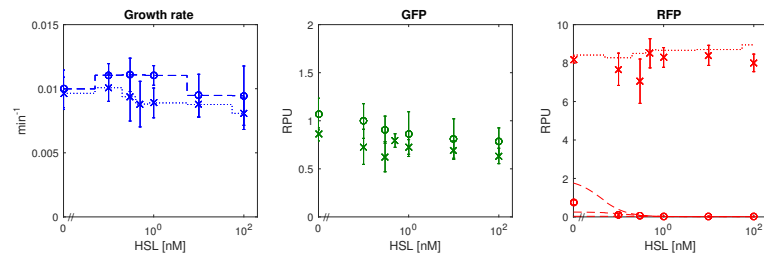
### C. Supplementary information for Chapter 3



(a) *IPTG-driven gPtet on MC target (E52)*.



(b) *HSL-driven gPlac on MC target (E62)*.

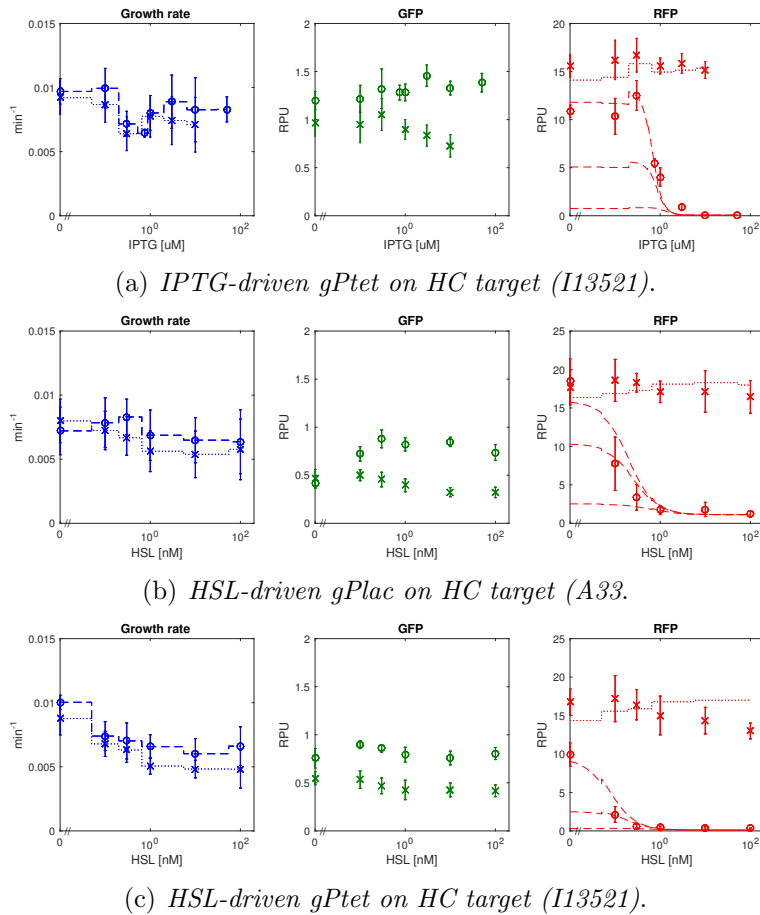


(c) *HSL-driven gPtet on MC target (E52)*.

**Figure C.12: No Burden model: fitting w/o predictions of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.



## C.7. Fitting and prediction from Section 3.3.4



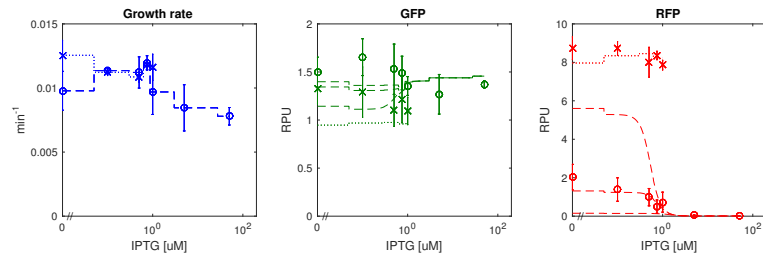
**Figure C.13: No Burden model: fitting w/o predictions of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## C. Supplementary information for Chapter 3

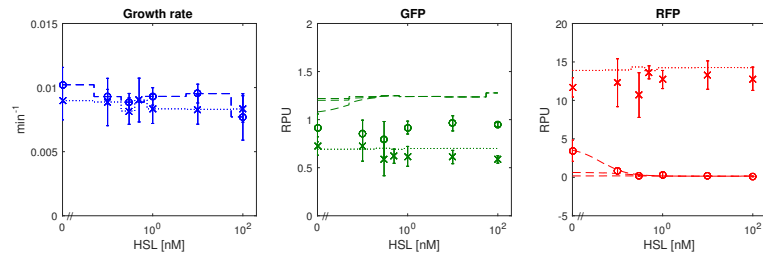
### Two-step with Training set and Test set for prediction

The first step is shared with the previous approach, sgRNA expression cassettes fittings are reported in Figures C.8 and C.11.

#### Burden model



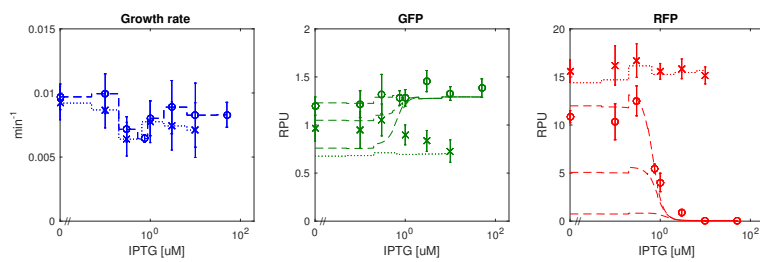
(a) IPTG-driven *gPtet* on MC target (E52).



