Novel Genetic Tools for Clostridia and a Synthetic Biology Approach to Control Spore Formation in Process Organisms

Ines Cristina Cañadas Blasco, MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2018

Abstract

Through the application of engineering principles, synthetic biology provides the means to harness the power of *living things* for the good of humankind. Central to this discipline is the ability to create systems through the assembly of standardised biological parts that may be incorporated into living cells. As a consequence, such synthetic chassis are endowed with biological properties of limitless, beneficial potential in applications as divergent as medicine, energy, food production and the environment.

The genus *Clostridium* holds particular potential as a chassis for synthetic biology applications. The beneficial attributes of its members range from their ability to convert industrial, greenhouse gases into chemicals and fuels to their use as anticancer agents. Despite their tremendous diversity, clostridial species are unified by their ability to form endospores. Dormant, non-reproductive structures, spores are increasingly finding use in a wide range of applications. These seek to exploit their resilient structure and high degree of resistance to all manner of chemical and physical agents. Such properties, however, are also of concern as they could enable recombinant spores to survive for prolonged periods of time if shed into the environment. To anticipate *potential* risks and facilitate future applications, innovative tools that address the biological containment of spores are therefore needed.

Regardless of the extraordinary diversity of clostridia, deriving both a greater understanding of their biology, while at the same time enhancing their exploitable properties, requires ever more sophisticated approaches. Hoping to accomplish this, the body of work described in this thesis explored a wide range of inducible systems that enabled precise control of gene expression in clostridia. The utility of these systems, that included riboregulators and a novel dual mechanism that acts at the level of transcription and translation, was exemplified by the creation of a conditionally sporulating strain of *Clostridium sporogenes*. This clostridial species, the most promising candidate in clostridial-based cancer therapies, was able to sporulate only in the presence of specific ligands, providing a solution to the prevention of the release of recombinant spores into the environment. Additionally, the clostridial synthetic biology toolbox was expanded by developing a novel and universal CRISPR-based genome editing tool. This has provided an extremely rapid and highly effective method for *Clostridium* engineering in synthetic biology applications.

Acknowledgements

I am very grateful to have so many people to acknowledge for the patience, help, support, wisdom and contribution that have made this thesis possible. First, I want to thank my supervisors, Dr Alan Cockayne and Professor Nigel Minton, for guiding me in this PhD adventure. Thanks for giving me the freedom to explore science and initiating me into the fascinating world of bacteria, genetic engineering and synthetic biology. Second, I would like to thank my bench supervisor, Aleksandra Bailey, for teaching me how to take the first steps in the lab and providing support whenever necessary. Many thanks to the clostridia group members who have made these four years in the lab (or the office) a fun place to investigate, chat, learn and laugh. Special thanks to my team, for being great friends and collaborators. In particular, thanks to my partner in science (and crime), Maria Zygouropoulou, for the amount of hours playing in the anaerobic cabinet, conversations and support. There are no words to express how grateful I am to have done this journey together, from induction to submission. To Daphne Groothuis, for all her support and wisdom as well as the friendship that we have built over these years. And to Raquel, for being always there. This thesis is as much theirs as it is mine.

I thank Dave Ortega and Andrew Dempster, for reading and commenting on this thesis and for helping whenever I needed it. I would also like to thank the support staff in the CBS and, in particular, to Evangelos Karnovis, for providing consistent media and buffers to my bacteria as well as to elaborate things *a la carta* every time I requested it.

Outside of the lab, I have to thank many other people who have been enormously supportive in incalculable ways. En primer lugar, a mis padres, por el apoyo incondicional en todas las decisiones que he tomado y el quererme siempre; gracias por batallar conmigo en la distancia. A mis hermanas, Maria, Aitana, Fuencisla, Paloma, Julia, Anita y progenie, por su inmensa fe en mí y por hacer de cada caída un trampolín desde el que impulsarme. A Francis, por su manera incondicional de empujarme. I have to thank as well all the friends and members of my family who provided support and accompanied me during this time. Gracias.

Last but not least, I want to thank JuanP, for absolutely everything.

This work was made possible by the BBSRC DTP programme to which I am tremendously thankful.

To JuanP and my amazing parents

Table of contents

Abstract i
Acknowledgements ii
Table of contents iv
List of Figures viii
List of Tables ix
Abbreviations x

Chapter 1 / Introduction

1.1 Designing biology: the Synthetic Biology era 2
1.2 Bacterial spores and the need for biological containment 5
1.3 *Clostridium*, a spore former relevant to biotechnology 9
1.4 To Sporulate (or not) 10
1.5 Tuning gene expression, a challenge in synthetic biology 15
1.6 The CRISPR Revolution 19
1.7 Aims of this study 22

Chapter 2 / Materials and Methods

2.1 Chemicals and suppliers 25
2.2 General growth methods 25
2.2.1 Growth media composition 25
2.2.2 Preparation of media 25
2.2.3 Preparation of media supplements 25
2.3 Culture and manipulation of *E. coli* strains 26
2.3.1 E. coli strains 26
2.3.2 Culture of E. coli cells 26
2.3.3 Storage of E. coli strains 27
2.3.4 Preparation of electro-competent E. coli cells 27
2.3.5 Transformation of electro-competent E. coli cells 28
2.4 Culture and manipulation of *Clostridium* strains 28
2.4.1 Clostridium strains 28
2.4.2 Clostridium cell culture 28
2.4.3 Storage of Clostridium strains 29
2.4.4 Methylation of plasmid DNA 29
2.4.5 Preparation of electrocompetent Clostridium cells 29
2.4.6 Transformation of Plasmid DNA from E. coli host into clostridia by conjugation 30
2.4.7 Transformation of Plasmid DNA into clostridia by electroporation 31
2.4.8 Plasmid curing 32
2.5 Molecular methods and techniques 32
2.5.1 Extraction and purification of genomic DNA 32
2.5.2 Extraction and purification of plasmid DNA 33
2.5.3 DNA amplification 33
2.5.4 Agarose gel electrophoresis and DNA extraction 35
2.5.5 Restriction endonuclease digestion 36
2.5.6 DNA ligation 36

2.5.7 DNA sequencing *36*
2.5.8 RNA extraction *37*
2.5.9 Amplification of 5'UTR *37*
2.5.10 RT-qPCR *38*
2.5.11 Deletions or integrations via Allelic Exchange *38*
2.5.12 Chromosomal integration and pyrE restoration via Allele-Coupled Exchange (ACE) *39*
2.5.13 Genome editing via CRISPR (RiboCas) *40*
2.6 Cloning of plasmid constructs **40**
2.6.1 Construction of plasmids used in Chapter 3 ..... *41*
2.6.2 Construction of plasmids used in Chapter 4 ..... *43*
2.6.3 Construction of plasmids used in Chapter 5 ..... *44*
2.6.4 Construction of plasmids used in Chapter 6 ..... *47*
2.6.5 Construction of plasmids used in Chapter 7 ..... *50*
2.7 Phenotypic analysis **52**
2.7.1 Growth curve experiments *52*
2.7.2 Sampling and optical density measurement *52*
2.7.3 Sporulation assays *53*
2.7.4 Transformation/conjugation efficiency assays *53*
2.7.5 Phase contrast microscopy *54*
2.8 Enzyme assays **54**
2.8.1 Preparation of cell lysates of C. sporogenes *54*
2.8.2 Determination of protein concentration using the BCA assay *55*
2.8.3 Enzymatic assay of Chloramphenicol Acetyltransferase (CAT) *55*
2.9 Analytics **56**
2.9.1 High performance liquid chromatography (HPLC) and HPLC/Mass spectrometry (MS) analysis *56*
2.10 List of bioinformatics tools **58**
2.11 Data and statistical analyses **58**

Chapter 3 / The lactose inducible system from *C. perfringens* strain 13

3.1 Background **61**
3.2 Chapter overview **64**
3.3 Results and Discussion **65**
3.3.1 Functionality of the LAC system in C. sporogenes *65*
3.3.2 Kinetics of the LAC system in C. sporogenes *68*
3.3.3 Lactose stability in C. sporogenes cultures *70*
3.3.4 Further characterization of the LAC system *73*
3.4 Key outcomes **87**

Chapter 4 / Theophylline responsive riboswitches in *C. sporogenes*

4.1 Background **89**
4.1.1 Riboswitches: RNA aptamers to control gene expression *89*
4.1.2 The theophylline responsive riboswitch *92*
4.2 Chapter overview **94**
4.3 Results and Discussion **95**
4.3.1 Theophylline responsive riboswitches for Clostridium *95*

4.3.2 Adjustability of the theophylline responsive riboswitches ... 101
4.3.3 Further characterization of the theophylline responsive riboswitches ... 104
4.3.4 Functional characterization of the theophylline responsive riboswitch in C. sporogenes NCIMB 10696 ... 108
4.4 Key outcomes ... 112

Chapter 5 / A dual mechanism to tightly control gene expression – the RiboLac

5.1 Background ... 114
5.2 Chapter overview ... 115
5.3 Results and Discussion ... 117
5.3.1 RiboLac, a dual mechanism with improved dynamic range ... 117
5.3.2 Functional characterization of the RiboLac ... 121
5.3.3 The RiboLac in the chromosome ... 124
5.3.4 The Lactoswitch, dual control over the target gene ... 129
5.4 Key outcomes ... 132

Chapter 6 / Development of a *C. sporogenes* conditionally sporulating strain

6.1 Background ... 134
6.1.1 A conditionally sporulating strain ... 134
6.1.2 Controlling spo0A to regulate sporulation ... 134
6.1.3 Other targets to generate conditionally sporulating strains ... 137
6.2 Chapter overview ... 138
6.3 Results and Discussion ... 140
6.3.1 Generation of an asporogeneous host to test conditional sporulation systems ... 140
6.3.2 spo0A complementation and pyrE repair ... 143
6.3.3 Characterization of C. sporogenes ΔspoA pyrE ... 144
6.3.4 Plasmid-based conditionally sporulating strains ... 147
6.3.5 Chromosomally stable conditionally sporulating strains ... 152
6.3.6 Characterization of the stable conditionally sporulating C. sporogenes::RiboLac-spo0A strain ... 155
6.3.7 Other targets to conditionally sporulate – spoIIR ... 159
6.4 Key outcomes ... 164

Chapter 7 / A universal Cas9-based editing tool for the genus *Clostridium*: RiboCas

7.1 Background ... 166
7.2 Chapter overview ... 167
7.3 Results and Discussion ... 168
7.3.1 Riboswitches to create an efficient Cas9-based tool ... 168
7.3.2 Construction of RiboCas ... 170
7.3.3 Analysis of RiboCas in C. sporogenes ... 171
7.3.4 RiboCas for more ambitious genome editing strategies ... 173
7.3.5 Demonstration of the universality of RiboCas ... 175
7.4 Key outcomes ... 181

Chapter 8 / General Discussion

8.1 Context 183
8.2 Original contributions 185
8.3 Perspectives 190

Bibliography 192
Appendix 205

List of Figures

Figure 1.1 The design-build-test phases in synthetic biology 3
Figure 1.2 First published "Recombinant DNA" figure 4
Figure 1.3 Diagrammatic representation of a typical bacterial endospore 11
Figure.1.4 Schematic representation of the sporulation cycle in *B. subtilis* 12
Figure 2.1 Schematic representation of SOE PCR 35
Figure 3.1 Schematic representation of the LAC system from *C. perfringens* strain 13 66
Figure 3.2 Schematic representation of pMTL-HZ1 67
Figure 3.3 LAC system dose-dependent response 67
Figure 3.4 Dose-response of LAC system in *C. sporogenes* 69
Figure 3.5 Time course of induction of LAC system 71
Figure 3.6 Schematic illustration of the 5´ RACE procedure 75
Figure 3.7 5´ RACE experiments on P_{bgaR} and P_{bgaL} 76
Figure 3.8 Phylogenetic tree of AraC/XylS-family TFs 78
Figure 3.9 Motifs identified within 300 bp upstream of the genes encoding the AraC/XylS-family TFs 79
Figure 3.10 Schematic representation of the LAC system and the hypothetical operator sequences O1 and O2 79
Figure 3.11 LAC mutations O1, O2 and O1/O2 80
Figure 3.12 LAC stepwise mutations and truncations 83
Figure 3.13 Hypothetical LAC system mechanism 86
Figure 4.1 The theophylline responsive riboswitch 94
Figure 4.2 Schematic illustration of the riboswitch-reporter plasmid 96
Figure 4.3 5′ RACE experiments on P_{fdx} 98
Figure 4.4 Riboswitch library CAT results 99
Figure 4.5 Schematic representation of a functional model of the theophylline responsive riboswitch 99
Figure 4.6 CAT results of theophylline responsive riboswitches with different 5′ UTRs 102
Figure 4.7 Nucleotide sequences of the theophylline responsive riboswitch located downstream of P_{fdx} and P_{fdx}* 103
Figure 4.8 RT-qPCR results 106
Figure 4.9 Predicted and observed behaviours for riboswitch-G downstream of P_{fdx} or P_{fdx}* in *C. sporogen*es 108
Figure 4.10 Dynamic and kinetic profiles of the theophylline responsive riboswitch located downstream of P_{fdx} or P_{fdx}* 110
Figure 5.1 Response of the dual theophylline-LAC systems (RiboLacs) in *C. sporogenes* to different concentrations of theophylline and lactose 118
Figure 5.2 Dose-response comparison 122
Figure 5.3 Time-course induction assays 123
Figure 5.4 Agarose gel electrophoresis of PCR products of the *pyrE* locus following integration of the Lac-*catP* and RiboLac-*catP* sequences 127
Figure 5.5 CAT assay of systems integrated into the chromosome 128
Figure 5.6 Lactoswitch evaluation 130

Figure 6.1 Allelic Exchange single crossovers 142
Figure 6.2 Allelic Exchange double crossovers 143
Figure 6.3 ACE double crossovers 144
Figure 6.4 Sporulation and growth profiles of *C. sporogenes* Δ*spo0A pyrE*$^{-}$, *C. sporogenes* Δ*spo0A*^{COMP}, *C. sporogenes* Δ*spo0A*, *C. sporogenes WT* and *C. sporogenes pyrE*$^{-}$ 145
Figure 6.5 Microscopic view of a *C. sporogenes* Δ*spo0A* and *C. sporogenes* WT cultures 147
Figure 6.6 Assay of plasmid-based conditionally sporulating strains 151
Figure 6.7 Plates of heat-treated, plasmid-based conditionally sporulating strain 152
Figure 6.8 Sporulation capacity of stable conditionally sporulating strains 153
Figure 6.9 Plates of heat-treated, stable conditionally sporulating strains 155
Figure 6.10 Dose-dependent increase in sporulation efficiency in *C. sporogenes*::RiboLac-*spo0A*. 156
Figure 6.11 Sporulation profiles of *C. sporogenes*::RiboLac-*spo0A* 159
Figure 6.12 Sporulation capacity of *C. sporogenes* Δ*spoIIR and C. sporogenes* Δ*spoIIR*^{COMP} 161
Figure 6.13 Sporulation capacity of *C. sporogenes*::P_{spoIIR}-*G-spoIIR* 162
Figure 7.1 RiboCas 169
Figure 7.2 Analysis of various RiboCas systems 172
Figure 7.3 Deletion of *spoIIE* and integration of ARC 175
Figure 7.4 Transformation efficiencies of RiboCas in *C. pasteurianum* 178

List of Tables

Table 2.1 Supplements used in this study 26
Table 4.1 Nucleotide sequences of the riboswitches used in this study 95
Table 4.2 Comparison of CAT activity results from different riboswitches 107
Table 7.1 Summary of RiboCas results in different *Clostridium* species 180

Abbreviations

5-FOA	5-fluoroorotic-acid
5´ RACE	Rapid amplification of cDNA ends
5´ UTR	5´ untranslated region
ACE	Allele-coupled exchange
ARA	Arabinose-inducible
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine serum albumin
Cas	CRISPR-associated
CAT	Chloramphenicol acetyltransferase
CDEPT	Clostridial directed-enzyme prodrug therapy
CRISPR	Clustered regularly interspaced palindromic repeats
dH_2O	Distilled water
DMSO	Dimethyl sulphoxide
DSB	Double strand break
DTNB	5,5´dithio-bis (2-nitrobenzoic acid)
EPB	Electroporation buffer
GMO	Genetically modified organism
HPLC	High performance liquid chromatography
HTH	Helix turn helix
LAC	Lactose-inducible
LB	Lysogeny broth
LHR	Left homologous region
MS	Mass spectrometry
PBS	Phosphate buffer saline
PCE	Prodrug-converting enzyme
PCR	Polymerase chain reaction
RBS	Ribosome binding site
RHR	Right homologous region
RNAP	RNA polymerase
RT-qPCR	Quantitative reverse transcription PCR
SD	Shine-dalgarno
SELEX	Systematic evolution of ligands by exponential enrichment
SOE	Splicing by overlap extension
TET	Tetracycline-inducible
TF	Transcription factor
TSS	Transcriptional start site
WT	Wild type
XYL	Xylose-inducible

Chapter 1
/ Introduction

"Give me a living protoplasm and I will re-make the whole animal and vegetable kingdom"

– Felix Le Dantec, 19th century

1.1 Designing biology: the Synthetic Biology era

In the last few decades, an approach to engineer biology was emerging alongside the experimental procedures that were nurturing our understanding of *living things*. From Jacques Loeb's hypothesis about artificially creating life in the laboratory circa 1900[1], to the complete design of a synthetic yeast genome (Sc2.0) in 2016[2], engineering biology has served to better understand certain aspects of the complex behaviour of natural cells.

The application of engineering principles to rationally design biological systems falls under the scope of synthetic biology, an emerging and interdisciplinary field that involves the collaboration of biologists, engineers, computer scientists, social scientists and designers, among others[3]. Although the consensus and precise definition of synthetic biology has yet to be reached, the official and most acknowledged one in the UK is that provided by the Royal Academy of Engineering[4]:

> "*Synthetic biology aims to design and engineer biologically based parts, novel devices and systems as well as redesigning existing, natural biological systems.*"
>
> (Royal Academy of Engineering, 2009)

The conflation of science with engineering (*knowing by making*[5]) is intended to shorten the time required to design, build and test novel synthetic gene networks (Figure 1.1). Thus, iterative cycles of the streamlined practice of engineering biology, can facilitate (and has already facilitated) the construction of biological systems for diverse applications, including the production of novel materials, drugs, foods and fuels.

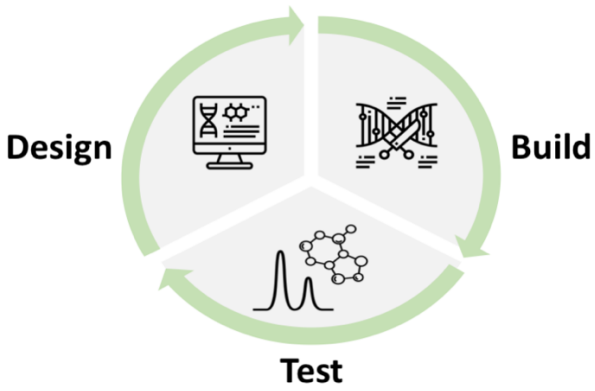


Figure 1.1 The design-build-test phases in synthetic biology

In synthetic biology, the design of biological programs is directed towards various epistemic concerns and foci. These are primarily grouped into three domains[3]: the production of commodity chemicals via re-designing metabolic pathways[6]; the rational design of logic genetic devices through the assembly of DNA parts[7], and; the generation of synthetic self-replicating living cells[2,8–10]. In these strategies, engineering principles such as abstraction and functional decoupling allow researchers to reduce biological complexity whereas modularization and standardization of biological parts serve to facilitate the construction of novel gene networks[11]. To engineer biology, biological functions must be simplified. The practice of biology, standardized.

In 1982, human insulin produced by engineered *Escherichia coli*, was the first recombinant DNA drug product to be marketed[12]. Such biotechnological application was made possible by the introduction of genetic engineering in the 1970s; the ability to cut and paste together DNA from different organisms (Figure 1.2)[13] and then introduce the recombinant elements produced into the chosen process organism[14]. Since then, the field of engineering living cells has had a huge impact in bringing advancement to

human life, attempting to address not only health-related problems, but also the rising global demand for food and various environmental issues[15]. However, in today's era of synthetic biology, it seems as if the entire field of genetic engineering had reached a conceptual breakthrough. Knowledge generates tools; tools generate more knowledge; more knowledge generates new tools (*knowing as making and making as knowing*[5]). In this context, living cells are seen as "chassis" harbouring the "cargo" needed to produce all manner of things. More efficient and fine-tuned tools for the manipulation of such chassis have been (and are constantly being) generated, in part, thanks to the modularization and standardization of biological parts. The circle closes when immense amounts of data, generated and shared via all type of databases, serve to create new chassis and tools, and therefore, more information. From and for hundreds of species. Worldwide.

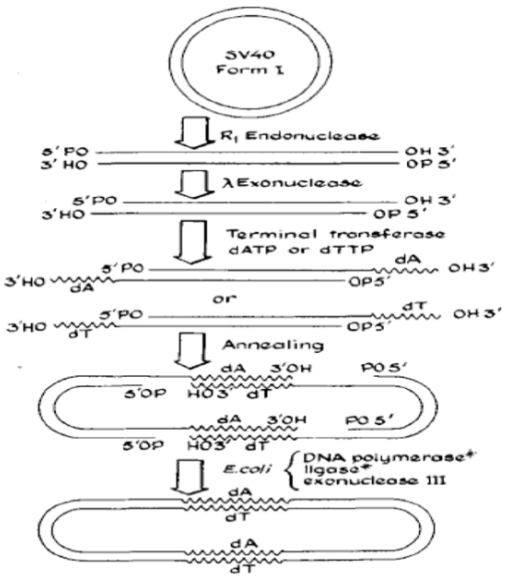


Figure 1.2 First published "Recombinant DNA" figure[13]. In 1972, two scientists, Paul Berg and David Jackson, joined the entire genome of SV40 to a piece of DNA; a chimera DNA molecule formed by the bacterial virus Lambda DNA and three genes from the bacterium *E. coli*. The hybrid DNA molecules were called "recombinant DNA".

1.2 Bacterial spores and the need for biological containment

Spores are metabolically inert entities produced by some vegetative bacterial cells as a survival response to adverse environmental conditions. As such, spores are well-known for their resilient structure, capable of defying extreme chemical and physical conditions and surviving in a wide variety of environments for long periods of time[16,17]. Although spores are the cause of much food spoilage, food poisoning and human diseases, their inherent resistance properties make them suitable for several beneficial applications. Of particular note is the growing use of bacterial spores in the area of biotechnology, where they have been used as probiotics, in vaccine formulations as well as in environmental bioremediation processes and biosensing systems*.

Probiotic products consisting of bacterial spores from *Bacillus* and *Clostridium* species, including *Bacillus subtilis, Bacillus cereus* and *Clostridium butyricum*, have been widely used in human and animal dietary supplements for several years[18–20]. However, the era of genetic manipulation has raised the concept of recombinant probiotics[18], establishing an exciting, growing field that exploits the beneficial properties of bacterial spores. One example of this is a study that has generated a strain of *B. subtilis* to expose on the spore surface the enzyme Phytase, a commonly used feed enzyme for animals[21]. Bacteria able to deliver enzymes directly to the gut can offer significant advantages to industry, in particular by eliminating the need of enzyme purification. Other examples include enzymes used for therapeutic purposes, such as the lipoprotein Lipase, widely employed in the treatment of cardiovascular and obesity-associated complications[22]. Immobilization of such enzymes on the spore surface would

*** Information linked to the 8th European Spores Conference, held in Egham (UK) on April 2018.**

allow their use as oral formulations for the treatment of diverse disorders and diseases.

Vaccines are also in the spotlight of spore applications. To a large extent, this interest lies in the extreme stability of spores, that could allow surface-displayed immunogens on recombinant spores (i.e., vaccines) to be delivered where cold storage is difficult[17]. Some such examples include the spore-associated recombinant expression of the B subunit of the heat-labile toxin of *E. coli* and the tetanus toxin fragment C from *Clostridium tetani* in *B. subtilis,* which were able to generate an immune response in a mouse model, and have been considered for oral administration[23]. Spore-based vaccines have also been demonstrated to confer protection against the white spot syndrome virus in shrimps[24,25] and against *Helicobacter pylori* or *Mycobacterium tuberculosis* in murine models[26,27].

Additionally, spores can be utilised in the fight against cancer. The uncontrolled growth of cancer cells within a tumour, and the deficient blood vessel architecture that supports it, leads to a hypoxic microenvironment that has been associated with some metastatic and aggressive forms of cancer[28]. However, such lack of oxygen offers great opportunities for anaerobic *Clostridium* strains, which can exploit this condition to selectively target solid tumours[29]. In particular, considerable research has focused on Clostridial Directed-Enzyme Prodrug Therapy, or CDEPT, whereby a clostridial host is genetically engineered to express a prodrug-converting enzyme (PCE)[30]. In this therapy, intravenously (or otherwise intratumorally) administered clostridial spores localize, and selectively germinate in, the hypoxic microenvironment of cancer tumours. Once germinated, the engineered strain expresses the PCE,

which metabolizes the non-toxic prodrug into its toxic derivative[30,31]. Since germination is restricted to the hypoxic regions of the tumour, expression of the PCE and activation of the prodrug with subsequent anticancer activity take place exclusively in such microenvironment. Most prominent among the clostridial species investigated as the "Trojan spore" that infiltrates into the tumour and destroys cancer cells is the proteolytic *Clostridium sporogenes*[32]. Apart from its inherent oncolytic properties, which stem from the expression of native cytosolic enzymes upon germination, *C. sporogenes* has demonstrated higher sporulation capacity and colonizing properties than other non-proteolytic species (i.e., *Clostridium acetobutylicum* and *Clostridium beijerinckii*)[33,34]. Additionally, *C. sporogenes* is entirely non-pathogenic, if compared with *Clostridium novyi-NT*, an attenuated proteolytic strain which has been used in several CDEPT studies[35,36]. Combinations of clostridial spores with radiation, vascular targeting agents and cytotoxic chemotherapeutic drugs at the pre-clinical level has demonstrated encouraging outcomes, resulting in tumour regression in all cases[37].

While the resilient structure of spores and their mechanisms of resistance make them suitable for the aforementioned applications, they are also the very cause of safety concerns associated with their use. These concerns are related to the use of genetically modified organisms (GMOs) and the risks of their release to the environment. Firstly, the engineering and growth of recombinant strains generally makes use of antibiotic resistance genes; the potential risk of antibiotic gene transfer following release is widely recognized. Secondly, the escape of engineered organisms into the environment could lead to unintentional colonization of others, even though most GMOs developed in the laboratory

seem to be less fit than wildtype (WT) strains[38]. Finally, the inherent robustness of spores results in the paradoxical situation of an unbreakable structure that needs to be "contained" or "destroyed" once used; recombinant spores could survive indefinitely in the environment after their deliberate release otherwise. To this day, it is not entirely clear whether the release of GMOs into the environment is completely harmful or safe. Thus, biocontainment, as the prevention of unintended environmental release[39], should be ensured to avoid any uncertain risk.

Physical containment, including the regulated disposal of GMO cultures, constitutes a standard procedure in research or industrial facilities[40]. However, if GMO applications continue to expand and intended to be used in open environments, further containment, including that achieved by biological means, might be required. A number of strategies have been developed over the last few years to provide biological containment (recently reviewed in[41]). These include the conditional activation of lethal genes (kill switches) to eliminate the engineered strains[42] or the use of "GeneGuard" plasmids, which are only functional in their intended host but disadvantageous for others[43]. Auxotrophic strains unable to survive if a specific compound is no longer supplied have also been demonstrated to reduce the spread of engineered cells, in particular, if compounds that are required in high concentrations are targeted (i.e., thymine in *thyA*$^{-}$)[44]. Recent attempts have also addressed biological containment by producing *E. coli* strains unable to survive in the absence of unnatural amino acids, which are only provided when grown in the laboratory[45]. Crucially, none of the approaches employed to provide biological containment have, to date, completely succeeded[46]; additionally, none of them have specifically addressed

the potential risk of recombinant spores shed into the environment. In this respect, it is expected that the development of multi-layered control systems incorporated into a single cell mitigate the probability of GMOs persisting in the environment[41].

One solution to provide an additional layer of biological containment in spore formers would be to create a strain that is conditional for sporulation, meaning a strain that can sporulate only under specific growth conditions. This conditionally sporulating strain could offer an ideal means of biocontainment for commercial applications; after fulfilling their purpose, e.g., as an antigen or probiotic vehicle, the engineered spores would germinate into vegetative cells with depleted ability to sporulate, and hence, be subject to cell death and DNA degradation (e.g., imposed by other layers of biocontainment). Additionally, biocontainment could be reinforced if non-pathogenic, anaerobic spore formers, such as those belonging to the genus *Clostridium* were used. These offer an intrinsic physiological restriction – their inability to grow in an aerobic environment.

1.3 *Clostridium*, a spore former relevant to biotechnology

Clostridium comprises more than 146 physiologically and genetically diverse Gram-positive, anaerobic, spore-forming bacteria[47]. The genus, notorious for the pathogenic properties of some of its members, includes several species relevant to a range of different biotechnological applications[48]. Representatives of the pathogenic branch of the genus include *Clostridium perfringens*[49], *Clostridium difficile*[50] and *Clostridium botulinum*[51], to which necrotizing infection, antibiotic-associated diarrhoea and the production of the

most potent natural toxin ever discovered, are attributed, respectively. However, the vast majority of clostridial species are non-pathogenic and many have been exploited at the industrial level. Applications of importance include the production of platform chemicals and biofuels (i.e., *Clostridium pasteurianum*[52,53], *C. acetobutylicum*[54]); cellulosic and hemicellulosic biomass degradation (i.e., *Clostridium cellulolyticum*[55]); carbon fixation (i.e., *Clostridium ljungdahlii*[56] and *Clostridium autoethanogenum*[57]); regulation of the microbiota (i.e., *C. butyricum*) and the previously described fight against cancer by *C. sporogenes*[30,32]. Regardless of their extraordinary diversity, deriving both a greater understanding of their biology, while at the same time enhancing their exploitable properties, offers a number of exciting possibilities for public and environmental benefit. Thus, addressing the critical problem of releasing spores of GMOs in the environment would facilitate the commercialization of genetically modified clostridial strains in the near future. Rewiring the sporulation pathway so that the ability to form spores (or not) is tightly controlled could be a potential solution, what leads the researcher to ask, Can a strain be engineered to conditionally sporulate?, and if so, How? Can synthetic biology provide the tools needed to engineer such functionality in a clostridial host?. As Herb Boyer, pioneer of genetic engineering said, "knowing the questions is knowing half" (quoted in Mukherjee, 2016[58]).

1.4 To Sporulate (or not)

A common characteristic of all clostridial species is their ability to form spores to avoid conditions of adversity[59]. Spores are multi-layered structures (Figure 1.3) whose overall resistance results from the sum of the specific

function of each layer[17]. The central core, enclosed by the spore inner membrane, is characterized by low water content, abundant small acid-soluble proteins and high levels of dipicolinic acid[60]. Such features, combined with the intrinsic structure of the spore proteins, allow the spore to ensure dormancy and confer resistance to some DNA-damaging agents, including genotoxic chemicals, desiccation, wet and dry heat, and UV and γ-radiation. Additionally, the coat, impermeable to large enzymes (i.e., lysozyme) and other biocidal chemicals, confers resistance via detoxification of such chemicals before they penetrate into the spore. The low permeability of the inner membrane and the presence of carotenoids in the outer membrane, contribute to resistance by restricting the access to some biocidal chemicals and by protecting the spore from UV radiation, respectively[17].

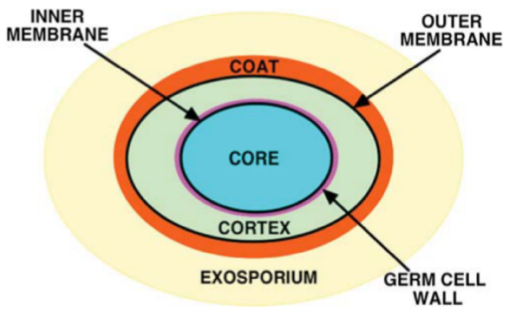


Figure 1.3 Diagrammatic representation of a typical bacterial endospore (not to scale; adapted from[17]).

In order to sporulate, the cell undergoes an ordered and complex program of cellular differentiation that leads to the dormant, metabolically-inert state[61]. The morphological stages of the sporulation program, conserved in all spore-forming bacteria, comprises seven stages (stage I-VII), as illustrated in

Figure.1.4. At the onset of sporulation (stage I), the cell carries two copies of the chromosome, one for the mother cell and one for the forespore. During stage II, the DNA forms an axial filament and the subsequent asymmetric cell division results in the formation of two compartments, the small pre-spore compartment and the large mother cell compartment. Of importance is the fact that not all the DNA is equally distributed between the mother cell and the pre-spore compartment at this stage. Formation of the sporulation septum results in enclosure of only one-third of the DNA in the pre-spore, allowing the differential expression of sporulation and vegetative-related genes in the pre-spore and mother cell, respectively; translocation of the rest of the chromosome into the pre-spore is then done via an active process requiring the translocase SpoIIIE[62]. Engulfment of the pre-spore by the mother cell characterizes stage III, where the now designated forespore is surrounded by the inner and outer membranes. In stage IV, the primordial cell wall (or peptidoglycan layer) and the cortex are formed. The spore coat, a complex structure of proteins surrounding the surface of the forespore is formed in stage V. Finally, the core of the spore is arranged so that resistance to heat and UV radiation is established (stage VI) and the mother cell lyses, liberating the spore into the environment (stage VII)[63,64].

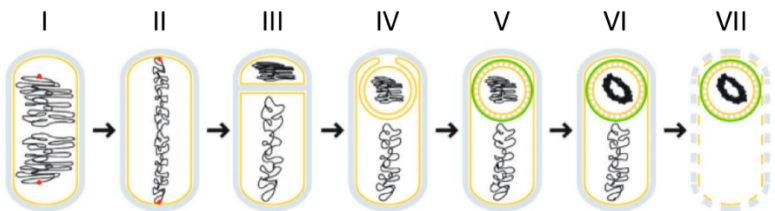


Figure.1.4 Schematic representation of the sporulation cycle in *B. subtilis*. Distinct stages of sporulation are denoted with a Roman numeral (I-VII). Adapted from[65].