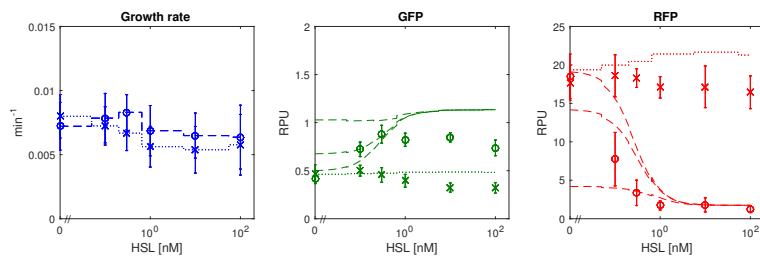
(b) HSL-driven *gPlac* on MC target (E62).

Figure C.14: **Burden model: fitting w/ predictions of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## C.7. Fitting and prediction from Section 3.3.4



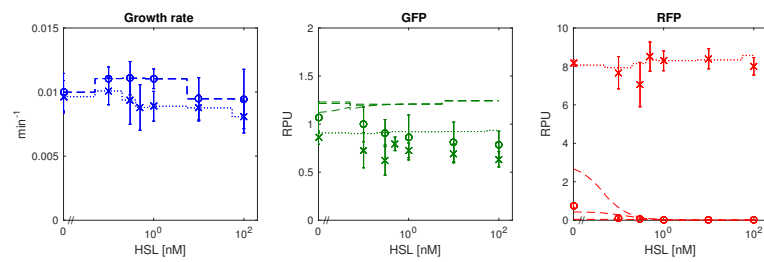
(a) *IPTG-driven gPtet on HC target (I13521).*



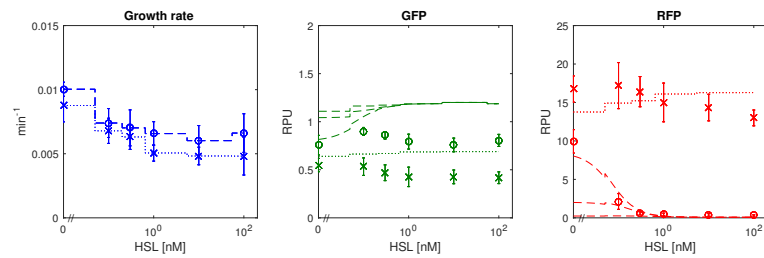
(b) *HSL-driven gPlac on HC target (A33).*

**Figure C.15: Burden model: fitting w/ predictions of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

### C. Supplementary information for Chapter 3



(a) *HSL-driven gPtet on MC target (E52).*

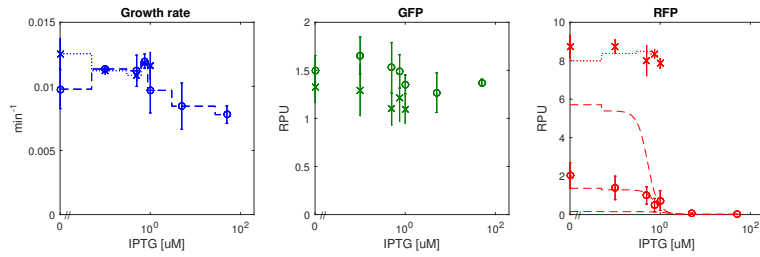


(b) *HSL-driven gPtet on HC target (I13521).*

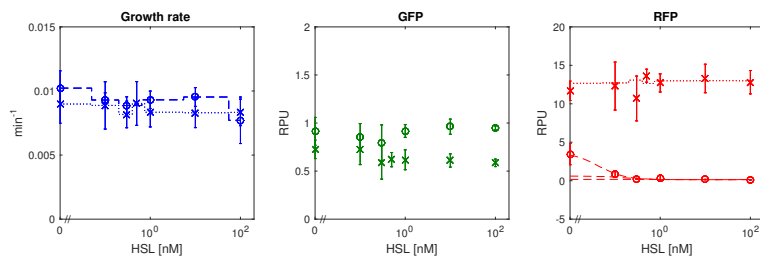
**Figure C.16: Burden model: predictions of HSL-driven gPtet.** Interpolation of Growth rate and prediction of fluorescent protein signals. **a.** MC target, **b.** HC target. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## C.7. Fitting and prediction from Section 3.3.4

### No Burden model



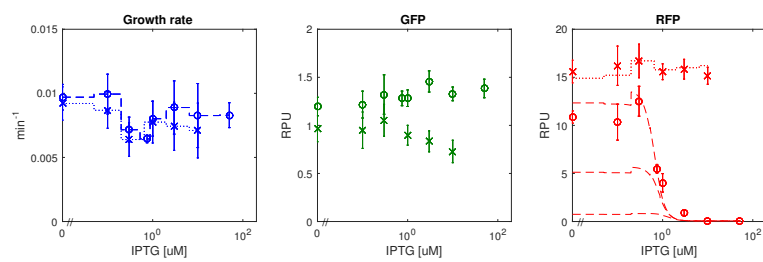
(a) *IPTG-driven gPtet on MC target (E52).*



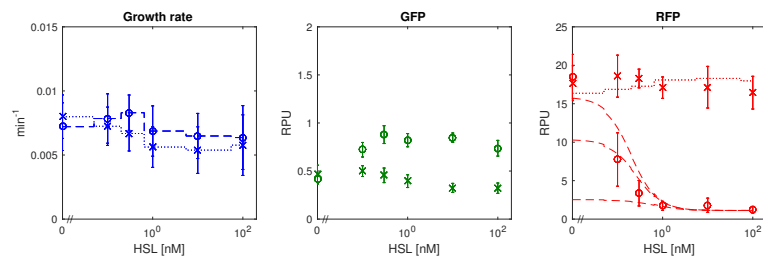
(b) *HSL-driven gPlac on MC target (E62).*

Figure C.17: **No Burden model: fitting w/ predictions of MC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

### C. Supplementary information for Chapter 3



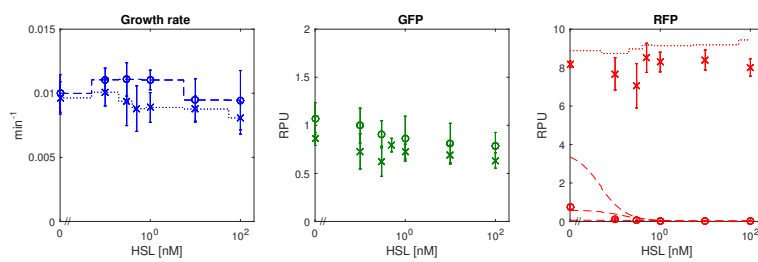
(a) *IPTG-driven gPtet on HC target (I13521).*



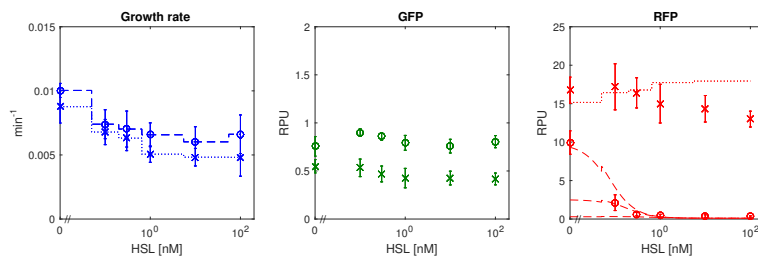
(b) *HSL-driven gPlac on HC target (A33).*

**Figure C.18: No Burden model: fitting w/ predictions of HC targets.** Interpolation of Growth rate and fitting of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

## C.7. Fitting and prediction from Section 3.3.4

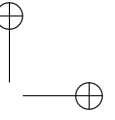
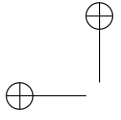


(a) *HSL-driven gPtet on MC target (E52).*



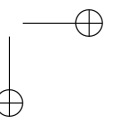
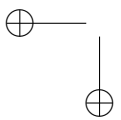
(b) *HSL-driven gPtet on HC target (I13521).*

**Figure C.19: No Burden model: predictions of HSL-driven gPtet** Interpolation of Growth rate and prediction of fluorescent protein signals. Amount of sgRNA in the cell depends on the inducer concentration, reported on the x-axis. Dashed lines represent the functional system while dotted ones are controls. Data are reported as mean values over at least 3 biological replicates while error bars represent the 95% confidence intervals of the mean.

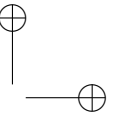
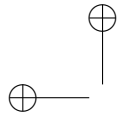


C. Supplementary information for Chapter 3

---

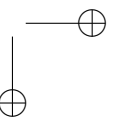
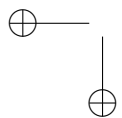






## Bibliography

- [1] Watson J.D and Crick F.H.C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 1953.
- [2] Cohen S., Chang A., Boyer H., and Helling R. Construction of biologically functional bacterial plasmids in vitro. In *Proceedings of the National Academy of sciences*, 1973.
- [3] Temin H. and Mizutani S. Rna-dependent dna polymerase in virions of rous sarcoma virus. *Nature*, 226(5252), 1970.
- [4] Baltimore D. Rna-dependent dna polymerase in virions of rna tumour viruses. *Nature*, 227(5253), 1970.
- [5] Crick F.H.C. Central dogma of molecular biology. *Nature*, 227(5258), 1970.
- [6] Kitney R., Calvert J., Challis R., Cooper J., Elck A., Freemont P.S., Haselo J., and Kelly M. amd Paterson L. Synthetic biology: scope and applications and implications. *The Royal Academy of Engineering*, 2009.





## BIBLIOGRAPHY

---

- [7] Heinemann M. and Panke S. Synthetic biology: putting engineering into biology. *Bioinformatics*, 22(22), 2006.
- [8] Chopra P. and Kamma A. Engineering life through synthetic biology. *In silico biology*, 6(5), 2006.
- [9] Moser F., Broers N.J., Hartmans S., Tamsir A. and Kerkman R., Roubos J.A., Bovenberg R., and Voigt C.A. Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synthetic Biology*, 1(11), 2012.
- [10] Voigt C.A. Genetic parts to program bacteria. *Current opinion in biotechnology*, 17(5), 2006.
- [11] Endy D. Foundations for engineering biology. *Nature*, 438(7067), 2005.
- [12] Andrianantoandro E., Basu S., Karig D.K., and Weiss R. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2(1), 2006.
- [13] Pasotti L., Politi N., Zucca S., Cusella De Angelis M. G., and Magni P. Bottom-up engineering of biological systems through standard bricks: a modularity study on basic parts and devices. *PLoS One*, 7, 2012.
- [14] Sauro H.M. Modularity defined. *Molecular System Biology*, 4, 2008.
- [15] Jayanthi S., Nilgiriwala K., and Del Vecchio D. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synthetic Biology*, 2, 2013.
- [16] Sprinzak D. and Elowitz M.B. Reconstruction of genetic circuits. *Nature*, 438(7067), 2005.



## BIBLIOGRAPHY

---

- [17] Chandran D., Copeland W.B., Sleight S.C., and Sauro H.M. Mathematical modeling and synthetic biology. *Drug Discovery Today: Disease Models*, 5(4), 2009.
- [18] Salis H.M., Mirsky E.A., and Voigt C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27, 2009.
- [19] Gyorgy A., Jimenez J.I., Yazbek J., Huang H.H., Chung H., Weiss R., and Del Vecchio D. Isocost lines describe the cellular economy of genetic circuits. *Biophys Journal*, 109, 2015.
- [20] Carr S.B., Beal J., and Densmore D.M. Reducing dna context dependence in bacterial promoters. *PLoS One*, 12, 2017.
- [21] Lou C., Stanton B., Chen Y. J., Munsky B., and Voigt C.A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology*, 30, 2012.
- [22] Ceroni F., Boo A., Furini S., Goroehowski T.E., Borkowski O., Ladak Y.N., Awan A.R., Gilbert C., Stan G.B., and Ellis T. Burden-driven feedback control of gene expression. *Nature Methods*, 15(5), 2018.
- [23] Pasotti L. and Zucca S. Advances and computational tools towards predictable design in biological engineering. *Computational Mathematical Methods in Medicine*, 2014.
- [24] Schwille P. Bottom-up synthetic biology: engineering in a tinkerer's world. *Science*, 333, 2011.
- [25] Lu T.K., Khalil A.S., and Collins J.J. Next-generation synthetic gene networks. *Nature Biotechnology*, 27, 2009.