Due to the availability of some techniques for genetic manipulation, scientists have been able to demonstrate the molecular bases of the sporulation process. Generally, this information has been obtained from the most studied spore-forming bacterium, *B. subtilis*; some features of the sporulation process, however, differ between *Clostridium* and *Bacillus* species. One common aspect in both genera is that the decision to enter the sporulation programme, albeit differently regulated and subject to multiple metabolic and environmental cues, is mediated by the conserved master regulator Spo0A[65,66]; its genetic inactivation has resulted in an asporogeneous phenotype in all species studied to date[67]. In *B. subtilis*, Spo0A is indirectly phosphorylated via a multicomponent "phosphorelay" whereby five histidine kinases (KinA-KinE), that respond to different environmental stresses, become phosphorylated and subsequently transfer the phosphoryl group to Spo0F. The phosphoryl group is then transferred to Spo0B, which mediates the phosphorylation of the master regulator Spo0A. Finally, the phosphorelay drains the phosphoryl groups via phosphatases that dephosphorylate Spo0F and Spo0A[65,68]. In *Clostridium*, on the other hand, no evidence for the existence of a phosphorelay have been yet found in any species; instead, activation of Spo0A seems to occur through direct interaction with multiple kinases[64]. Once initiated, the sporulation program is driven via the sporulation-specific sigma factors σ^F, σ^E, σ^G and σ^K, which regulate the temporal (different stages of sporulation) and spatial (different cell compartment) expression of sporulation- and vegetative cell-specific genes until the process is completed[69]. In *Bacillus*, the first sigma factor activated during the sporulation cascade is σ^F, which is regulated by Spo0A~P[69]. During stage I, the *spoIIAA* operon, which includes the genes encoding the sigma factors σ^F and

σ^E, is expressed in the forespore. During the course of stage II, activation of σ^F takes place in the pre-spore whereas σ^E is activated in the mother cell. In *Bacillus*, this is considered the point at which the cells are committed to sporulatation[70]. Activation of σ^E leads to the subsequent activation of σ^G within the spore, which in turn, is involved in the regulation of sporulation- and germination-related genes. Among these genes are those responsible for the activation of σ^K in the mother cell, which finally mediates the transcription of genes codifying coat- or germination-related genes[69]. In *C. difficile*, a recent study showed that more than 300 genes are induced by sporulation-specific sigma factors during spore formation[71]. Inactivation of such sigma factors in both, *C. difficile*[71] and *B. subtilis*[69], abrogates the formation of spores. However, the different phenotypes of the individual sigma factor mutants differed between these two firmicutes, indicating the role that evolution has played to fine-tune such a complex process.

One possible way of controlling 'when' and 'where' spores are formed, while maintaining a "healthy" phenotype in the vegetative form, is to fine-tune the complex sporulation cascade, so that the pathway is activated/completed in the presence or absence of an "input" (i.e., a ligand). From a genetic point of view, this applies to a system that mediates changes in gene expression levels and, as a consequence, establishes different (temporal or spatial) network connections within the pathway. As such, the main "ingredients" to engineer a strain with the ability to sporulate (or not) in a controlled manner include: (i) a tightly regulated inducible system that permits levels of key components of the sporulation cascade to be regulated so that the desired "output" is obtained (i.e., spores are formed or not), and (ii) identified key sporulation genes whose

activation/inactivation results in the desired sporogeneous/asporogeneous phenotype. Iterative cycles of design, building and testing of synthetic biology parts (i.e., promoters, operators) to create a device (i.e., a system capable of finely regulating the sporulation pathway) could lead to a strain with the desired functionality; a strain with conditional control of spore formation.

1.5 Tuning gene expression, a challenge in synthetic biology

Synthetic biologists often want to predictably tune the regulatory components of a system to elicit the desired output. When referring to ligand-inducible gene expression, this translates into a system exhibiting appropriate OFF and ON states, providing low leakiness and adequate activation in the absence and presence of the inducer, respectively. However, the dynamic range of natural inducible promoters is difficult to control. The most highly evolved and regulated expression systems in nature seem to be inevitably linked to promoter leakage (i.e., expression in the absence of inducer; also termed leakiness), which, in turn, benefits survival when encountering changes in the environment[72]. Synthetic regulatory systems, by contrast, can be specifically designed to overcome these natural constraints.

Transcription, translation and mRNA degradation play a crucial role in gene expression. All three processes are controlled by a combination of genetic elements, most notably promoters, Ribosome Binding Sites (RBSs) and *cis*-regulatory sequences located within the 5′ untranslated region (5′ UTR)[73]. One of the greatest barriers to designing an efficient synthetic regulatory system is the lack of standard genetic elements that can be reliably combined to obtain the desired functionality[74]. For this reason, several studies over the last few years

have focused on the standardization of such elements[75–79]; examples include the construction of large libraries of constitutive promoters[76], the characterization of libraries of natural and synthetic RBSs[78] and the manipulation of 5´ UTRs[79]. Additionally, constructing synthetic regulatory systems, which are often incorporated into a plasmid, requires other indispensable elements. These include: (i) origins of replication, preferably allowing a wide range of copies per cell[80]; (ii) insulators, which serve to counteract the effects that adjacent sequences might have in a specific biological part[81], and; (iii) transcriptional terminators, which prevent read-through transcription into the synthetic gene construct and insulate gene expression units[82]. As a consequence of such process standardization, synthetic regulatory systems have been successfully created for various species, allowing a predictable control over the dynamic range of the regulatory components[83]. However, well-established regulatory elements have been mostly characterized in model organisms (i.e., *E. coli, B. subtilis* and *Saccharomyces cerevisiae*) with only a limited collection of standard genetic elements available for non-model organisms, such as *Clostridium* spp. (recently reviewed in[84]).

Inducible systems capable of activating/repressing gene expression in response to small ligands might act at the level of transcription (i.e., LAC system[85]) or translation (i.e., theophylline responsive riboswitches[86]). Tuning gene expression in a transcriptional system can be achieved by changing the copy number of the plasmid encoding the transcription factor (TF)[87] or by modifying the elements involved in its regulation, generating a so-called synthetic inducible system. This has been achieved, for example, by mutating TFs to either reduce cross-induction from undesired metabolites[88] or to vary

their affinity for the operator sequences[89,90]. In most cases, improvements of transcriptional systems occur through the modification of the promoter, including inducible promoters with mutated operator sequences[91–93] or the presence of various operator sequences within the same inducible promoter[94]. However, the dynamic range of these synthetic inducible systems seems to be compromised when optimized for low levels of background expression; their outputs often show only a small difference between the OFF and ON states in response to the inducer.

Most recently, complex genetic circuits that respond to two or more TFs have demonstrated very tight input-output responses. For example, two operators within the same promoter that respond to different TFs (AND gate)[95], or various TFs that respond to different ligands that bind to the same operator (NOR gate)[96]. While these systems seem to be highly effective at controlling gene expression, tuning these regulatory pathways requires building and testing pre-characterized systems until the desired properties are empirically achieved, which are often functional only in the chassis strains in which they were developed. Additionally, due to the large number of elements required (i.e., various TFs and promoters), this approach can be extremely laborious and costly, making it unsuitable for strain engineering.

Compared to protein-based inducible systems, RNA regulators still suffer from very a low dynamic range[97]. In most cases, improvements of the dynamic range have been achieved in synthetic riboswitches that act at the transcriptional level[98]. In particular, transcriptional switches based on aptamers for theophylline and tetracycline showed increased dynamic range when placed in tandem upstream of a reporter gene[99,100]. However, the improved

performance is a direct consequence of having multiple termination points, that results in very low levels of gene expression in the presence of inducer and requires high doses of ligand to bind multiple aptamer copies. Additionally, swapping the order of the riboswitches also resulted in the system being responsive only with one of the two inducers, which indicates that even if the design of the individual elements is successful, their combination is far from trivial. Improvements of riboswitches regulating translation, on the other hand, have been mainly obtained by modifying the nucleotide residues located between the RBS and the translational start site[101]. Optimization of these riboswitches, however, results in very low gene expression levels if repression is tight, with higher expression levels inevitably linked to increased background expression.

In the case of *Clostridium*, there are only a limited number of methods available for inducing gene expression in a ligand-dependent manner. These include a lactose-inducible (LAC) system from *C. perfringens*[102], an arabinose-inducible (ARA) system, based on the ARAi regulon from *C. acetobutylicum*[103], the tetracycline-inducible (TET) system, originally adapted from *E.coli*[104] and a xylose-inducible (XYL) repressor system from *Staphylococcus xylosus*[105,106]. Despite evidence of a dose-dependent induction, the LAC, ARA and XYL systems are hampered by a significant level of gene expression in the absence of the inducer, compromising their dynamic range and making them unsuitable for applications where a tight control of gene expression is needed. On the other hand, the TET system exhibits very low basal expression and has the highest inducing efficiency among the reported inducible promoters investigated in *Clostridium spp.*[102–104,106–109] thus far. However, the optimal working conditions

of the TET system require high doses of the inducer anhydrotetracycline, a tetracycline-analogue. The levels required, however, have significant inhibitory effects on cell growth[104].

Apart from controlling sporulation, there are many appealing applications for tightly inducible systems functioning in a clostridial host. In any cell factory, environmental signals are integrated into a novel biological behaviour. Inducible systems are key components of the synthetic pathways that modulate such behaviour, refining enzyme and, therefore, pathway function. Just as the length and thickness of a wire determine its resistance and the output of an electric device, inducible systems serve to modulate a genetic (and metabolic) network so that a valuable technological function is obtained. Moreover, inducible genetic circuits are very useful for basic science applications, where hypotheses can be tested by mimicking natural processes in a simplified context or by controlling the individual components of the network at a time.

1.6 The CRISPR Revolution

CRISPR, which is short for "clustered regularly interspaced palindromic repeats", is definitely revolutionizing the synthetic biology industry. Discovered as a "strange" microbial repeat sequence in 1993[110], CRISPR has taken the scientific community by storm, and 25 years after its discovery, thousands of researchers are using CRISPR-based technologies for applications that range from agriculture to industrial biotechnology and biomedicine. CRISPR, composed of direct repeat sequences of 21-37 nucleotides, interspaced by similarly sized non-repetitive sequences (spacers) and flanked on one side by a

common 300-500 nucleotide leader sequence, are included within the genome of diverse bacteria and archaea[111]. The finding that CRISPR-associated (Cas) genes were shown to contain helicase and exonuclease domains[111], that the identical CRISPR repeats showed sequence homology to bacteriophages and conjugative plasmids[112,113] and that a phage-resistance phenotype could be altered by modifying the CRISPR spacer sequences[114], revealed not only its nature as a defence mechanism, but also possible ways of exploiting it.

Specifically, the effectiveness of mutant isolation by allelic exchange has been revolutionized through the use of CRISPR/Cas to effectively counter select non-mutant WT cells through the introduction of double-strand breaks (DSBs) in the DNA[115,116]. In particular, the type II CRISPR-Cas9 system derived from *Streptococcus pyogenes*, has been the most extensively CRISPR system employed for genome editing purposes[115]. The basis of the Cas9 editing tools relies on the fact that Cas9 nucleases can be guided to a specific DNA target sequence, termed protospacer, by a CRISPR RNA (crRNA):trans-activating crRNA (tracrRNA) hybrid. Accordingly, the hybrid CRISPR RNA (crRNA):trans-activating crRNA (tracrRNA), or the simplified chimeric synthetic single guide RNA (sgRNA), combines with Cas9 to form a ribonucleoprotein complex. This complex then recognizes the target site, based on the protospacer adjacent motif (PAM) sequence, and induces a DNA DSB. In most bacteria, selection of mutant cells is achieved when the DNA editing template, which lacks the recognition site, is introduced into the genome via homologous recombination, enabling these mutant cells to “escape” from the cutting activity of Cas9. Additionally, DSBs can be repaired by the non-

homologous end joining pathway, resulting in random sized deletions in the vicinity of the targeted DNA sequence[117].

The beauty of CRISPR-Cas9 technologies, if compared with other genome editing technologies such as zinc finger nucleases (ZFNs)[118] and transcription activator-like effector nucleases (TALENs)[119], is that it allows the modification of the genomic target by simply replacing the 20 nucleotides of the chimeric sgRNA complementary to the target sequence of interest. For this reason, CRISPR-Cas9 technologies have been exploited for purposes other than the precise deletion, replacement and insertion of genomic DNA. Examples include CRISPR interference (CRISPRi), where gene expression can be regulated by using a nuclease-deactivated Cas9 (dCas9), which only maintains its DNA-binding function, to target the promoter region of a gene, and CRISPR activation (CRISPRa), which can activate gene expression by fusing dCas9 to a transcriptional activator[120].

Various CRISPR-based systems have now been described and employed to edit prokaryotic and eukaryotic cells. However, despite the extraordinary power and potential applications of the technique, there is concern about CRISPR altering regions of the genome inadvertently[121]. Whilst such concern mainly refers to the potential pathogenic consequences CRISPR-Cas9 editing might have in clinical applications, microbial synthetic biology can also suffer from lack of CRISPR-Cas9 specificity. When a synthetic biological pathway is built, further modifications that could affect its designed function must be prevented. Therefore, alongside the search for more precise gene editing techniques, improved methods that confer a more accurate control of CRISPR-Cas9 technologies must be developed.

1.7 Aims of this study

Genetically modified bacterial spores offer a great number of exciting applications in the field of biotechnology. However, the problem for containing recombinant spores in the environment remains unsolved. In this study, a synthetic biology approach is applied to solve this problem. Specifically, *C. sporogenes*, the most prominent candidate for clostridial-based cancer therapies, is engineered so that it only forms spores under specific growth conditions. To achieve this, tightly regulated inducible systems, which overcome the leakiness and low dynamic range associated with natural systems, are generated and implemented not only in the construction of a conditionally sporulating strain, but also in the creation of a novel Cas9-based editing tool for *Clostridium.*

In **Chapter 3**, one of the best inducible systems hitherto tested in clostridial organisms, the LAC system from *C. perfringens* strain 13, is exhaustively characterized with the aim of its further optimization to reduce its basal expression and maintain a wide dynamic range. **Chapter 4** describes that previously existing, as well as newly designed, theophylline responsive riboswitches function in *C. sporogenes*, being consolidated as novel devices to regulate gene expression in clostridia. These data are exploited together in **Chapter 5**, which describes the design and implementation of a synthetic dual mechanism to regulate gene expression. In **Chapter 6**, the systems characterized in the previous chapters are used to generate conditionally sporulating strains via regulating the expression of genes that are essential for spore formation. Post-hoc testing of the engineered strain provides clues to how sporulation functions in *C. sporogenes*. Finally, and to develop new mutagenesis tools to

facilitate synthetic biology approaches in clostridial chassis, a novel and universal Cas9-based editing tool is created, which is the focus of **Chapter 7**. Conclusions and future perspectives of this work are outlined in **Chapter 8**.

Chapter 2
/ Materials and Methods

2.1 Chemicals and suppliers

A full list of the chemicals and kits used in this study, including supplier and catalogue number, is shown in the Appendix, Table S1.

2.2 General growth methods

2.2.1 Growth media composition

A list of the composition of the growth media used in this study is given in the Appendix, Table S2.

2.2.2 Preparation of media

All media and buffers were prepared in dH_2O and autoclaved for 20 minutes at 121°C or sterilized by filtration through a 0.2 μm pore size filter (Ministart®). Medium 2xYPG was adjusted to pH 6.2 prior to autoclaving and sterile 50% w/v glucose solution was added subsequent to autoclaving. Agar media were allowed to cool to 50°C in a water bath before the addition of any supplement. Supplements were added to the indicated working concentrations and the melted agar poured into 9 cm Petri dishes, with approximately 20 mL per plate. Plates were dried at 37°C for 1 hour before use.

2.2.3 Preparation of media supplements

All supplements were prepared as stock solutions and added to growth medium to give the final concentrations described in Table 2.1. Stock solutions were sterilized by filtration through a 0.2 μm pore size filter (Minisart®).

Table 2.1 Supplements used in this study

Supplement	Stock concentration (mg/mL)	Solvent*	Working concentration for *E. coli* (μg/mL)	Working concentration for *Clostridium* (μg/ml)
Ampicillin	50	H_2O	100	-
Cloramphenicol	25	100% ethanol	12.5 (broth) 25 (agar)	-
D-Cycloserine	50	H_2O	-	250
Erythromycin	50	100% ethanol	500	10
Thiamphenicol	15	50% ethanol	-	15
5-Fluoroorotic acid (5-FOA)	200	DMSO	-	4500
Theophylline	48	DMSO	-	0.1-10 mM
Lactose	100 mM	H_2O	-	0.1-10 mM

*DMSO (dimethyl sulphoxide)

2.3 Culture and manipulation of *E. coli* strains

2.3.1 E. coli strains

A detailed list of all *E. coli* strains used in this study can be found in the Appendix, Table S3.

2.3.2 Culture of E. coli cells

E. coli strains were cultured aerobically at 37°C on lysogeny broth (LB), both in liquid and on solid agar plates. Cultures were incubated overnight (approximately for 16 hours) at 30 or 37°C with shaking at 200 rpm for broths. *E. coli* cultures on solid medium were prepared by a variety of methods. For generation of biomass (e.g., -80°C stocks), a 10 μL loopful of liquid culture or a single colony of the strain was re-streaked onto LB-medium. For isolation of single colonies, cells suspensions were pipetted onto LB-agar and immediately spread using a sterile plastic spreader. Plates were typically incubated for 1-3

days depending on the strain and cargo. Media were supplemented with appropriate antibiotics for plasmid selection.

2.3.3 Storage of E. coli strains

E. coli strains were stored at -80°C using the Microbank™ Long Term Bacterial Storage vials (Pro-lab Diagnostics). The strain of interest was streaked onto a plate of LB-agar with appropriate antibiotics. After substantial growth was obtained, cells were scraped from the plate and used to inoculate the Microbank™ vials, in triplicates.

2.3.4 Preparation of electro-competent E. coli cells

Overnight cultures of *E. coli* TOP10 and CA434 were prepared by addition of a single bead from a Microbank™ (or single loop of culture) to 5 mL LB medium supplemented with relevant antibiotics. The culture was incubated overnight at 37°C with shaking at 200 rpm and 1 mL of the seed culture was used to inoculate 200 mL of sterile, pre-warmed LB broth (37°C) supplemented with appropriate antibiotics. Flasks were incubated at 37°C with shaking at 200 rpm until an OD_{600} of 0.5-0.7 was reached. Cell cultures were transferred to pre-chilled Falcon tubes and centrifuged at 6000 x g for 10 minutes at 4°C, and supernatants were discarded. Pellets were subsequently washed by gentle resuspension in 20 mL of sterile, ice-cold dH_2O and centrifuged under the same conditions. This washing step was repeated three times while combining all pellets into only one Falcon tube. Finally, the pellet was resuspended in 1 mL of 1 mM MOPS with 10% (v/v) glycerol. The cells were aliquoted into 50 μL samples and stored at -80°C.