## BIBLIOGRAPHY

---

- [26] Arkin A.P. A wise consistency: engineering biology for conformity, reliability, predictability. *Current Opinion in Chemical Biology*, 17, 2013.
- [27] Muers M. Synthetic biology: quality and quantity. *Nature Reviews Genetics*, 14, 2013.
- [28] Del Vecchio D., Ninfa A.J., and Sontag E.D. Modular cell biology: retroactivity and insulation. *Molecular System Biology*, 4, 2008.
- [29] Kelly J.R., Rubin A.J., Davis J.H., Ajo-Franklin C.M., Cumbers J., and Czar M.J. Measuring the activity of biobrick promoters using an in vivo reference standard. *Journal of Biological Engineering*, 3, 2009.
- [30] Espah Borujeni A., Channarasappa A.S., and Salis H.M. Translation rate is controlled by coupled trade-offs between site accessibility, selective rna unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42, 2014.
- [31] Chen Y.J., Liu P., Nielsen A.A., Brophy J.A., Clancy K., Peterson T., and Voigt C.A. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods*, 10, 2013.
- [32] Cambray G., Guies J.C., Mutalik V.K., Lam C., Mai Q.A., and Thimmaiah T. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Research*, 41, 2013.
- [33] Hajimorad M., Gray P.R., and Keasling J.D. A framework and model system to investigate linear system behavior in escherichia coli. *Journal of Biological Engineering*, 5(3), 2011.



## BIBLIOGRAPHY

---

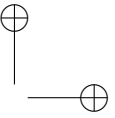
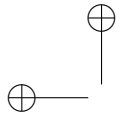
- [34] Guido N.J., Wang X., Adalsteinsson D., McMillen D., Hasty J., and Cantor C.R. A bottom-up approach to gene regulation. *Nature*, 439, 2006.
- [35] Ceroni F., Furini S., Stefan A., Hochkoeppler A., and Giordano E. A synthetic post-transcriptional controller to explore the modular design of gene circuits. *ACS Synthetic Biology*, 2012.
- [36] Anderson J.C., Voigt C.A., and Arkin A.P. Environmental signal integration by a modular and gate. *Molecular System Biology*, 3, 2007.
- [37] Wang B., Kitney R.I., Joly N., and Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2, 2011.
- [38] Moon T.S., Lou C., Tamsir A., Stanton B.C., and Voigt C.A. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491, 2012.
- [39] Ellis T., Wang X., and Collins J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology*, 27, 2009.
- [40] Kosuri S., Goodman D.B., Cambray G., Mutalik V.K., Gao Y., and Arkin A.P. Composability of regulatory sequences controlling transcription and translation in escherichia coli. *PNAS*, 110, 2013.
- [41] Goodman D.B., Church G.M., and Kosuri S. Causes and effects of n-terminal codon bias in bacterial genes. *Science*, 342, 2013.
- [42] Davis J.H., Rubin A.J., and Sauer R.T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research*, 39, 2011.



## BIBLIOGRAPHY

---

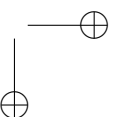
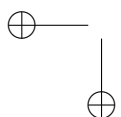
- [43] Mutalik V.K., Guimaraes J. C., Cambray G., Lam C., Christofersen M.J., and Mai Q.A. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10, 2013.
- [44] Mishra D., Rivera P.M., Lin A., Del Vecchio D., and Weiss R. A load driver device for engineering modularity in biological networks. *Nature Biotechnology*, 32, 2014.
- [45] Mutalik V.K., Guimaraes J.C., Cambray G., Mai Q.A., Christofersen M.J., and Martin L. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nature Methods*, 10, 2013.
- [46] Nielsen A.A., Der B.S., Shin J., Vaidyanathan P., Paralanov V., Strychalski E.A., Ross D., Densmore D., and Voigt C.A. Genetic circuit design automation. *Science*, 352, 2106.
- [47] Borkowski O., Ceroni F., Stan G.B., and Ellis T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinions in Microbiology*, 33, 2016.
- [48] Shachrai I., Zaslaver A., Alon U., and Dekel E. Cost of unneeded proteins in e. coli is reduced after several generations in exponential growth. *Molecular Cell*, 38, 2010.
- [49] Pasini M., Fernandez-Castane A., Jaramillo A., de Mas C., Caminal G., and Ferrer P. Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant escherichia coli. *Nature Biotechnology*, 33, 2016.
- [50] Lo T. M., Chng S. H., Teo W. S., Cho H. S., and Chang M. W. A two-layer gene circuit for decoupling cell growth from metabolite production. *Cell Systems*, 3, 2016.



## BIBLIOGRAPHY

---

- [51] Ceroni F., Algar R., Stan G. B., and Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods*, 12, 2015.
- [52] Dragosits M Nicklas D Tagkopoulos I. A synthetic biology approach to self-regulatory recombinant protein production in escherichia coli. *Journal of Biological Engineering*, 6(2), 2012.
- [53] Qian Y. and Del Vecchio D. Effective interaction graphs arising from resource limitations in gene networks. In *Proc American Control Conference*, 2015.
- [54] Carbonell-Ballester M., Garcia-Ramallo E., Montanez R., Rodriguez-Caso C., and Macia J. Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the ohm's law. *Nucleic Acids Research*, 44, 2016.
- [55] Algar R.J.R., Ellis T., and Stan G.B. Modelling essential interactions between synthetic genes and their chassis cell. In *Proc. 53rd IEEE Conference on Decision and Control*, 2014.
- [56] Weisse A. Y., Oyarzun D. A., Danos V., and Swain P. S. Mechanistic links between cellular trade-offs, gene expression, and growth. *PNAS*, 112, 2015.
- [57] Ang J., Harris E., Hussey B.J., Kil R., and McMillen D.R. Tuning response curves for synthetic biology. *ACS Synthetic Biology*, 2, 2013.
- [58] Chandra F.A. and Del Vecchio D. The effects of ribosome autocatalysis and negative feedback in resource competition. *bioRxiv*, 2016.





## BIBLIOGRAPHY

---

- [59] Qian Y., Huang H.H., and Jimenez J.I. and Del Vecchio D. Resource competition shapes the response of genetic circuits. *ACS Synthetic Biology*, 2017.
- [60] Marraffini L.A. and Sontheimer E.J. Crispr interference: Rna-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*, 11(3), 2010.
- [61] Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., and Horvath P. Crispr provides acquired resistance against viruses in prokaryotes. *Science*, 2007.
- [62] Horvath P. and Barrangou R. Crispr/cas, the immune system of bacteria and archaea. *Science*, 315(5819), 2010.
- [63] Mojica F., Diez-Villasenor C., Soria E., and Juez G. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. *Molecular Microbiology*, 36(1), 2002.
- [64] Farasat I. and Salis H.M. A biophysical model of crispr/cas9 activity for rational design of genome editing and gene regulation. *PLoS Computational Biology*, 12(1), 2016.
- [65] Qi L.S., Larson M.H., Gilbert L.A., Doudna J.A., Weissman J.S., Arkin A.P., and Lim W.A. Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 2013.
- [66] Silva G., Poirot L., Galetto R., Smith J., Montova G., Duchateau P., and Paques F. Meganucleases and other tools for targeted genome engineering: Perspectives and challenges for gene therapy. *Current Gene Therapy*, 11(1), 2011.





## BIBLIOGRAPHY

---

- [67] Urnov F.D., Rebar E., Holmes M., Zhang H., and Gregory P. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9), 2010.
- [68] Chandrasegaran S. and Carroll D. Origins of programmable nucleases for genome engineering. *Journal of Molecular Biology*, 428(5 Pt B), 2016.
- [69] Pattanayak V., Guilinger J.P., and Liu D.R. Determining the specificities of talens, cas9, and other genome editing enzymes. *Methods Enzymology*, 2015.
- [70] Brookhouser N., Raman S., Potts C., and Brafman D. May i cut in? gene editing approaches in human induced pluripotent stem cells. *Cells*, 6(1), 2017.
- [71] Srikanth P., Han K., Callahan D.G., Makovkina E., Muratore C.R., Lalli M.A., Zhou H., Boyd J.D., Kosik K.S., Selkoe D.J., and Young-Pearse T.L. Genomic disc1 disruption in hipsocs alters wnt signaling and neural cell fate. *Cell Reports*, 12(9), 2015.
- [72] Lee H., Sundberg B., Sigafos A., and Clark K. Genome engineering with tale and crispr systems in neuroscience. *Frontiers in Genetics*, 2016.
- [73] Chavez A., Scheiman J., Vora S., Pruitt B., Tuttle M., Lyer P., Lin S., Kiani S., Guzman C., Weigand D., Ter-Ovanesyan D., Braff J., Davidsohn N., Housden B. and Perrimon N., Weiss R., Aach J., and Collins J. and Church G. Highly efficient cas9-mediated transcriptional programming. *Nature Methods*, 12(4), 2015.
- [74] Lo A. and Qi L. Genetic and epigenetic control of gene expression by crispr-cas systems. *F1000Research*, 6(747), 2017.



## BIBLIOGRAPHY

---

- [75] Bikard D., Euler C., Jiang W., Nussenzweig P., Goldberg G., Duportet X., Fischetti V., and Marraffini L. Exploiting crispr-cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology*, 32(11), 2014.
- [76] Citorik R., Mimee M., and Lu T. Sequence-specific antimicrobials using efficiently delivered rna-guided nucleases. *Nature Biotechnology*, 32(11), 2014.
- [77] Yin C., Zhang T., Qu X., Zhang Y., Putatunda R., Xiao X., Li F., Xiao W., Zhao H., Dai S., Qin X., Mo X., Young W.B., Khalili K., and Hu W. In vivo excision of hiv-1 provirus by sacas9 and multiplex single-guide rnas in animal models. *Molecular Therapy*, 25(5), 2017.
- [78] Chen B. and Zou W. and Huang B. Crispr-tag: an efficient dna tagging system in living cells. *bioRxiv preprint*, 2018.
- [79] Pasotti L., Bellato M., Casanova M. and Zucca S., Cusella De Angelis M.G., and Magni P. Re-using biological devices: a model-aided analysis of interconnected transcriptional cascades designed from the bottom-up. *Journal of Biological Engineering*, 11(1), 2017.
- [80] Hooshangi S., Thiberge S., and Weiss R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *PNAS*, 102, 2005.
- [81] Shetty R.P., Endy D., and Knight T.F. Engineering biobrick vectors from biobrick parts. *Journal of Biological Engineering*, 2, 2008.
- [82] Elowitz M.B. and Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403, 2000.



## BIBLIOGRAPHY

---

- [83] Fuqua W. C., Winans S. C., and Greenberg E. P. Quorum sensing in bacteria: the luxR-luxI family of cell density-responsive transcriptional regulators. *Journal of Bacteriology*, 176, 1994.
- [84] Zucca S., Pasotti L., Politi N., Casanova M., Mazzini G., Cusella De Angelis M. G., and Magni P. Multi-faceted characterization of a novel luxR-repressible promoter library for escherichia coli. *PLoS One*, 10, 2015.
- [85] Lutz R. and Bujard H. Independent and tight regulation of transcriptional units in escherichia coli via the lacR/o, the tetr/o and arac/i1-i2 regulatory elements. *Nucleic Acids Research*, 25, 1997.
- [86] Andersen J. B., Sternberg C., Poulsen L. K., Bjorn S. P., Givskov M., and Molin S. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and Environmental Microbiology*, 64, 1998.
- [87] Zucca S., Pasotti L., Mazzini G., Cusella De Angelis M.G., and Magni P. Characterization of an inducible promoter in different dna copy number conditions. *BMC Bioinformatic*, 13(Suppl 4), 2012.
- [88] Politi N., Pasotti L., Zucca S., Casanova M., Micoli G., Cusella De Angelis M. G., and Magni P. Half-life measurements of chemical inducers for recombinant gene expression. *Journal of Biological Engineering*, 8, 2014.
- [89] Massaiu I., Pasotti L., Casanova M., Politi N., Zucca S., Cusella De Angelis M. G., and Magni P. Quantification of the gene silencing performances of rationally designed synthetic small rnas. *Systems and Synthetic Biology*, 9, 2015.