2.3.5 Transformation of electro-competent E. coli cells

For each electroporation, an aliquot of frozen electro-competent *E. coli* cells was thawed on ice for 10 minutes. Dialyzed (0.025 μM Merck Millipore dialysis discs over dH_2O for at least 20 minutes) or purified plasmid DNA (0.5-1 μg) was added to the cells and mixed by pipetting. The DNA-cell mixture was then transferred to a pre-chilled electroporation cuvette (Flowgen Biosciences, 0.2 cm) and pulsed with conditions set to 22.5 kV/cm, 200Ω and 25 μF using the Bio-Rad Gene Pulser. Cells were immediately resuspended in 400 μL SOC medium, transferred to a 1.5 mL microcentrifuge tube and incubated at 37°C under agitation for 1 hour (or 30°C for 2 hours). After incubation, suitable dilutions of cells were plated onto LB agar medium supplemented with the appropriate antibiotic and incubated as required.

2.4 Culture and manipulation of *Clostridium* strains

2.4.1 Clostridium strains

A detailed list of all *Clostridium* strains used in this study is given in the Appendix, Table S4.

2.4.2 Clostridium cell culture

All work involving *Clostridium* cultures was carried out in a MG1000 Mark II Anaerobic Work Station (Don Whitley Scientific Ltd) with an atmosphere of 80% nitrogen, 10% hydrogen, and 10% carbon dioxide; a humidity of 75% and a temperature of 37°C. All media and reagents were reduced prior to use for a minimum of 4 hours in the anaerobic workstation.

C. sporogenes and *C. botulinum* were cultured on TYG or defined MACC media[122], both in liquid and on solid agar plates. *C. difficile* was cultivated on BHIS medium, both in liquid and on solid agar plates. *C. pasteurianum* was cultivated on 2xYPG liquid broth and RCM-agar plates.

2.4.3 Storage of Clostridium strains

All *Clostridium* strains except *C. pasteurianum* were stored as glycerol stocks made with their specific medium containing 10-20% (v/v) glycerol, frozen at -80°C. *C. pasteurianum* was stored with medium containing 10% (v/v) DMSO. Strains were recovered from the stocks by streaking of a loop of frozen suspension onto the specific medium, supplemented with appropriate antibiotics and incubated in the anaerobic cabinet overnight.

2.4.4 Methylation of plasmid DNA

Plasmids intended for transformation into *C. pasteurianum* were methylated by transformation into *E. coli* pCR1[53]. *E. coli* transformations were carried out as previously described. DNA was purified from cells grown on selective media.

2.4.5 Preparation of electrocompetent Clostridium cells

To produce *C. pasteurianum* electrocompetent cells, a seed culture was first prepared by inoculating 10 mL of reduced 2xYPG with approximately 200 μL of a thawed DMSO frozen stock and two 10-fold dilutions were made. Broths were incubated overnight. An overnight culture that had not reached the stationary growth phase (generally the most diluted culture), was used to

inoculate 200 mL anaerobic 2xYPG liquid broth to an initial OD_{600} of approximately 0.05. Cells were grown to early exponential phase (OD_{600} of 0.3-0.4), at which time 75 mL of a 2 M sucrose solution and 26.7 mL of a 18.8% (w/v) glycine solution were added. Growth was resumed until the culture attained an OD_{600} of between 0.6 and 0.7 (approximately 2-3 hours). At this point, the culture was divided into six 50 mL centrifuge tubes and centrifuged at 6000 x g for 10 minutes at 4°C. Pellets were washed by resuspension in 15 mL ice-cold anaerobic sodium phosphate buffer (SPM) and subjected to the same centrifugation step. SPM was prepared by mixing 1 mL filter sterilized 500 mM sodium phosphate solution pH 6.5, 1 mL filter-sterilised 100 mM magnesium chloride solution and 13.5 mL 2 M sucrose solution. Pellets were resuspended in 1.5 mL SPM and used immediately. If stored, electrocompetent cells were aliquoted with the addition of 10% (v/v) DMSO into 2 mL screw-cap micro tubes and frozen at -80°C.

2.4.6 Transformation of Plasmid DNA from *E. coli* host into clostridia by conjugation

C. sporogenes, *C. botulinum* and *C. difficile* were generally transformed via conjugation. For such conjugation, 5 mL of LB containing appropriate antibiotic selection were inoculated with *E. coli* CA434 transformed cells and incubated overnight. Similarly, TYG (if the recipient was *C. sporogenes* or *C. botulinum*) or BHIS (if the recipient was *C. difficile*) was inoculated with the clostridial recipient strains and incubated overnight. The following day, 1 mL of the donor *E. coli* culture was centrifuged for 1 minute at 6000 x g and the pellet was washed by gentle resuspension in phosphate buffer saline (PBS) and

centrifuged again under the same conditions. The *E. coli* pellet was then transferred to the anaerobic workstation and resuspended in 200 μL of *Clostridium* culture. The mixture of donor and recipient was spotted onto a plain TYG or BHIS plate and incubated at 37°C for at least 7 hours. Cells were harvested from the mating plate and resuspended in 500 μL PBS. Finally, cells were plated on media supplemented with appropriate antibiotic for the selection of the the transferred plasmid and D-cycloserine to counter select the *E. coli* donors. After 24-48 hours of incubation, random cells that had grown on selected medium were PCR screened to confirm the plasmid had been transferred.

2.4.7 Transformation of Plasmid DNA into clostridia by electroporation

For transfer of plasmids to *C. pasteurianum*, methylated plasmid DNA (4-5 μg) was introduced to a pre-chilled, 4 mm wide, electroporation cuvette (Flowgen Biosciences). *C. pasteurianum* electrocompetent cells (570 μL) were added to the cuvette and incubated for 2 minutes on ice. The cell-DNA mixture was then electroporated using a BioRad GenePulser Xcell^{TM} electroporator with conditions 2.0kV, 25μF and ∞Ω. Immediately following pulse delivery, 1 mL warm (37°C) 2xYPG medium was added to the cuvette and the suspension was transferred to a tube containing 9 mL of the same medium. After a recovery period of 3-4 hours, volumes of 100 μL of neat culture, and tof wo 10-fold dilutions of this, were plated into RCM plates containing the appropriate antibiotic.

2.4.8 Plasmid curing

To cure the plasmid, the plasmid-harbouring mutants were cultivated into appropriate non-selective liquid media for 24-48 hours. Then, appropriate dilutions were plated onto agar to obtain discrete colonies. Colonies were randomly picked and patch-plated onto agar plates in the presence and absence of selection. Colonies susceptible to the antibiotic were PCR screened to confirm the plasmid had been loss. The efficiency of plasmid curing was calculated by dividing the total number of colonies tested by the number of selection-sensitive colonies.

2.5 Molecular methods and techniques

2.5.1 Extraction and purification of genomic DNA

For general purposes, genomic DNA was extracted and purified from *Clostridium* strains using the Sigma-Aldrich GenElute^{TM} Bacterial Genomic DNA Kit in accordance with the manufacturer's instructions.

For whole genome sequencing, *C. sporogenes* genomic DNA extraction was carried out according to the phenol:choloform:isoamyl alcohol (25:24:1) method. To do so, 2 mL cultures of the target strain were removed from the anaerobic cabinet after overnight incubation and centrifuged at 12,000 x g for 3 minutes. Pellets were resuspended in 180 μL lysis buffer (PBS supplemented with 10 mg/mL lysozyme) followed by incubation at 37°C for 30 minutes, undergoing periodic gentle agitation. Then, 25 μL proteinase K solution, 85 μL dH_2O and 110 μL 10% (w/v) SDS solution were added, mixed by inversion and incubated at 65°C for 30 minutes. A volume of 400 μL of

phenol:chloroform:isoamyl alcohol (25:24:1), saturated with 10 mM Tris, pH 8.0, 1 mM EDTA was added to the samples. After thoroughly mixing, the mixture was transferred to phase-lock tubes and centrifuged for 3 minutes at 12,000 x g. After centrifugation, the upper layer was transferred to a fresh phase-lock tube repeating the phenol:cloroform:isoamyl alcohol and centrifugation process a further 2 times. The resultant top layer of supernatant was transferred to a microcentrifuge tube containing 40 μL of 3 M sodium acetate and 800 μL of ice cold 100% ethanol. The samples were gently mixed by inversion and incubated at -80°C for 30 minutes, followed by 15 minutes of centrifugation at 12000 x g. Pellets were washed by gently addition of 70% ethanol and centrifuged for 3 minutes at 12000 x g, followed by removal of the remaining supernatant. Finally, samples were left to air-dry at room temperature and DNA was resuspended in 50 μL of diethylpyrocarbonate (DEPC) treated water and quantified on the NanoDrop (Thermo Fisher Scientific).

2.5.2 Extraction and purification of plasmid DNA

Plasmid DNA was extracted from overnight *E. coli* TOP10 cultures using the Monarch Plasmid Miniprep Kit (New England Biolabs) following the manufacturer's instructions.

2.5.3 DNA amplification

2.5.3.1 Oligonucleotide primers

A detailed list of all the primers used in this study is provided in the Appendix, Table S6.

2.5.3.2 Polymerase Chain Reaction

All sequences of interest were amplified using KOD Hot Start Master Mix or Expand High Fidelity PCR System if high yield and fidelity were required. All PCR reactions were performed according to the manufacturer's instructions in a thermocycler block (Applied Biosystems Gene Amp PCR system 2400).

2.5.3.3 Colony PCR

For rapid screening, single colonies were picked from agar plates and resuspended in 10 μL of sterile dH_2O. Generally, 1 μL of this colony suspension was added to the 2xDream Taq Green PCR cocktail, and PCR performed according to the manufacturer's instructions.

2.5.3.4 Splicing by overlap extension (SOE) PCR

All knockout cassettes, as well as some constructs requiring DNA deletions/replacements were generated by SOE PCR. The procedure has been previously outlined[123] and is illustrated in Figure 2.1. The basic principle relies on the design of oligonucleotides with overlapping sections, allowing the deletion of specific sequences in the DNA.

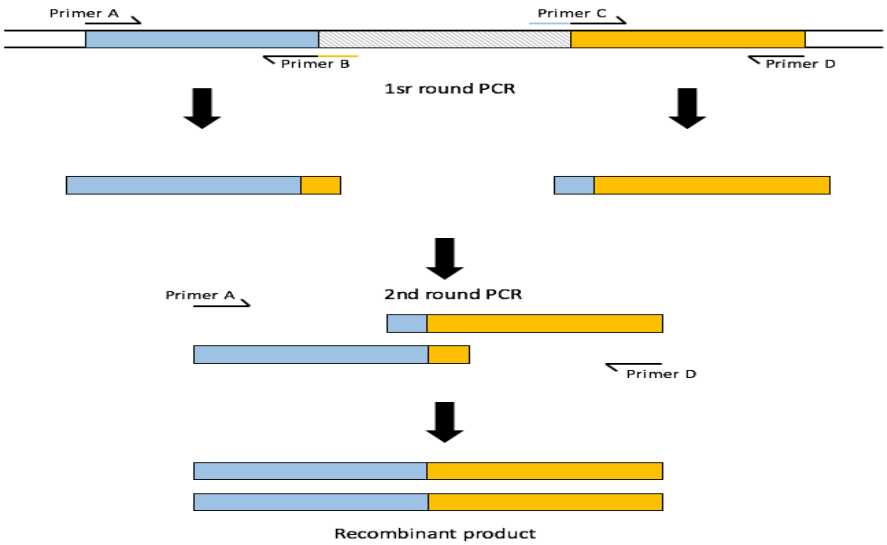


Figure 2.1 Schematic representation of SOE PCR. Primers A and D are normally designed to incorporate restriction sites for further cloning. Primers B and C are designed with overlapping regions to each other at their 5′ and 3′ terminals, respectively (> 10 nucleotides in length). Following the amplification of the two DNA fragments in the 1st PCR round, these are combined in a 2nd PCR along with primers A and D. The region of homology anneals and the final recombinant product is completed by PCR following the annealing of the primers.

2.5.4 Agarose gel electrophoresis and DNA extraction

All PCR products and restriction fragments were separated by performing electrophoresis in 1-1.5% w/v agarose dissolved in TAE buffer (40 mM Tris, 1 mM EDTA, 0.1% v/v glacial acetic acid) containing SYBR™ Safe DNA gel stain. DNA samples were combined with 6X loading dye and run at 5-10 volts per cm of distance between the two electrodes in TAE buffer. The separated DNA was subsequently visualized under blue light and gel images obtained using a Gel Doc™ XR+ system (Bio Rad).

If needed, DNA fragments were extracted from excised agarose gel pieces using the Monarch Gel Extraction Kit, following the manufacturer's instructions.

2.5.5 Restriction endonuclease digestion

Double stranded DNA was specifically cleaved, and when needed de-phosphorylated/phosphorylated, with restriction enzymes and Antarctic phosphatase/T4 polynucleotide kinase respectively, following the manufacturer's instructions.

2.5.6 DNA ligation

The assembly of DNA fragments and plasmid construction was carried out using T4 DNA ligase. Linearized fragments were combined in a reaction along with 1 μL T4 DNA ligase and 1X ligase buffers, made up to a total reaction volume of 10 μL with DEPC treated water. Ligations were placed on thawing ice overnight. Ligation mixtures were dialysed using 0.025 μm Millipore dialysis discs over dH_2O for 20 minutes prior to electroporation.

2.5.7 DNA sequencing

All PCR and cloned products were verified by nucleotide sequencing performed by Eurofins (Germany), using supplied oligonucleotide primers.

Genome sequencing was provided by MicrobesNG, using their Illumina-based, Standard Whole-Genome Service (http://www.microbesng.uk). Whole-genome sequencing data were assembled and analysed using the CLC Genomics Workbench version 10 (Qiagen). Resequencing analysis for the identification of single nucleotide variants (SNVs) was carried out using the Basic Variant Detection tool and the program's default parameters.

2.5.8 RNA extraction

For the amplification of 5´ UTRs, RNA was extracted from 2 mL overnight *C. sporogenes* cultures, grown in TYG, using the High Pure RNA Isolation Kit, following the manufacturer's instructions.

For quantitative reverse transcription PCR analysis (RT-qPCR), prior to RNA extraction, 2 mL of *C. sporogenes* cultures grown in TYG to early stationary phase ($OD_{600} \approx 2.5$) were mixed with 4 mL of RNA Protect Bacteria reagent. Total RNA was extracted using the FastRNA Pro Blue kit, according to the manufacturer's instructions. Purified RNA was DNase-treated using the RQ1 RNase-Free DNase kit. RNA integrity was determined using a 2100 Bioanalyzer (Agilent) and the concentration determined with a NanoDrop (Thermo Fisher Scientific).

2.5.9 Amplification of 5'UTR

Poly-T-cDNA was obtained according to an approach termed 5´-Rapid Amplification of cDNA Ends (5´ RACE), using the 5´ RACE kit 2nd Generation and the Expand High Fidelity PCR System as per manufacturer's instructions. The kit requires three specific primers (IC263-3, IC264-r and IC265-r), all of them complementary to the mRNA transcript (*catP*). After amplification, the transcriptional start site (TSS) was determined according to 5´ UTR and the putative -10 and -35 boxes on the basis of typical characteristics of bacterial promoters.

2.5.10 RT-qPCR

cDNA synthesis was performed on 1 µg of RNA using the Omniscript RT Kit, primed with random hexamer primers. qRT-PCR analyses were performed with 5 µL of 1:10 diluted cDNA mixtures using the Power SYBR Green Master Mix on a Light Cycler 480 II (Roche). cDNA synthesis reactions lacking reverse transcriptase were performed as a control for genomic DNA contamination. All calculations were automatically obtained using the LightCycler® 480 software, version 1.5. The efficiency of the primers was calculated before used. qRT-PCR was performed on cDNA isolated from three biological replicates and in technical duplicates for each cDNA sample and primer pair. Results were calculated according to the E-method[124] and normalized to the *16Srrn* and *gyrA* genes.

2.5.11 Deletions or integrations via Allelic Exchange

For genome editing via allelic exchange, knockout cassettes were generated by SOE PCR and cloned into the pMTL-ICD vector. Plasmid constructs were transferred into *C. sporogenes ΔpyrE* from the respective *E. coli* donors by conjugative plasmid transfer and transconjugants were selected by acquired resistance to thiamphenicol. Single crossover integrants at either homologous region were identified by their increased growth rate, exhibiting larger colony sizes after incubation for 24-48 hours. Colonies were streaked to purity and screened via colony PCR to confirm the integration event. The primers used for the screening were specific combinations of one primer aligning to the chromosome and a second primer aligning to the deletion vector. Following confirmation, selected integrants were streaked on TYG agar plates

containing 4.5 mg/ml fluoroorotic acid. Colonies were selected, re-streaked for purity onto TGY agar plates and confirmed via colony PCR. On this occasion, both primers annealed to the chromosome, 100-200 bp upstream and downstream of the site of deletion (or integration). Plasmid loss was confirmed by replica plating onto TYG agar supplemented with thiamphenicol and by colony PCR. In all cases, three biological replicas of the mutant strain were obtained from three independent conjugations and the identity of the mutants confirmed by Sanger sequencing.