## BIBLIOGRAPHY

---

- [90] Kwok R. Five hard truths for synthetic biology. *Nature*, 463, 2010.
- [91] Carbonell-Ballester M., Duran-Nebreda S., Montanez R., Sole R., Macia J., and Rodriguez-Caso C. A bottom-up characterization of transfer functions for synthetic biology designs: lessons from enzymology. *Nucleic Acids Research*, 42, 2014.
- [92] Pasotti L., Quattrocchi M., Galli D., Cusella de Angelis M. G., and Magni P. Multiplexing and demultiplexing logic functions for computing signal processing tasks in synthetic biology. *Biotechnol Journal*, 6, 2011.
- [93] Canton B., Labno A., and Endy D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, 26, 2008.
- [94] Purcell O., Grierson C. S., Di Bernardo M., and Savery N. J. Temperature dependence of ssra-tag mediated protein degradation. *Journal of Biological Engineering*, 6, 2012.
- [95] Pedraza J. M. and van Oudenaarden A. Noise propagation in gene networks. *Science*, 307, 2005.
- [96] Politi N., Pasotti L., Zucca S., and Magni P. Modelling the effects of cell-to-cell variability on the output of interconnected gene networks in bacterial populations. *BMC System Biology*, 9 (Supplementary 3), 2015.
- [97] Sleight S.C., Bartley B.A., Lieviant J.A., and Sauro H.M. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, 4, 2010.
- [98] Cookson N.A., Mather W.H., Danino T., Mondragon-Palomino O., Williams R.J., Tsimring L.S., and Hasty J. Queueing up



## BIBLIOGRAPHY

---

- for enzymatic processing: correlated signaling through coupled degradation. *Molecular System Biology*, 7, 2011.
- [99] Cho S., Choe D., Lee E., Kim S.C., Palsson B., and Cho B.K. Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *ACS Synthetic Biology*, 7(4), 2018.
- [100] Griffiths A.J.F., Wessler S.R., Carroll S.B., and Doebley J. *An Introduction to Genetic Analysis*. Freeman/Worth, 11 edition, 2015.
- [101] Hansen L.H., Knudsen S., and Sorensen S.J. The effect of the lacy gene on the induction of iptg inducible promoters, studied in escherichia coli and pseudomonas fluorescens. *Curr. Microbiol.*, 36(6), 1998.
- [102] Mit: Registry of standard biological parts. <http://partsregistry.org>.
- [103] Doench J., Fusi N., Sullender M., Hedge M., Vaimberg E., Donovan K., Smith I., Tothova Z., Wilen C., Orchard R., Virgin H., Listgarten J., and Root D. Optimized sgrna design to maximize activity and minimize off-target effects of crispr-cas9. *Nature Biotechnology*, 34(2), 2016.
- [104] Hsu P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X., Shalem O., Cradick T.J., Marraffini L.A., Bao G., and Zhang F. Dna targeting specificity of rna-guided cas9 nucleases. *Nature Biotechnology*, 31(9), 2013.
- [105] Bikard D., Jiang W., Samai P., Hochschild A., Zhang F., and Marraffini L.A. Programmable repression and activation of bacterial gene expression using an engineered crispr-cas system. *Nucleic Acid Research*, 41(15), 2013.



## BIBLIOGRAPHY

---

- [106] Sternberg S.H., Redding S., Jinek M., Greene E.C., and Doudna J.A. Dna interrogation by the crispr rna-guided endonuclease cas9. *Nature*, 507(7490), 2014.
- [107] Bernstein J.A., Khodursky A.B., Lin P.H., Lin-Chao S., and Cohen S.N. Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *Proceedings of the National Academy of Sciences*, 99(15), 2002.
- [108] Chen P., Qian Y., and Del Vecchio D. A model for resource competition in crispr-mediated gene repression. *bioRxiv*, 2018.
- [109] Fusi N., Smith I., Doench J., and Listgarten J. In silico predictive modeling of crispr/cas9 guide efficiency. *bioRxiv*, 2015.
- [110] Pasotti L., Bellato M., Politi N., Casanova M., Zucca S., Cusella De Angelis M.G., and Magni P. A synthetic close-loop controller circuit for the regulation of an extracellular molecule by engineered bacteria. *IEEE Transactions on Biomedical Circuits and Systems*, 2018.
- [111] Smits W.K., Kuipers O.P., and Veening J.W. Phenotypic variation in bacteria: the role of feedback regulation. *Nature Reviews Microbiology*, 4(4), 2006.
- [112] Montefusco F., O.E. Akman O.E., Soyer O.S., and Bates D.G. Ultrasensitive negative feedback control: a natural approach for the design of synthetic controllers. *PLoS ONE*, 11(8), e0161605, Aug. 2016.
- [113] Harris A.W.K. and. Dolan J.A, Kelly C.L., Anderson J., and Papachristodoulou A. Designing genetic feedback controllers. *IEEE Transactions Biomedical Circuits Systems*, 9(4), 2015.



## BIBLIOGRAPHY

---

- [114] Del Vecchio D., Dy A.J., and Qian Y. Control theory meets synthetic biology. *Journal of the Royal Society Interface*, 13(120), 2016.
- [115] Vignoni A., Oyarzun D.A., Pico J., and Stan G.B. Control of protein concentrations in heterogeneous cell populations. *Proceedings of the 2013 European Control Conference*, 2013.
- [116] Kopniczky M.B., Moore S.J., and Freemont P.S. Multilevel regulation and translational switches in synthetic biology. *IEEE Transactions Biomedical Circuits Systems*, 9(4), 2015.
- [117] Cosentino C., Ambrosino R., Ariola M., Bilotta M., Pironti A., and Amato F. On the realization of an embedded subtractor module for the control of chemical reaction networks. *IEEE Transactions Automatic Control*, 61(11), 2016.
- [118] Holtz W.J. and Keasling J.D. Engineering static and dynamic control of synthetic pathways. *Cell*, 140(1), 2010.
- [119] Becksei A. and Serrano L. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786), 2000.
- [120] Oyarzun D.A., Lugagne J.B., and Stan G.B. Noise propagation in synthetic gene circuits for metabolic control. *ACS Synthetic Biology*, 4(2), 2014.
- [121] Austin D.W., Allen M.S., McCollum J.M., Dar R.D., Wilgus J.R., Sayler G.S., Samatova N.F., Cox C.D., and Simpson M.L. Gene network shaping of inherent noise spectra. *Nature*, 439(7076), 2006.
- [122] Rosenfeld N., Elowitz M.B., and Alon U. Negative autoregulation speeds the response times of transcription networks. *Journal Molecular Biology*, 323, 2002.

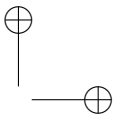
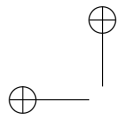


## BIBLIOGRAPHY

---

- [123] You L., Cox R.S.3rd, Weiss R., and Arnold F.H. Programmed population control by cell-cell communication and regulated killing. *Nature*, 428(6985), 2004.
- [124] Bloom R.J., Winkler S.M., and Smolke C.D. Synthetic feedback control using an rnai-based gene-regulatory device. *Journal of Biological Engineering*, 9(5), 2015.
- [125] Stapleton J.A., Endo K.and Fujita Y., Hayashi K., Takinoue M., Saitoand H., and Inoue T. Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synthetic Biology*, 1(3), 2012.
- [126] Franco E., Giordano G., Forsberg P.O., and Murray R.M. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synthetic Biology*, 3(8), 2014.
- [127] Farmer W.R. and Liao J.C. Improving lycopene production in escherichia coli by engineering metabolic control. *Nature Biotechnology*, 18(5), 2000.
- [128] Zhang F., Carothers M., and Keasling J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, 30(4), 2012.
- [129] Liu D., Xiao Y., Evans B.S., and Zhang F. Negative feedback regulation of fatty acid production based on a malonyl-coa sensor-actuator. *ACS Synthetic Biology*, 4(2), 2015.
- [130] Xu P., Li L., Zhang F., Stephanopoulos G., and Koffas M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *PNAS*, 111(31), 2014.

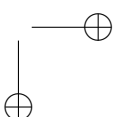
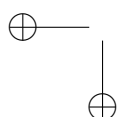




## BIBLIOGRAPHY

---

- [131] David F. and. Nielsen J and Siewers V. Flux control at the malonyl-coa node through hierarchical dynamic pathway regulation in *saccharomyces cerevisiae*. *ACS Synthetic Biology*, 5(3), 2016.
- [132] Teo J.J.Y., Woo S.S., and Sarpeshkar R. Synthetic biology: a unifying view and review using analog circuits. *IEEE Transactions Biomedical Circuits Systems*, 9(4), 2015.
- [133] Danino T., Mondragon-Palomino O., Tsimring L., and Hasty J. A synchronized quorum of genetic clocks. *Nature*, 463(7279), 2010.
- [134] Kotula J.W., Kerns S.J., Shaket L.A., Siraj L., Collins J.J., Way JC, and Silver P.A. Programmable bacteria detect and record an environmental signal in the mammalian gut. *PNAS*, 111(13), 2014.
- [135] Menolascina F., M. di Bernardo, and Di Bernardo D. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica*, 47(6), 2011.
- [136] Uhlendorf J., Miermont A., Delaveau T., Charvin G., Fages F., Bottani S., Batt G., and Hersen P. Long-term model predictive control of gene expression at the population and single-cell levels. *PNAS*, 109(35), 2012.
- [137] Fracassi C., Postiglione L., Fiore G., and Di Bernardo D. Automatic control of gene expression in mammalian cells. *ACS Synthetic Biology*, 5(4), 2016.
- [138] Zhang F. and Keasling J.D. Biosensors and their applications in microbial metabolic engineering. *Trends in Microbiology*, 19(7), 2011.

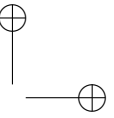
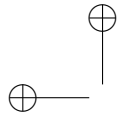




## BIBLIOGRAPHY

---

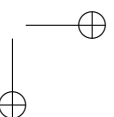
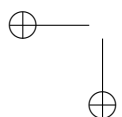
- [139] Qian S. and Cirino S.P. Using metabolite-responsive gene regulators to improve microbial biosynthesis. *Current Opinions in Chemical Engineering*, 14, 2016.
- [140] Dahl R.H., Zhang F., Alonso-Gutierrez J., Baidoo E., Batth T.S., Redding-Johanson A.M., Petzold C.J., Mukhopadhyay A., Lee T.S., Adams P.D., and Keasling J.D. Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology*, 31(11), 2013.
- [141] Bandiera L., Pasini A., Pasotti L., Zucca S., Mazzini G., Magni P., Giordano E., and Furini S. Experimental measurements and mathematical modeling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits. *Journal of Theoretical Biology*, 395, 2016.
- [142] Oyarzun D.A. and Stan G.B. Synthetic gene circuits for metabolic control: design trade-offs and constraints. *Journal of The Royal Society Interface*, 10(78), 2013.
- [143] Stevens A.M. and Greenberg E.P. Quorum sensing in vibrio fischeri: essential elements for activation of the luminescence genes. *Journal of Bacteriology*, 179(2), 1997.
- [144] Haseltine E.L. and Arnold F.H. Implications of rewiring bacterial quorum sensing. *Applied and Environmental Microbiology*, 74(2), 2008.
- [145] Dong Y.H., Xu J.L., Li X.Z., and Zhang L.H. Aiaa, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of erwinia carotovora. *PNAS*, 97(7), 2000.



## BIBLIOGRAPHY

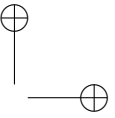
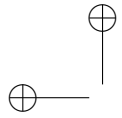
---

- [146] Kim K.H., Choi K. and Bartley B., and Sauro H.M. Controlling e. coli gene expression noise. *IEEE Transactions Biomedediacl Circuits Systems*, 9(4), 2015.
- [147] Knight T.F. *Idempotent vector design for standard assembly of biobricks*, 2003.
- [148] Lee T.S. Biobrick vectors and datasheets: A synthetic biology platform for gene expression. *Journal of Biological Engineering*, 5(12), 2011.