2.5.12 Chromosomal integration and pyrE restoration via Allele-Coupled Exchange (ACE)

The ACE method[125] was used to integrate DNA into the genome of *C. sporogenes* concomitant with *pyrE* restoration (and uracil prototrophy). The integration vectors pMTL-JH29 and pMTL-JH29T, with the specific "cargos", were transferred into *C. sporogenes ΔpyrE* from the respective *E. coli* donors by conjugative plasmid transfer and transconjugants were selected by acquired resistance to thiamphenicol and D-cycloserine. After 24-48 hours of incubation under anaerobic conditions, a loopful of transconjugant cells was grown overnight in liquid TYG medium supplemented with thiamphenicol and D-cycloserine. Overnight cultures were centrifuged at 6000 x g for 10 minutes at room temperature, pellets resuspended in 500 μL PBS and appropriate dilutions plated onto defined MACC agar plates lacking uracil. Mutants in which *pyrE* had been restored were selected by their acquired ability to grow on medium lacking uracil. Colonies were re-streaked for purity onto TGY agar plates and colony PCR screened. On this occasion, both primers annealed to the

chromosome, 100-200 bp upstream and downstream of the *pyrE* locus. Plasmid loss was confirmed by replica plating onto TYG agar supplemented with thiamphenicol and by colony PCR. In all cases, three biological replicas of the mutant strain were obtained from three independent conjugations and the identity of the mutants confirmed by Sanger sequencing.

2.5.13 Genome editing via CRISPR (RiboCas)

For genome editing purposes using CRISPR, the RiboCas vectors (pRXCasN), with the specific sgRNA and/or editing templates, were transferred into the clostridial host by conjugative plasmid transfer (*C. sporogenes, C. difficile and C. botulinum*) or by electroporation (*C. pasteurianum*) and transformants were selected by acquired resistance to thiamphenicol (and D-cycloserine if transforming by conjugation). After 24-48 hours of incubation under anaerobic conditions, colonies were picked and transferred to selective plates in the presence of theophylline. Cells with the ability to grow on plates containing the inducer theophylline were randomly picked and screened for desired mutations using colony PCR and specific flanking primers. Plasmid loss was confirmed by replica plating onto TYG agar supplemented with thiamphenicol and by colony PCR. In all cases, three biological replicas of the mutant strain were obtained from three independent conjugations and the identity of the mutants confirmed by Sanger sequencing.

2.6 Cloning of plasmid constructs

For a full list of plasmid constructs and resulting strains, see Appendix, Tables S4 and S5. Maps of the different constructs and primers used for sequencing can be found in the Appendix, Figures S1 to S6.

2.6.1 Construction of plasmids used in Chapter 3

pMTL-IC001

To create the promoter-less reporter plasmid pMTL-IC001, *catP* was amplified with primers IC105-f and IC103-r from pMTL83151[126] and inserted into pMTL82251[126] through the NotI and XhoI restriction sites.

pMTL-IC1201

P_{bgaR} was amplified from pMTL-HZ1 with primers IC301-f and IC302-r, to introduce the restriction sites NotI and BsaI, respectively. *catP* was amplified from pMTL83151 with primers IC103-f and IC104-r, which introduce the restriction sites BsaI and XhoI, respectively. Promoter and gene were ligated in a scar-less manner using the restriction site BsaI and introduced into the reporter plasmid pMTL82251 via the restriction sites NotI and XhoI.

pMTL-ICMO1, pMTL-ICMO2 and pMTL-ICMO1O2

pMTL-ICMO1 and pMTL-ICMO2, where the sequences O1 or O2 are deleted from the LAC system, respectively, were generated via inverse PCR using pMTL-HZ1 as a template and the pairs of primers IC303-f/Ic304-r and IC305-f/IC306-r for O1 and O2, respectively. After amplification, the PCR products were digested with the restriction enzyme DpnI, which digests only at specific methylated sites and serves to eliminate the template vector. pMTL-ICMO1O2, with deletions of both O1 and O2, was generated in a similar manner but using pMTL-ICMO1 as a template and the primers used to create pMTL-ICMO2.

pMTL-ICMB

Removal of *bgaR* from the plasmid harbouring the LAC system (pMTL-HZ1) was carried out by amplifying the intergenic region of the LAC reporter system, up to the translational start site of the reporter *catP*, using primers IC307-f and IC308-3, which introduced the restriction sites NotI and NdeI, respectively. The amplified fragment was inserted into the NotI/NdeI linearized pMTL-HZ1 vector.

pMTL-ICT1 to pMTL-ICT11 and pMTL-ICM1 to pMTL-ICM11

Vectors pMTL-ICT1 to pMTL-ICT11 and pMTL-ICM1 to pMTL-ICM11 were generated by stepwise deleting or otherwise mutating blocks of five nucleotides between positions -89 and -34, upstream of P_{bgaL}, within the intergenic region of the LAC reporter system. In all cases, the truncated/mutated intergenic region was generated by ligating two DNA fragments via the restriction enzyme BsaI. One of the two fragments, amplified from pMTL-HZ1 with primers IC309-f and IC310-r, which insert the NotI and BsaI restriction sites, respectively, was common to all the ligations. This DNA fragment included *bgaR* and its native terminator, as it is in pMTL-HZ1, and part of the intergenic region between *bgaR* and *catP*. The second DNA fragment, comprising the remaining part of the intergenic region, is specific for each truncated/mutated variant, and this was amplified using the forward primers IC311-f, IC312-f, IC313-f, IC314-f, IC315-f, IC316-f, IC317-f, IC318-f, IC319-f, IC320-f, IC321-f, IC322-f, IC323-f, IC324-f, IC325-f, IC326-f, IC327-f, IC328-f, IC329-f, IC330-f, IC331-f or IC332-f for pMTL-ICT1, -T2, -T3, -T4, -T5, -T6, -T7, -T8, -T9, -T10, –T11, -M1, -M2, -M3, -M4, -M5, -M6, -M7, -M8, -M9, -M10 or -M11, respectively, and the reverse primer IC308-r. The forward

and reverse primers included the restriction sites BsaI and NdeI, respectively. Both fragments were then ligated by the BsaI restriction site and introduced into the pMTL-HZ1 backbone using the restriction sites NotI and NdeI.

2.6.2 Construction of plasmids used in Chapter 4

pMTL-IC101, pMTL-IC201 and vectors carrying theophylline responsive riboswitches

In order to compare theophylline responsive riboswitches, the modular vector pMTL82251 was used as a chassis for all the vectors. In all cases, the promoters (P_{fdx}, P_{fdx}*, P_{fdx4} and P_{fdx4}*), the different riboswitches and the reporter *catP* were assembled using the restriction endonuclease BsaI and introduced into the application-specific module via NotI and XhoI restriction sites. Oligonucleotides IC103-f and IC104-r were used to amplify *catP* from pMTL83151. Riboswitches were amplified with the forward primer IC107-f and the reverse IC108-r, IC109-r, IC110-r, IC111-r, IC112-r, IC113-r, IC114-r and IC115-r for riboswitches -D, -E, -E*, -F, -G, -H, -I and –J, respectively. P_{fdx} was amplified using genomic DNA from *C. sporogenes* NCIMB 10696 as a template and the pairs of primers IC101-f/IC102-r and IC101-f/IC106-r when placed upstream of *catP* (pMTL-IC101) or a riboswitch (pMTL-IC111-D to –J), respectively. P_{fdx4} was generated by SOE PCR using primer pairs IC101-f/IC119-r and IC118-f/IC102-r and pMTL-IC101 as a template; P_{fdx4} was inserted into the backbone similarly to P_{fdx} generating the vectors pMTL-IC201 and pMTL-211-E, -G and -H. P_{fdx}* (for vectors pMTL-121-E, -G and -H) and P_{fdx4} * (for vectors pMTL-221-E, -G and -H) were amplified using the primers IC101-f and IC116-r and pMTL-IC101 and pMTL-IC201 as templates,

respectively; riboswitches were amplified using the forward primer IC117-f and the same reverse primers listed above.

2.6.3 Construction of plasmids used in Chapter 5

pMTL-ICE1, pMTL-ICG1, pMTL-ICH1

Vectors containing RiboLacs where riboswitches were placed downstream of P_{bgaR} were constructed by fusing three DNA sequences via the restriction enzyme BsaI. The first DNA fragment comprises the *bgaR* coding sequence and its native terminator; this was amplified from pMTL-HZ1 using the primers pair IC307-f and IC333-r, which introduced the restriction sites NotI and BsaI, respectively. The second DNA fragment contains the riboswitch, which was amplified with the forward primer IC107-f and the reverse primers IC334-r, IC335-r and IC336-r, for riboswitch-E (pMTL-ICE1), -G (pMTL-ICG1) and -H (pMTL-ICH1), respectively. This amplification was done by aligning the two oligonucleotide primers in the absence of DNA template. The forward primer includes a BsaI restriction site compatible with fragment one and the reverse primer includes a BsaI restriction site compatible with fragment three. The third DNA fragment comprises P_{bgaR}, the region between P_{bgaR} and P_{bgaL}, P_{bgaL} and the downstream *catP* coding sequence. This fragment was amplified with primers IC337-f and IC308-r, which introduced the restriction sites BsaI and NdeI, respectively, using pMTL-HZ1 as a template. The three fragments were ligated via compatible BsaI restriction sites and introduced into the linearized pMTL-HZ1 through the restriction sites NotI and NdeI.

pMTL-ICE1*, pMTL-ICG1*, pMTL-ICH1*

Vectors containing RiboLacs where P_{bgaR}* is placed upstream of the riboswitch were generated in a similar fashion to those containing P_{bgaR}; that is, by fusing three DNA sequences via the restriction endonuclease BsaI. The first and second sequences were identical to the ones described in the previous section for pMTL-ICE1, -G1 and -H1. The third fragment, which contains P_{bgaR}* instead of P_{bgaR}, was amplified with primers IC338-f and IC308-r, using pMTL-HZ1 as a template. The three fragments were then ligated via compatible BsaI restriction sites and introduced into the linearized pMTL-HZ1 vector using the restriction sites NotI and NdeI.

pMTL-JH29T

The vector pMTL-JH29T differs from the original pMTL-JH29 [125] in that the former includes a double terminator downstream of the 300 bp of the *pyrE* gene, which are missing in a *pyrE* deficient strain, and serve to repair the gene after chromosomal integration. The entire *pyrE* was amplified using primers IC339-f and IC340-r, which include the restriction sites SbfI and BsaI, respectively, using pMTL-JH29 as a template. The double terminator was generated by annealing the oligonucleotides IC341-f and IC342-r, which include the restriction sites BsaI and NotI, respectively. The two amplified fragments were then introduced into the SbfI/NotI linearized pMTL-JH29 vector.

pMTL-JH29T-L, pMTL-JH29T-R

Vectors intended to integrate the LAC system-*catP* and RiboLac-*catP* reporter systems into the genome of *C. sporogenes* were generated via digesting the sequences encoding such reporter systems from the reporter plasmids

pMTL-HZ1 and pMTL-ICG1* with the restriction enzymes NotI and XhoI and introducing them into the vector pMTL-JH29T, linearized by using the same restriction sites.

pMTL-ICLG1* and pMTL-ICLH1*

Expression vectors harbouring the Lactoswitch were generated by fusing three DNA sequences via the restriction enzyme BsaI. The first DNA fragment comprises the *bgaR* coding sequence with its native terminator and the intergenic region of the LAC system up to the Shine-Dalgarno (SD) sequence of P_{bgaL} (corresponding with P_{bgaL}*, see Appendix, Fig. S7). This DNA sequence was amplified from pMTL-HZ1 using the primer pair IC307-f and IC343-r, which introduced the restriction sites NotI and BsaI, respectively. The second DNA fragment contains the riboswitch, which was amplified with the forward primer IC117-f and the reverse primers IC112-r and IC113-r for riboswitch-G (pMTL-ICLG1*) and -H (pMTL-ICLH1*), respectively. This amplification was done by aligning the two oligonucleotide primers in the absence of DNA template. The forward primer introduced a BsaI restriction site compatible with fragment one and the reverse primer introduced a BsaI restriction site compatible with fragment three. The third DNA fragment contains the reporter *catP*, amplified with primers IC103-f and IC104-r, which introduced the restriction sites BsaI and XhoI, respectively, using pMTL-83151 as a template. The three fragments were ligated via compatible BsaI restriction sites and cloned into the linearized pMTL-HZ1 vector using the restriction sites NotI and XhoI.

2.6.4 Construction of plasmids used in Chapter 6

pMTL-ICD

The deletion vector, pMTL-ICD, was generated by replacing the pCD6 replicon by the pIM13 replicon through restriction digestion with FseI and AscI of pMTL-CH14 and pMTL-85151, for backbone and replicon, respectively.

pMTL-ICD-spo0A

The *spo0A* knockout cluster was generated by SOE PCR, using genomic DNA from *C. sporogenes* and the primer pairs IC401-f/IC402-r and IC403-f/IC404-r, for left and right homologous regions (1 kbp each), respectively. External primers IC401-f and IC404-r were used to amplify the joined cassette. The knockout cassette was cloned into pMTL-ICD using restriction sites AscI and NheI.

pMTL-JH29-spo0A

The vector used to integrate *spo0A* in the *pyrE* locus was generated by amplifying *spo0A* with its native promoter and terminator from the genomic DNA of *C. sporogenes* with primers IC405-f and IC406-r, which inserted the restriction sites NotI and NheI, respectively. The digested fragment was then ligated into the NotI/NheI linearized pMTL-JH29 backbone.

pMTL-IC321-E, -G and –H

Vectors harbouring *catP* under the control of P_{araE}* located upstream of riboswitches -E, -G and -H were constructed in the same fashion as the theophylline responsive expression plasmids created for Chapter 4. Briefly, oligonucleotides IC103-f and IC104-r were used to amplify *catP* from pMTL83151. Riboswitches were amplified with the forward primer IC107-f and

the reverse IC108-r, IC112-r and IC113-r for riboswitches -E, -G and -H, respectively. P_{araE}* was amplified using genomic DNA from *C. acetobutylicum* ATCC 824 as a template and the pair of primers IC407-f and IC408-r. The amplified DNA fragments, which included compatible BsaI restriction sites, were ligated and inserted into the linearized NotI/XhoI pMTL82251 vector.

pMTL-ICSPO-series vectors

Expressing vectors containing *spo0A* under the control of a riboswitch (pMTL-ICSPO-P_{fdx4}-E, -P_{fdx4}-G, -P_{fdx4}-H, -P_{fdx4}*-E, -P_{fdx4}*-G, -P_{fdx4}*-H, -P_{fdx}-E, -P_{fdx}-G, -P_{fdx}-H, -P_{fdx}*-E, -P_{fdx}*-G, -P_{fdx}*-H) were amplified in a similar fashion to those expressing *catP* (Construction of plasmids from Chapter 4) but containing *spo0A* as the target gene instead of *catP*. To do so, *spo0A* was amplified from *C. sporogenes* using primers IC409-f and IC409-r, ligated to the promoter and the riboswitch via compatible BsaI restriction sites and inserted into the linearized NotI/XhoI pMTL82251 backbone. On the other hand, expressing vectors containing *spo0A* under the control of the LAC system or the RiboLac were generated by digesting pMTL-HZ1 and pMTL-ICG1* with the restriction enzymes NdeI and XhoI, which excised *catP* from the construct. *spo0A* was amplified using primers IC411-f and IC410-r and cloned into the linearized backbones via the same restriction sites.

pMTL-JH29T-*spo0A*-series vectors

All vectors intended to integrate *spo0A* under the control of an inducible system in the *pyrE* locus were generated by restriction cloning of the corresponding pMTL-ICSPO-series vector using the restriction sites NotI and

XhoI and subsequent ligation into the NotI/XhoI linearized pMTL-JH29T backbone.

pMTL-JH29-*spoIIR*

The vector used to integrate *spoIIR* in the *pyrE* locus was generated by amplifying *spoIIR* with its native promoter from the genomic DNA of *C. sporogenes* with primers IC412-f and IC413-r, which inserted the restriction sites NotI and NheI, respectively. The digested fragment was then ligated into the NotI/NheI linearized pMTL-JH29 backbone.

pMTL-ICD-P$_{spoIIR}$*-G-*spoIIR*

The vector used to integrate the sequence P$_{spoIIR}$-G-*spoIIR* in the *spoIIR* locus was constructed by cloning the knock-in cluster into pMTL-ICD. The knock-in cluster was formed by two 1 kbp regions, homologous to the sequences upstream and downstream of *spoIIR* in the genome of *C. sporogenes,* and by the integration "cargo", P$_{spoIIR}$-G-*spoIIR*. The left homologous region was amplified with primers IC420-f and IC421-r, which introduced the restriction sites NheI and AatII, respectively. The right homologous region was amplified with primers IC422-f and IC423-r, which introduced the restriction sites XhoI and AscI, respectively. The integration cargo was generated by linear-ligating three DNA fragments with compatible BsaI restriction sites. As such, the promoter P$_{spoIIR}$ was amplified with primers IC414-f and IC415-r; riboswitch-G was amplified with primers IC416-f and IC417-r, and; *spoIIR* was amplified with primers IC418-f and IC419-r. The three DNA fragments (P$_{spoIIR}$-G-*spoIIR*) were ligated via compatible BsaI restriction sites and amplified with the external primers IC414-f and IC419-r, introducing the restriction sites AatII and XhoI.

The homologous regions and the integration cargo were then cloned into the NheI/AscI-digested pMTL-ICD plasmid.