## BIBLIOGRAPHY

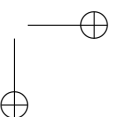
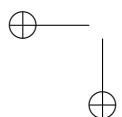
---



# List of publications

## Articles in peer reviewed journals

- L. Pasotti, **M. Bellato**, N. Politi, M. Casanova, S. Zucca, M.G. Cusella De Angelis and P. Magni  
*A synthetic close-loop controller circuit for the regulation of an extracellular molecule by engineered bacteria*, IEEE Transactions on Biomedical Circuits and Systems, 2018
- L. Pasotti, **M. Bellato**, M.Casanova, S. Zucca, M. G. Cusella De Angelis and P. Magni  
*Re-using biological devices: a model-aided analysis of interconnected transcriptional cascades designed from the bottom-up*, Journal of Biological Engineering 11:50, 2017
- L.Pasotti, **M. Bellato**, D. De Marchi and P.Magni  
Mechanistic models of inducible synthetic circuits for the joint description of DNA copy number, regulatory protein level and cell load, Processes, submitted

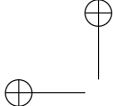
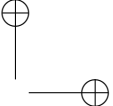




## Contributions to conference proceedings

- **M. Bellato**, L. Pasotti, D. De Marchi, C. Gualtieri, P. Magni, A. Balestrazzi and A. Macovei  
*A bioinformatic approach to predict the cross-kingdom potential of plant miRNA in humans*, Proceeding of BITS 2018. Turin, Italy, 2018
- L. Pasotti, **M. Bellato**, M. Casanova, S. Zucca, M.G. Cusella De Angelis and P. Magni  
*No stress for the host: model-based and in-vivo strategies to improve predictability and reliability of synthetic biological systems without overloading host cells*, Congresso Nazionale di Bioingegneria - Atti. Milan, Italy, 2018
- **M. Bellato**, L. Pasotti, E. Salibi, A. Frusteri Chiacchiera, P. Chen, Y. Qian, M. Casanova, M.G. Cusella De Angelis, D. Del Vecchio and P. Magni  
*CRISPRi for rational design of genetic circuits in Synthetic Biology*, Congresso Nazionale di Bioingegneria - Atti. Milan, Italy, 2018
- L. Pasotti, I. Massaiu, E. Rama, D. De Marchi, M. Cavalletti, M. Casanova, **M. Bellato**, A. Zebre, G. Mazzini, M.G. Cusella De Angelis, C. Calvio and P. Magni  
*In-silico and in-vivo metabolic engineering strategies to optimize waste bioconversion pathways in microbial strains: two case studies on ethanol and poly-gamma-glutamic acid biosynthesis*, Proceeding of SEED 2018. Scottsdale, Arizona, 2018
- P. Chen, Y. Qian, **M. Bellato**, A.J. Dy and D. Del Vecchio  
*A model for resource competition in CRISPRi circuits*, Proceeding of SEED 2018. Scottsdale, Arizona, 2018

- **M. Bellato**, L. Pasotti, E. Salibi, M. Casanova, M.G. Cusella De Angelis and P. Magni  
*Overcome Metabolic Burden in Rational Design of Genetic Synthetic Circuits: Deep Characterization of a CRISPRi Alternative*, Proceeding of SEED 2018. Scottsdale, Arizona, 2018
- L. Pasotti, S. Zucca, I. Massaiu, M. Casanova, **M. Bellato**, Mazzini, M.G. Cusella De Angelis, C. Calvio and P. Magni  
*Synthetic Biology Approaches for the Rational Engineering of Novel E. coli and B. Subtilis Biocatalysts for the Conversion of Industrial Bio-Waste into Biofuels and Bioproducts*, Proceeding of SEED 2017. Vancouver, Canada, 2017
- L. Pasotti, S. Zucca, I. Massaiu, M. Casanova, **M. Bellato**, Mazzini, M.G. Cusella De Angelis, C. Calvio and P. Magni  
*Efficient conversion of industrial bio-waste into biofuels and bioproducts through E. coli and B. Subtilis synthetic biology*, Proceeding of Synthetic Biology 7.0. Singapore, 2017
- **M. Bellato**, L. Pasotti, M. Casanova, I. Massaiu, M.G. Cusella De Angelis and P. Magni  
*Rational engineering of protein- and CRISPRi- mediated regulation devices to design predictable interconnected circuits with reduced cell load*, Proceeding of Synthetic Biology 7.0. Singapore, 2017
- L. Pasotti, S. Zucca, M. Casanova, **M. Bellato**, I. Massaiu, M. Arbuschi, M. Murgiano, A. Serra, M.G. Cusella De Angelis and P. Magni  
*Definition and in vivo evaluation of mathematical models to predict the effects of copy number variations and cell burden in interconnected synthetic circuits*, Proceeding of SEED 2016. Chicago, Illinois, 2016

- 
- 
- **M. Bellato**, L. Pasotti, A. Serra, M. Casanova, S. Zucca, M.G. Cusella De Angelis and P. Magni  
*Mathematical model-based prediction of metabolic burden effects in interconnected synthetic gene circuits*, Proceeding of the 3rd International Synthetic and Systems Biology Summer School. Volterra (Pisa), Italy, 2016
  - **M. Bellato**, L. Pasotti, F. Castronuovo, N. Politi, M. Casanova, I. Massaiu, S. Zucca, M.G. Cusella De Angelis and P. Magni  
*Study of a genetic negative feedback controller via bottom-up approach and mathematical modelling*, Congresso Nazionale di Bioingegneria - Atti. Naples, Italy, 2016
  - M. Casanova, L. Pasotti, S. Zucca, N. Politi, I. Massaiu, **M. Bellato**, G. Mazzini, M.G. Cusella De Angelis and P. Magni  
*Mathematical model-based modular design of genetic circuits to obtain expression systems with predictable mean and cell-to-cell variability of the protein output*, Proceeding of BITS 2015. Milan, Italy, 2015