2.6.5 Construction of plasmids used in Chapter 7

RiboCas-series plasmids

The RiboCas plasmids pRECas1, pRECas2, pRGCas1 and pRGCas2 were first generated to facilitate the construction of plasmids for targeting and editing of specific genes. These plasmids include the nuclease module, this is, the Cas9 coding sequence downstream of P_{fdx} fused to either riboswitch-E or –G, and the promoters P_{araE} (indicated by "1" in the plasmid name) or J23119 (indicated by "2" in the plasmid name), which are intended to drive the expression of the sgRNA. Briefly, Cas9 coding sequence (amplified with primers IC201-f and IC202-r from pMTLCas[127]), riboswitches -E or -G (amplified using the forward primer IC107-f and the reverse IC109-r and IC1012-r, respectively) and P_{fdx} (amplified with primers IC203-f and IC106-r) were assembled using BsaI restriction sites, with the outer flanking sites being NotI and XbaI. P_{araE} was amplified from *C. acetobutylicum* ATCC 824 using the primers IC205-f and IC206-r, which include the restriction site XbaI and a multiple cloning site SalI-AatII-AscI, respectively. J23119 was generated with primers IC207-f and IC208-r, which include the restriction site XbaI and a multiple cloning site SpeI-AatII-AscI, respectively. For vector assembly, the Cas9 module and either P_{araE} or J23119 were ligated by the XbaI restriction site and inserted into the pMTL83151 chassis via NotI and AscI restriction sites.

Targeting vectors

In targeting vectors, the sgRNA module was amplified by PCR using the primer IC209-r, which inserts the AatII restriction site, and the forward primer, specific for each target, flanked by either SalI or SpeI. The editing template was assembled by SOE PCR and inserted into the RiboCas backbone using the restriction sites AatII and AscI. Ligation of both fragments, sgRNA and editing template, was carried out in one step. Primers used for each target on every organism are specified in the Appendix, table S6.

Control plasmids

pRECasC and pRGCasC, which only contain the nuclease module, were generated by amplifying such module with primers IC204-f and IC201-r, using pRECas1 as a template, and inserting it into the backbone using the NotI and AscI restriction sites.

Integration vector pRECas1-IIEin

In pRECas1-IIEin, the homology arms flank an antibiotic resistance cassette (ARC), composed of the gene encoding the erythromycin resistance marker (*ermB*) followed by a sequence of 2.2 kbp of DNA that includes the *lacZα* multiple cloning site (MCS) of pMTL20[128] and the gene encoding the reporter enzyme ß-glucuronidase (GusA) from *E. coli*. In this study, our only intention was to demonstrate the successful integration of large fragments of DNA; all DNA fragments (*ermB*, *lacZα* MCS and *gusA*) have been previously used in *C. sporogenes*[109,129]. The *ermB* gene and the *lacZα* MCS were amplified from pMTL82251 using the primer pairs IC224-f/IC225-r and IC226-f/IC227-r, respectively, and fused together by SOE PCR. The resulting product was fused

by SOE PCR to the gene encoding GusA, amplified from *E.coli* TOP10 strain with primers IC228-f and IC229-r. Left and right homology arms were amplified from *C. sporogenes* with primer pairs IC222-f/IC223-r and IC230-f/IC231-r, respectively, fused to the ARC via SapI restriction sites and inserted into the backbone through the restriction sites AatII and AscI.

2.7 Phenotypic analysis

2.7.1 Growth curve experiments

Growth experiments were performed in 20 mL universal tubes. Three biological replicas of bacterial stocks were inoculated into appropriate media supplemented with antibiotics and allowed to grow overnight. Cultures were then used to set up fresh cultures; calculations were made to ensure that the starting OD_{600} was approximately 0.01 and time point measurements were taken every 1-2 hours. Cultures were grown for 24 hours or early death phase.

2.7.2 Sampling and optical density measurement

For measurement of optical density at 600 nm at specific time points, the CLARIOstar 96 well-plate reader (BMG Labtech, Germany) was used. During growth experiments, samples entering exponential phase were diluted 1:4 with appropriate medium to allow OD measurements to be read within the spectrophotometric range. The appropriate medium was also used as a reference blank.

For solvent analysis, 1mL of sample was centrifuged at 10,000 x g at 4°C for 10 minutes. The supernatant was stored at -80°C for later chromatographic analysis.

2.7.3 Sporulation assays

To quantify heat resistant spores, 5 mL volumes of TYG were inoculated with a loopful of *C. sporogenes* frozen stocks and grown overnight in the anaerobic workstation. Overnight cultures (50 μL) were inoculated into 5 mL of liquid TYG broth and grown to logarithmic growth phase (OD_{600} ≈ 0.5-0.6). Seven 10-fold serial dilutions were then made in TYG and allowed to grow overnight. The most diluted culture that showed growth was then used to inoculate 20 mL of TYG (supplemented with antibiotics if needed). This served to synchronize the cultures and minimize the presence of spores in the starting culture. Directly after inoculation, and every 24 hours for 5 days, total and heat-resistant colony forming units per mL of culture (colony forming units (CFU)/mL) were enumerated. To this end, at each time point serial 10-fold dilutions were prepared in sterile PBS and 3 discrete 20 μL replicas were spotted onto TYG plates and incubated for 24 hours. To determine spore counts, 300 μL of each culture was transferred to a microcentrifuge tube, centrifuged at 6000 x g for 20 seconds and heat treated at 80°C for 20 minutes before spotting onto plates. In the case of conditional sporulation assays, cultures were induced with appropriate concentrations of inducer (lactose and/or theophylline) at early exponential growth phase (OD_{600} ≈ 0.5; 4 hours after inoculation) before further incubation.

2.7.4 Transformation/conjugation efficiency assays

For conjugation efficiency assays, the donor strain, *E. coli* CA434 (in triplicates), was grown overnight at 30°C in LB supplemented with kanamycin

and chloramphenicol. For each replica, two 1 mL cultures were centrifuged at 6,000 x g for 1 minute and then washed once with PBS. After a second centrifugation step, one of the two cultures was used to quantify the donor by plating the appropriate serial dilutions onto LB plates. The second culture was used for plasmid transfer into the clostridial host, as previously described. After 24-48 hr of incubation at 37°C, colonies were counted and conjugation efficiency was calculated as total transconjugants per 1 mL of donor strain.

Transformation efficiency in *C. pasteurianum* was determined as the average of three independent transformations with 4 µg of plasmid DNA.

2.7.5 Phase contrast microscopy

Bacterial samples were observed using a Nikon Eclipse Ci-E microscope (Nikon Instruments Europe B.V., UK) using standard microscopy procedures.

2.8 Enzyme assays

2.8.1 Preparation of cell lysates of C. sporogenes

For each tested system, three independent 5 mL cultures of *C. sporogenes* were grown for 12 hours (with selection if needed) before being diluted to a starting OD_{600} of 0.01 in fresh medium. After selected times (and induction if needed), 2 mL of cultures were taken to prepare cell pellets. Samples were centrifuged for 10 minutes at 12,000 x g and 4°C, supernatants removed and cell pellets stored at -20°C. Cells were lysed on the day of assay using the Novagen® BugBuster Master Mix according to the manufacturer's instructions and soluble fractions were used in further experiments.

2.8.2 Determination of protein concentration using the BCA assay

Total protein concentration of cell lysates was evaluated by the BCA method, using the Pierce BCA Protein Assay Kit, according to the manufacturer's instructions. The test was conducted in a 96-well plate and absorbance measured on the CLARIOstar® microplate reader.

Bovine serum albumin (BSA) standards were prepared from 100 to 1400 μg protein per mL and used to generate a standard curve. For all test samples, 25 μL of diluted cell lysates (or BSA standards) were added to 200 μL of the working reagent and mixed on a plate shaker for 30 sec. Then, the plate was covered and incubated at 37°C for 30 minutes. After this time, the absorbance at 562 nm was measured relative to BugBuster Master Mix as a blank.

2.8.3 Enzymatic assay of Chloramphenicol Acetyltransferase (CAT)

CAT activity was determined according to the Sigma-Aldrich protocol, based on Silverstein and Shaw publications[130,131]. The assay relies on the enzyme CAT, which catalyzes the acetylation of chloramphenicol in a reaction that requires acetyl-CoA as a cofactor and, as a result, produces HS-CoA. The rate of this reaction can be measured by addition of 5,5′dithio-bis (2-nitrobenzoic acid) (DTNB), which reacts with the thiol group of HS-CoA to produce the yellow compound TNB. The rate of the reaction is then calculated by measuring the change in absorbance at 412 nm over time.

To perform the assay, reagents were mixed to produce a reagent master mix to give final concentrations in a 200 μL reaction volume of 94 mM Tris (pH

7.8), 0.083 mM DTNB, 0.17 mM acetyl coenzyme A and 0.005% (w/v) chloramphenicol. A volume of 40 µL of 100 mM Tris, pH 7.9 and 10 µL of cell lysates were added to individual wells of 96 well-plates, placed in the CLARIOstar® plate reader (BMG Labtech) and allowed to equilibrate at 25°C. The CLARIOstar® device master mix dispenser was used to inject 150 µL of master mix per well and the reaction was allowed to run for 1 minute; during this time, the absorbance change at 412 nm was recorded every second.

Analysis of the data and calculation of CAT Units per ml were performed by defining the maximum linear rate using MARS data analysis software and using the equation:

$$Units/ml\ CAT = \frac{(\Delta A_{412}/\text{ minute}\ test - \Delta A_{412}/\text{ minute}\ blank)(0.2)(df))}{(0.0136)(0.01)}$$

0.2 = Total volume (in ml) of assay
df = Dilution factor
0.0136 = Micromolar extinction coefficient for DTNB at 412 nm
0.01 = Volume (in ml) of cell lysate

In order to define the CAT activity in U/mg, quantification of the protein concentration was carried out and the U/mg protein calculated according to the following equation:

$$\frac{Units}{mg}protein = \frac{\frac{Units}{ml}enzyme}{mg\ protein/ml\ enzyme}$$

2.9 Analytics

2.9.1 High performance liquid chromatography (HPLC) and HPLC/Mass spectrometry (MS) analysis

To analyse if *C. sporogenes* consumes lactose or theophylline, 12 hour TYG cultures were diluted to a staring OD_{600} of 0.01. At an OD_{600} of approximately 0.5, cultures were induced with 2 mM theophylline or 10 mM lactose. Sample volumes of 1 mL were taken immediately and then at 2, 4, 6, 8, 10 and 20 hours after supplementation of theophylline. Samples were centrifuged for 10 minutes and 10000 x g.

For lactose determination, supernatants were diluted with an equal volume of mobile phase containing 50 mM valeric acid as an internal standard. The mobile phase was composed of 5 mM H_2SO_4. Samples and standards were filtered using a 0.2 µM pore size filter before injection. The analysis of the samples was performed with a Dionex UltiMate 3000 HPLC system (Thermo Scientific), using a Rezex^{TM} ROA-Organic Acid H^+ 8%, 150 mm x 7.8 mm x 8 µm column (Phenomenex), operated at 35°C and utilizing refractive index detection and a diode array detector at 210 nm. Flow rate was 0.5 mL/minute and the injection volume was 20 µL. All concentrations were determined by calibration against known concentration standards.

For theophylline determination, supernatants were analysed with an Alliance HT 2795 HPLC (Waters Corporation) coupled to a Micromass Quattro LC Mass Spectrometer (Waters Wilmslow) equipped with an electrospray source in positive mode. The MS parameters were as follows: capillary voltage 2000 V, sampling cone 30 V, source temperature 130°C, desolvation temperature 350°C, cone gas flow 64 L/hour and desolvation gas flow 621 L/hour. Prior to ionization, chromatographic separation was achieved using a Supelco Ascentis Express HPLC column, 100 mm x 3 mm, 2.7 µM (Sigma Aldrich). The mobile phase consisted of (A) water with 0.1% (v/v) formic acid

and (B) methanol with 0.1% (v/v) formic acid. In all HPLC runs the elution gradient started at 95% A, 5% B increasing to 10% A, 90% B, followed by a 5-minute re-equilibration period. A sample volume of 10 µL was injected for each HPLC run. The column was operated at 40°C with a flow rate of 0.4 mL/minute. The HPLC run contained blanks and the sample-relevant standard solution. Samples and standards were filtered using a 0.2 µM filter. Concentrations were again determined by calibration against a set standard curve.

2.10 List of bioinformatics tools

Table 2.9 List of bioinformatics tools used in this study

Bioinformatic tool	Use/Description	Source
ARNold algorithm	Terminator sequences and prediction of mRNA secondary structures	Institut de Génétique et Microbiologie
Blastp algorithm	Protein database searches	NCBI
CLC Genomics Workbench	Whole genome assembling and detection of SNVs	Qiagen
GraphPad Prism Software	Design of all graphically presented results and statistical analysis	GraphPad Prism
LightCycler® 480 software, version 1.5	Analysis of quantitative PCR	Roche
Megablast, discontiguous megablast and blastn algorithms	Search similarities between sequences	NCBI
MEME Version 4.11.2	Discover consensus motifs	MEME
Neighbor joining algorithm	Genetic distance between orthologous proteins	MEGA
Snap Gene Software	DNA visualization, cloning simulations, primer design, sequence annotations	SnapGene®

2.11 Data and statistical analyses

Numerical data was stored, calculated, analysed and presented graphically using Graphpad Prism 7 (GraphPad Software inc, USA) and Microsoft Excel 2016 (Microsoft). Calculations included determination of standard deviations of the mean and statistical significance values (P value) between data sets were determined using paired or unpaired two-tailed Student's *t-tests*.

Chapter 3
/ The LAC system from *C. perfringens* strain 13

"The proof of the explicability of any single life phenomenon is furnished as soon as it is successfully controlled unequivocally through physical or chemical means or is repeated in all details with nonliving materials"

– Jacques Loeb, 1903

3.1 Background

In the LAC system from *C. perfringens* strain 13, transcription from P_{bgaL} is stimulated by lactose[102]. Divergently oriented, and located 363 bp from *bgaL*, is *bgaR*, the gene encoding the TF BgaR. BgaR is a protein composed of 279 amino acids that is classified as a member of the AraC/XylS family of transcriptional regulators. Members of this family are widely distributed among Gram-positive and Gram-negative bacteria, where they modulate the transcription of genes involved in stress response, carbon metabolism or virulence[132,133]. They are typically 250-300 amino acids long and share with AraC, the first transcriptional activator of this family to be described[134], a conserved DNA-binding domain of approximately 100 residues located at the C-terminus. The majority of the family members also have a non-conserved N-terminal domain responsible for effector recognition and, if applicable, dimerization[132]. AraC/XylS family TFs associated with carbon metabolism, such as AraC and XylS, often are active as dimers[133], whereas those involved in stress response, such as MarA, Rob and SoxS, are active as monomers[135].

The DNA binding domain of AraC/XylS-like TFs is composed of two helix turn helix (HTH) motifs connected by a central helix which insert into the major grooves of two adjacent DNA turns[133]. Binding sites within (or in close proximity) to the regulated promoters have been identified for several members of the family. MarA and related TFs, such as Rob and SoxS, bind asymmetric DNA sequences of approximately 15 bp within the intergenic region located between the transcriptional regulator and the regulated gene. These three transcriptional activators can regulate the same set of promoters (>40), but to different extents[136]. For this reason, their binding sites share some homology,

and it is the affinity between the transcriptional activator and the specific binding site that determines the extent to which each activator regulates a specific promoter. The consensus binding site for each of these regulators has been identified as GCAC-N_8-AAA[137], GCAC-N_7-CAA[138] and GCAC-N_7-TAAA[139] for MarA, Rob and SoxS, respectively. Dimeric AraC/XylS regulators bind to consecutive sites, each bound to one monomer through the HTH motif. Characterized binding sites of this type are typically composed of combinations of similar or otherwise perfect repeats. As examples, the AraC binding site was identified as TAGC-N_7-TCCATA and TAGC-N_7-ACCTGA[140], whereas XylS binds to the two perfect half sites TGCA-N_6-GGNTA and TGCA-N_6-GGNTA[141].

The mechanisms involved in the regulatory process differ between different members of the AraC/XylS type regulators. In all cases, transcriptional modulation results from a structural change upon effector binding; however, this modulation might occur by means of DNA bending, through DNA binding affinity or via interactions with the RNA polymerase (RNAP)[133]. For example, the *E. coli* XylS activator does not bind to its target site in the absence of xylose. Upon ligand binding, a structural change occurs in the D-xylose-binding domain that results in dimer rearrangement, which is transmitted to the DNA-binding domain. XylR then binds to two DNA sites per dimer and, according to recent results, activates gene expression via a looping mechanism[142]. This activation looping is in contrast to the repressive mechanism of DNA looping by AraC, which in the absence of arabinose, loops DNA and inhibits P_{BAD} transcription[143]. Conversely, MarA binding activates gene expression through direct interactions with RNAP[144].

The aforementioned work done by Hartman and coworkers[102] is, as yet, the only published information regarding the mechanistic characterization of the LAC system of *C. perfringens* strain 13. In that study, the system was also shown to function in two strains of *C. perfringens* other than its native host – strains SM101 and JGS4143. The demonstration of BgaR being the transcriptional factor that stimulates P_{bgaL}-driven expression was established through the construction of a *bgaR*-deficient variant, which resulted in decreased P_{bgaL} activity. However, 60% of the activity remained. Since all the experiments were performed in rich medium containing high levels of glucose, it was also demonstrated that the system is not subject to catabolite repression. Analysis of its kinetics also revealed that induction takes place between five and ten minutes after lactose is added, indicating that lactose, and not a derivative, is the inducer that directly interacts with the BgaR transcriptional regulator. An 80-fold increase in expression at maximum induction was reported through coupling the promoter to the *E.coli gusA* reporter gene. In the absence of the inducer, variable levels of *gusA* expression were obtained in the various strains.

The utility of the system has now been investigated in a number of different clostridial species. One example is the generation of a system for the isolation of allelic exchange mutants in *C. acetobutylicum* through the conditional expression of the toxin encoding *mazF* gene[145]. Other examples include its exploitation in metabolic engineering studies in *C. ljungdahlii*[107] and its use to control the expression of a gene coding for an endogenous CRISPR array in *Clostridium tyrobutyricum*[146].

3.2 Chapter overview

Although functional in several *Clostridium* species, the LAC system from *C. perfringens* strain 13 exhibited considerable expression in the absence of inducer in all published studies. However, due to its large dynamic range, it seemed to be one of the most suitable inducible systems hitherto tested in clostridial organisms. In this chapter, the successful implementation of the LAC system is demonstrated in *C. sporogenes* NCIMB 10696. Due to the importance of having inducible systems where gene expression can be tightly controlled, the system was exhaustively characterized with the aim of further optimizing it to reduce basal expression levels while not compromising its dynamic range. Characterization of the system provided information on its functionality that had not previously been reported. This included the identification of the transcriptional start site and the putative -10 and -35 promoter elements of both, the transcriptional regulator and the regulated gene, as well as the region to which the transcriptional regulator binds. The work carried out in this chapter established the basis for the development of a novel and synthetically regulated inducible system for clostridia.

3.3 Results and Discussion

3.3.1 Functionality of the LAC system in C. sporogenes

Initial studies focussed on determining the functionality of the aforementioned LAC system in *C. sporogenes* NCIMB 10696. A DNA fragment encompassing *bgaR*-P_{bgaL}, provided by Hengzheng Wang (SBRC), had previously been synthesized by ATUM. This sequence includes the *bgaR* native terminator and the flanking restriction sites NotI/NdeI, incorporated to facilitate cloning. To preserve the original distance between the RBS and the translational start site when inserting the restriction site NdeI, the three nucleotides prior to the translational start site of P_{bgaL} had been excluded (Figure 3.1). The *bgaR*-P_{bgaL} fragment was then introduced into the vector pMTL82254[109] between the restriction sites NotI/NdeI, which positioned the promoter cassette upstream of the *catP* reporter gene, which encodes the reporter CAT (EC: 2.3.1.28), yielding the construct pMTL-HZ1 (Figure 3.2). This reporter plasmid includes the following elements: the pBP1 Gram-positive origin of replication, the Gram-negative ColE1 origin of replication, the erythromycin resistance gene (*ermB*) and the terminators of the *C. difficile* CD0164 gene, T_{CD0164}, and that of the *C. sporogenes* ferredoxin gene, T_{fdx}. The two terminators are placed downstream of *bgaR* and *catP,* respectively, and serve to prevent background reporter expression. The pBP1 Gram-positive origin of replication was chosen as it had been previously demonstrated to be one of the most stable replicons in this strain of *C. sporogenes*[34]. As a control, a promoter-less backbone, pMTL-IC001, was assembled to detect background reporter gene expression. The designated plasmid constructs were transferred into *C. sporogenes* from their respective *E. coli* donors by conjugative plasmid transfer and cultures of the resultant

transconjugant cells exposed to increasing concentrations of lactose, ranging from 0.1 to 100 mM, when they had reached early exponential growth phase (OD_{600} ≈ 0.5). CAT activity was determined in cell lysates derived from stationary phase cultures grown in the presence and absence of the inducer (Figure 3.3).

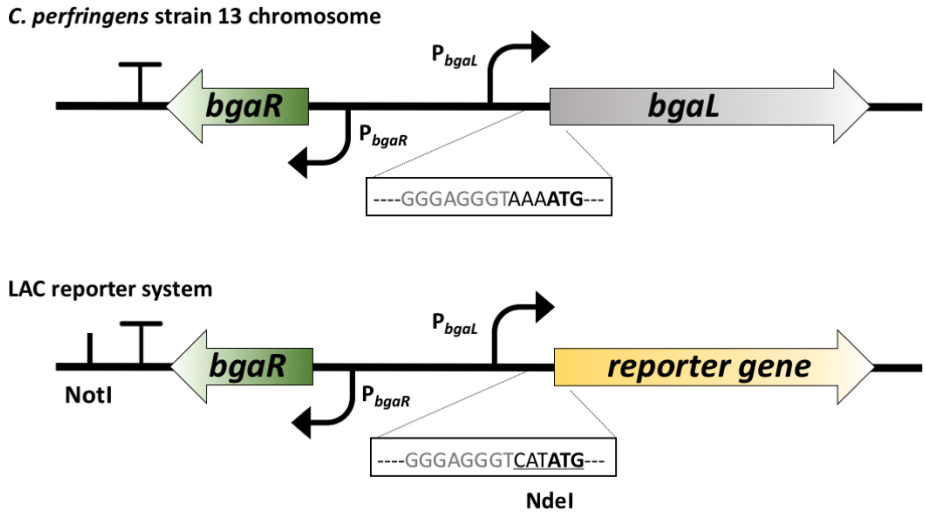


Figure 3.1 Schematic representation of the LAC system from *C. perfringens* strain 13. The system includes the transcriptional activator gene – *bgaR* – and the divergently oriented *bgaL* gene, which encodes a ß-galactosidase involved in lactose metabolism. As a reporter system, the transcriptional activator, *bgaR*, and the corresponding inducible promoter, P_{bgaL}, are flanked by the restriction sites NotI/NdeI for easy cloning into the modular pMTL8000 series vector[109].

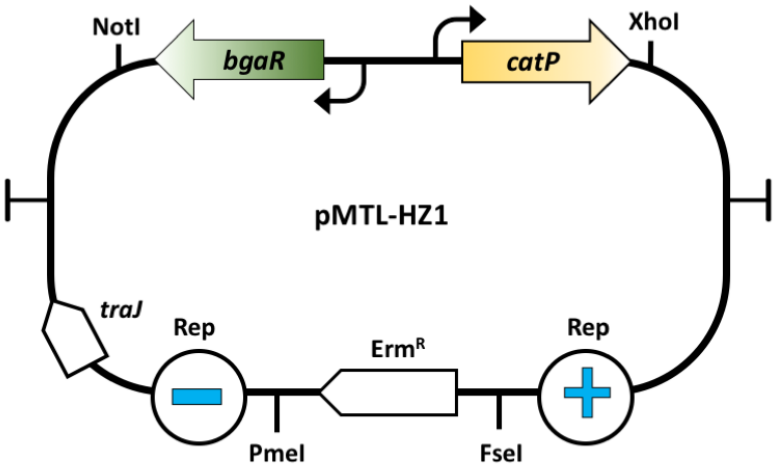


Figure 3.2 Schematic representation of pMTL-HZ1. The reporter gene *catP* is placed under the control of the LAC system. (-) is the Gram-negative origin of replication. *traJ* encodes the TraJ protein, needed for conjugal transfer. (+) is the Gram-positive origin of replication.

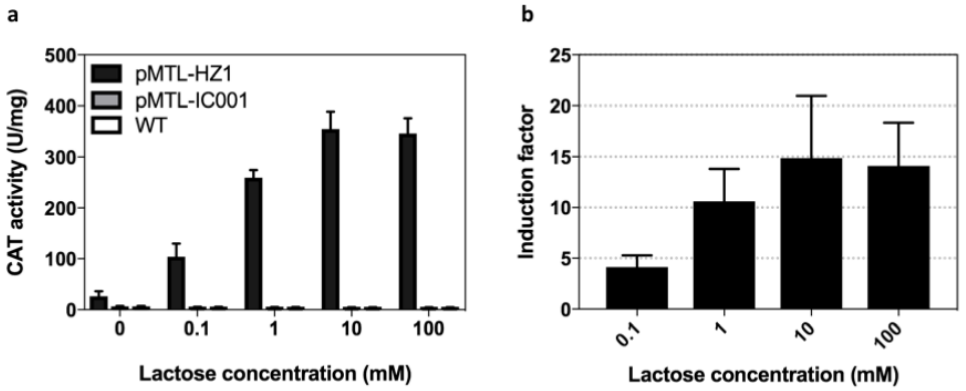


Figure 3.3 LAC system dose-dependent response. **(a)** Response of the LAC system in *C. sporogenes* to different concentrations of lactose. Pre-cultures of cells containing the plasmid pMTL-HZ1 were inoculated into fresh TYG medium to a starting OD_{600} = 0.05; each culture was split into five equal volumes, and each volume was then supplemented with different concentrations of lactose (0, 0.1, 1, 10 and 100 mM). The promoter-less vector, pMTL-IC001, and WT were used as negative controls. **(b)** The activation ratio at each concentration was calculated by dividing the value of CAT activity measured in the presence of the inducer by the value of that in the absence of the inducer. Error bars represent the standard deviation of three independent biological replicates.

As shown in Figure 3.3, increasing doses of lactose resulted in higher levels of the reporter CAT, with the maximum level of expression at a concentration of 10 mM. CAT activity of cells harbouring the control pMTL-IC001 was similar to that from cultures lacking the reporter backbone (WT), indicating no detectable transcriptional read-through. Additionally, basal expression from pMTL-IC001 was lower than that of the cells harbouring the LAC system in the absence of lactose. These data illustrate the detectable background expression of the system, which affords a modest 15-fold dynamic range – the absolute difference between the ON and OFF states – in *C. sporogenes*.

3.3.2 Kinetics of the LAC system in C. sporogenes

Having established the functionality of the LAC system, further experiments were undertaken to determine the kinetics of induction, specifically, the time that is required for the system to respond to a change in lactose levels as well as the time-course of induction. Accordingly, *C. sporogenes* cells harbouring the plasmid pMTL-HZ1, were cultivated in TYG medium in the presence and absence of 10 mM lactose and their optical density (OD_{600}) monitored over time (Figure 3.4b). Lactose addition was made at 4 hours, when cells had reached the early exponential growth phase ($OD_{600} \approx 0.5$); samples were collected at 0, 2, 4, 6, 8, 10, 12 and 24 hours and the level of activity in cell lysates determined (Figure 3.4a).

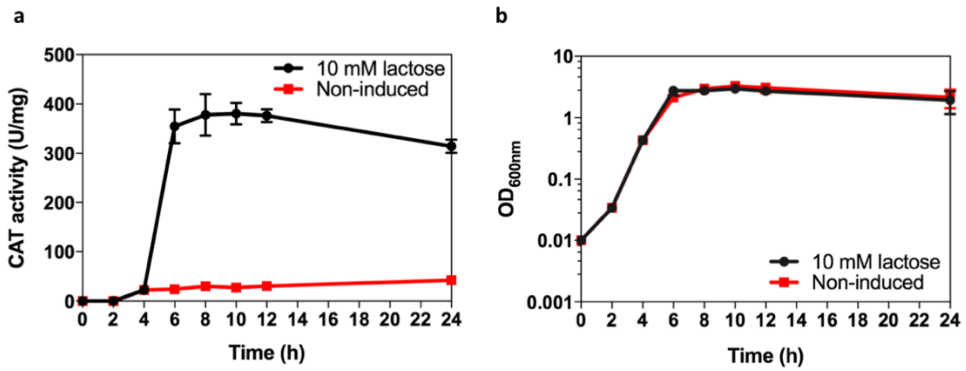


Figure 3.4 Time course of induction of LAC system **(a)** CAT expression profiles over time of cells harbouring pMTL-HZ1 in the presence or absence of 10 mM lactose. At 4 hours, samples were split into two equal volumes and one of them was induced with 10 mM lactose. Samples at 4 hours were collected immediately after induction. **(b)** Growth of *C. sporogenes* monitored over time in the presence and absence of 10 mM lactose. Cultures are the same as used in (a). Error bars represent the standard deviation of three independent biological replicates.

Expression of *catP* above the level of the uninduced culture was detectable immediately upon addition of lactose (at 4 hours), supporting the notion that induction occurs within a few minutes of lactose being added. Maximum CAT activity was observed at 8 hours. The levels of reporter began to decrease after 10 hours and had been reduced by approximately 15% after 24 hours. Since no differences in growth rates were observable between the cells grown in the presence and absence of lactose, the observed reduction in expression was not a consequence of a decrease in growth rate caused by the addition of inducer. It is known that the availability of the cell's transcriptional and translational machineries is growth-rate dependent, a factor that can affect gene expression[147,148]. In particular, this dependency is more pronounced in the case of positive regulation mediated by constitutively expressed activators, since the expression of the activator itself – required to activate the expression of the target gene – is already growth-rate dependant[147]. Basal CAT expression in the

absence of the inducer slightly increased with time, which is indicative of the leaky transcription from P_{bgaR} and/or P_{bgaL}.

3.3.3 Lactose stability in C. sporogenes cultures

The most effective inducers remain constant during the course of the experiment. To establish whether this was the case for lactose in *C. sporogenes*, cells were cultivated in TYG medium – the medium used to perform previous experiments — supplemented with 10 mM lactose. The concentration of lactose was monitored by HPLC (high performance liquid chromatography) analysis of cell-free supernatant samples over time (Figure 3.5).

Results showed that whilst the concentration of lactose initially decreased after induction, from two hours onwards it remained constant over the course of the experiment, indicating that lactose seems to be not metabolized by *C. sporogenes* in rich medium. The initial decrease in lactose concentration could be associated with the transport of lactose into the cells, becoming undetectable in the cells supernatants. Since the concentration of lactose remained constant in the culture over time, it seems plausible that the decrease in gene expression observed in the time course experiment (Figure 3.4a) was merely growth-rate dependent. Although lactose remains stable in rich medium, further experiments (growth in minimum/defined media) could show if lactose is consumed in the absence of other carbon sources (i.e., without glucose).

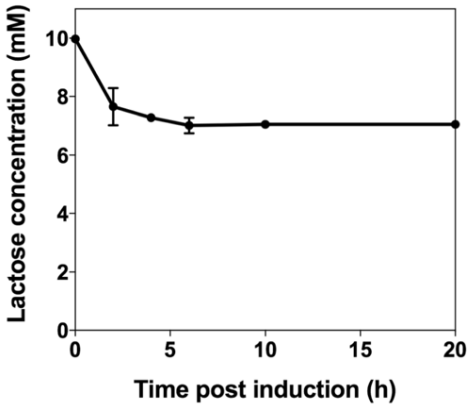


Figure 3.5 Lactose stability in cultures of *C. sporogenes* harbouring pMTL-HZ1. Cells were grown in TYG medium supplemented with 10 mM lactose at early exponential phase ($OD_{600} \approx 0.5$). Samples were collected prior to induction and at 2, 4, 6, 8, 10, 12 and 20 hours after induction; cell-free supernatants were analysed by HPLC analysis (see section 2.9.1). Error bars represent the standard deviation of three independent biological replicates.

● ● ●

The LAC system was shown to be functional in *C. sporogenes*. However, for applications where very tight control of gene expression is required (i.e., to control sporulation), the system remains far from ideal. Of the inducible systems available for *Clostridium*, only the TET system[104] exhibits very low basal expression while maintaining a high induction efficiency. Nevertheless, the optimal working conditions of the TET system require high doses of the inducer anhydrotetracycline, a tetracycline-analogue. The levels required, however, have significant inhibitory effects on cell growth[104]. Thus, it seemed necessary to obtain (or develop) a more tightly regulated inducible system for *Clostridium* that does not rely on any antibiotic and is suitable for any potential application.

In this sense, there are two overall strategies that could have been considered: the native strategy, based on the search of undiscovered/untested

inducible systems, or the synthetic strategy, that relies on the engineering of novel biological systems to achieve the desired functionality.

In nature, the ideal ON/OFF state for the expression of a particular gene might not exist. For years, several studies demonstrated that within a cell population, phenotypic diversity benefits survival in a fluctuating environment[149,150]. Recent studies on dynamics of gene expression suggest that promoter leakage is a key aspect in this respect, as it provides a mechanism to control diverse phenotypes and allows survival in diverse habitats[151]. In agreement with this, several studies support the notion that gene expression noise – variation in protein content among identical cells experiencing the same environment – provides a selective advantage under stress[152] and that many natural promoters, in particular those that are highly regulated by TFs, exhibit high levels of noise[72]. For this reason, inducible systems found in nature might inevitably be linked to expression noise and its associated leakage. Synthetic systems, on the other hand, can be designed to have a reduced background expression, combining both sophisticated regulation and low transcriptional noise.

Hence, the strategy adopted in this study was to synthetically optimize an inducible system to achieve the aforementioned characteristics. One of the most obvious options to start with was the LAC system from *C. perfringens* strain 13. However, modification of an inducible system requires, at the very least, an understanding of its most important regulatory elements. For this reason, and as this LAC system has been, as yet, poorly characterised, this chapter will focus on its characterization.

3.3.4 Further characterization of the LAC system

During transcription initiation, RNAP and one or more transcriptional initiation factors bind to the promoter sequence via sequence-specific interactions to form an RNAP-promoter open complex, initiating transcription at the TSS[153]. Although the distance between the TSS and the promoter core elements can vary, there is a general preference for TSS selection at a position 7 bp downstream of the promoter -10 element[154]. Although there are several tools for predicting promoters available, the vast majority rely on sequence databases from model organisms, such as *E.coli*, and are not yet very efficient for non-model chassis like *Clostridium*. Thus, accurate TSS location makes determination of promoters more accurate.

In order to determine the TSS of the promoters involved in the functioning of the LAC system, the approach 3′ RACE[155], was used. This technique allows the synthesis of cDNA from a specific mRNA target. A polyA-tail, termed adaptor, is then added to the 5′ end of the cDNA, which allows further amplification and determination of the sequence of the 5′ UTR. Analysis of the *bgaR* promoter required a new reporter vector containing P_{bgaR} upstream of the *catP* reporter, termed pMTL-IC1201. As with previous reporter constructs, it is based on the pMTL82251 chassis[109]. The general LAC reporter vector, pMTL-HZ1, was used to determine the TSS of P_{bgaL}. Total RNA was extracted from stationary cultures of *C. sporogenes* harbouring the designated constructs and 5′ RACE was performed using *catP* specific primers for the cDNA reverse transcription and extension steps (see section 2.5.9) (Figure 3.6). The extension step also required the utilisation of a primer that binds to the

adaptor, named anchor primer. As a result, the amplification of the 5′ UTR sequence downstream of P_{bgaL} and P_{bgaR}, was successfully obtained (Figure 3.7).

Sequencing of the PCR products resulted in 5′ UTR sequences with the TSS located 66 and 109 nucleotides upstream of the translational start site (ATG) of P_{bgaR} and P_{bgaL}, respectively. It has to be noted that one limitation of the 5′ RACE procedure is the inaccurate determination of the TSS of mRNAs ending with one or several thiamine residues, since during 5′ RACE a poly-A tail adaptor is added to the cDNA 3′ end (complementary to the 5′ UTR in the mRNA). Therefore, discrimination between a thiamine residue belonging to either the original 5′ UTR sequence, or to the anchor primer that binds to the poly-A adaptor, is impossible. Following the rule of the 7 bp upstream of the TSS, and based on the consensus 17-bp spacer between the -10 and -35 regulatory elements[59], the putative -10 and -35 boxes were manually annotated. It is worth noting that, despite the determination of the putative -10 and -35 elements being possibly accurate, they can only be approximations, since the analysis of the importance of every nucleotide residue nearby would be required for an exact determination.

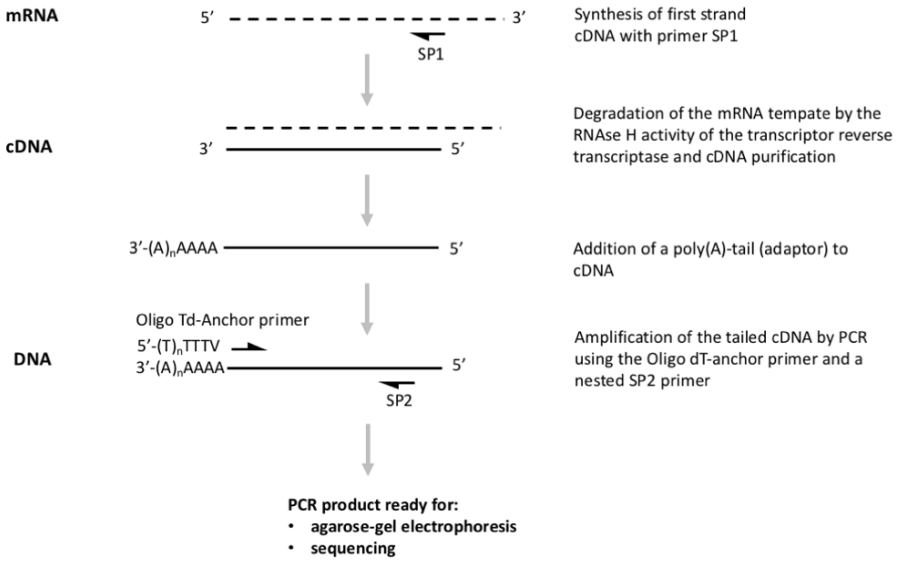


Figure 3.6 Schematic illustration of the 5′ RACE procedure. In this study, SP1 and SP2 refer to the *catP* specific primers 1 (IC263-r) and 2 (IC264-r), respectively. SP1 and SP2 aligned 580 and 245 bp downstream of the translational start site, respectively. PCR products were sequenced with a third nested *catP* specific primer, SP3 (IC265-r), which aligns 190 nucleotides downstream of the translational start site.

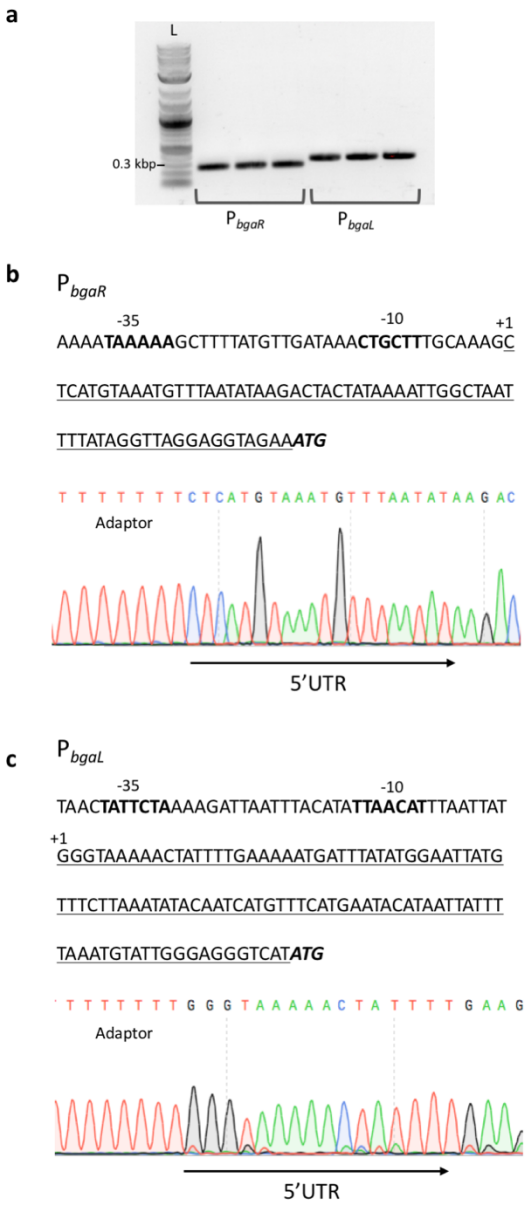


Figure 3.7 5′ RACE experiments on P_{bgaR} and P_{bgaL}. **(a)** The agarose gel of PCR amplifications of three independent DNA samples (from each promoter) that were reverse-transcribed from mRNA template according to the 5′ RACE protocol (section 2.5.9). cDNA transcription was performed with a primer aligning 250 bp downstream of the translational start site. Extension PCR was performed with a primer aligning to the adaptor at the 3′ end of the cDNA (complementary to the 5′ UTR of the mRNA) and a primer aligning 190 bp downstream of the translational start site, producing a product of at least 200 bp. L is 2-log ladder. **(b)** Sequence of P_{bgaR} and 5′ RACE sequencing results. **(c)** Sequence of P_{bgaL} and 5′ RACE sequencing results. In both cases, the sequence is the reverse complementary DNA of the adaptor. The putative -10 and -35 boxes were annotated manually in agreement with the determined TSS (+1). The translational start site is in bold and italicised. The 5′ UTR is underlined.

As previously discussed, regulation of gene expression in response to inducing ligands (i.e., lactose) is mainly mediated by TFs[156]. These interact with conserved cis-acting elements, or operator motifs, that are located in close proximity to or within the promoter sequence of genes involved with the metabolism of the inducer. Since *bgaR* encodes the TF that activates gene expression in the LAC system[102], the operator sequence/s linked to this regulation should be situated in close proximity to the annotated regulatory elements of P_{bgaL}. One simple approach to identify regulatory elements within the promoters of regulated genes is to compare sequences from divergent species[157]. By multiple alignment, conserved elements can be distinguished from sequences that have diverged more rapidly. Accordingly, an approach to predict TF-specific motifs, previously reported by Francke and colleagues[158], but applied to a different family of TFs, was chosen. The main premise behind this approach is that equivalent regulatory sites will be shared only if orthologous sequences share genomic context. In this sense, synteny – conserved gene order – was considered to strongly indicate functional equivalency[159]. Thus, only TFs divergently oriented in relation to the regulated gene were considered in this study. Since BgaR (accession number WP_003480500), is classified as an AraC/XylS-famly of TFs, a total of 66 non-redundant AraC/XylS-family TFs from taxonomically related Firmicute genomes, sharing at least 20% of homology with BgaR, were aligned and grouped through a Neighbor Joining algorithm using the program MEGA version 6[160] (Figure 3.8). The 300 bp upstream of the most phylogenetically related sequences (red branch) were used to identify consensus motifs using the program MEME[161]. As a result, two consensus motifs of 20 bp in length, termed

O1 and O2, were identified as hypothetical regulatory elements of the LAC system (Figure 3.9). Figure 3.10 illustrates the LAC system from *C. perfringens* strain 13 and the hypothetical operator sequences O1 and O2.

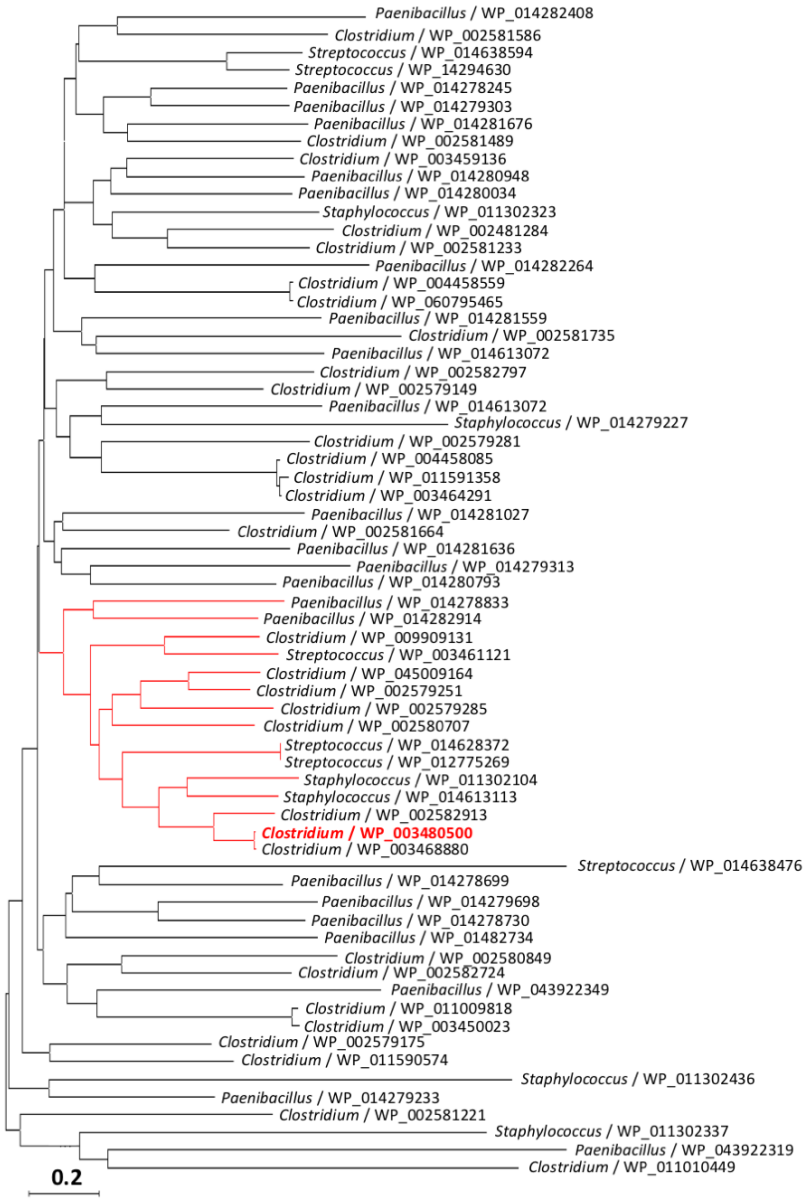
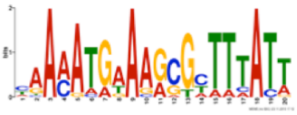


Figure 3.8 Neighbour-Joining tree with 66 homologous AraC/XylS-family TFs from species belonging to the division Firmicutes, specifying the genus and the protein accession number. The tree was aligned based on the respective protein sequences. Branch lengths correspond to the relative divergence in amino acid sequences. BgaR, with accession number WP_003480500 is highlighted in red. The 14 TFs most related to BgaR in phylogeny (red branch) were used for the identification of consensus motifs.

a

AraC family TF	Strand	Identified consensus motif
WP_012775269	+	TCTTATTGTA CAAAATGAAAACGCTTTATT TAATGGTCTT
WP_014638372	+	TCTTATTGTA CAAAATGAAAACGCTTTATT TAATGGTCTT
WP_011302104	+	TAAAAACAAC AAACATGAAAGCGCTTTACT TAAAAAGGAT
WP_009909131	-	ACTATAACTA TAACATGTAAGCGCTTTATA ATCAAATAGG
WP_003468880	+	AGCAGTTTAT CAACATAAAAGCTTTTTATT TTAAAATAAA
WP_014613113	+	TATCAATTAA CGAAAAGAACGCGTTATATT CCGTACATGA
WP_002579251	+	AAAATGATTA CAAAATGGAGGGGAATTATG TTAATAAATT
WP_045009164	-	ACATTATAAA TGAAATGAAAAGGATTAATT TTACTTTAAA
WP_014278833	-	AAGCTTCAAG GCAAAAGGAGGGGCTTCATT GTCTTTATTA
WP_002580707	+	ACTTTTTAAG TGAAACAAAAGAGTTTTATT ATCTATCTCG
WP_014282914	+	TGCTTTCGAT ATAAATGAAAGCGTTAAATA ATGTAATGCA
WP_002582913	+	TATTCCATTT AAAAGTGTAAAGGAATCACT AAAAAATATT
WP_003480500	+	AGCAGTTTAT CAACATAAAAGCTTTTTATT TTAAAATAAA



b

AraC family TF	Strand	Identified consensus motif
WP_012775269	-	AATCCTCCTA GGAAACATTATAACATGTTT TAAATACAAA
WP_014638372	-	AATCCTCCTA GGAAACATTATAACATGTTT TAAATACAAA
WP_002582913	+	TATATATTTT GGAAACATTATAACATATTT GAGCACTATA
WP_009909131	-	CCGCAAGCAA GGTAACATAGCGACCAATTT TCTTCATGAT
WP_003468880	+	TAAGATTATA AGAAACATTATAACATATTT TAGAACTTTT
WP_014613113	+	CTTTAAATAT GGAAACATTTTAACATATAC AGTGACGATT
WP_011302104	+	TAATATGAAA TGAAACATTTTGCCATATAT CATCACCATT
WP_003480500	+	TAAGATTATA AGAAACATTATAACATATTT TAGAACTTTT



Figure 3.9 Motifs identified within 300 bp upstream of the genes encoding the AraC/XylS-family TFs. **(a)** Motif O1, found in 13 out of the 14 sequences selected. **(b)** Motif O2, found in 8 out of the 14 sequences selected. BgaR, with accession number WP_003480500, is highlighted in red. All sequences containing O2, also contained O1. In only 5 cases, highlighted in blue, was O1 present in the absence of O2.

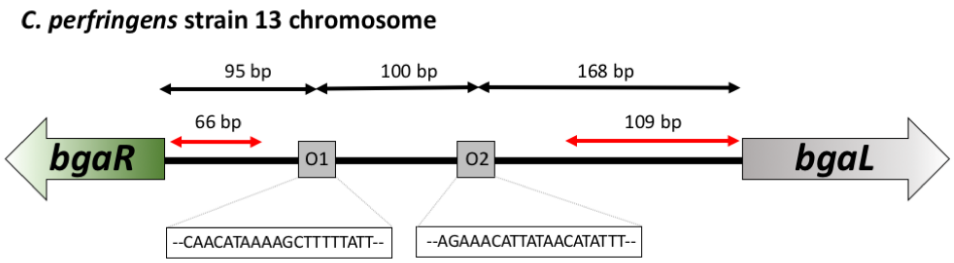


Figure 3.10 Schematic representation of the LAC system and the hypothetical operator sequences O1 and O2. Distances calculated from the position of the 10th base of the 20 bp hypothetical operator sequences. Red arrows indicated the length of the 5´ UTRs, experimentally determined in the previous section.

Following the identification of O1 and O2 as hypothetical operator sequences, 3 new constructs were designed in which either O1 or O2 and both

O1 and O2 had been deleted, generating vectors pMTL-ICMO1, pMTL-ICMO2 and pMTL-ICMO1O2, respectively. Additionally, a construct lacking *bgaR*, termed pMTL-ICMB, was also created, which served to determine to what extent activation of gene expression depended on the presence of BgaR.

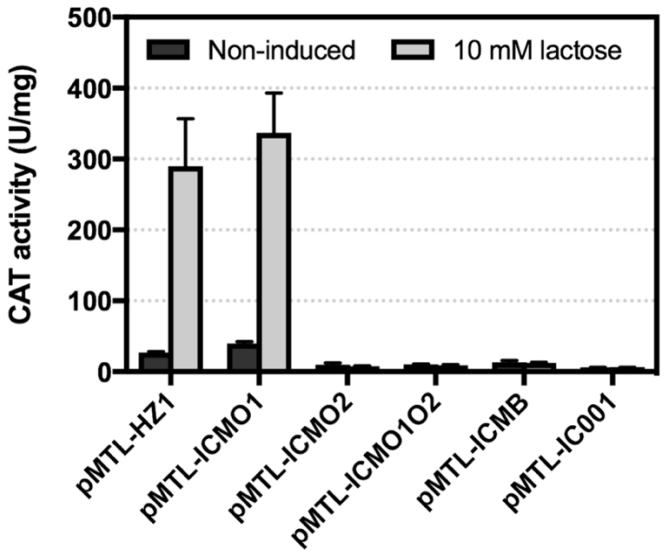


Figure 3.11 LAC mutations O1, O2 and O1/O2. Plasmid-based CAT assay of *C. sporogenes* strains expressing *catP* under the control of the LAC system (pMTL-HZ1) and the former system lacking either the hypothetical operator sequences O1 or O2 (pMTL-ICMO1 and pMTL-ICMO2, respectively) or both operator sequences (pMTL-ICMO1O2). The vector pMTL-ICMB contains the LAC system with a deletion of the entire *bgaR* gene. The promoter-less reporter plasmid, pMTL-IC001, was used as negative control. Cells containing the designated plasmids were cultivated in TYG medium lacking lactose as well as a culture to which lactose (10 mM) was added when they had reached the early exponential growth phase ($OD_{600} \approx 0.5$). CAT activity was determined on cell lysates derived from stationary phase cultures. Error bars represent the standard deviation of three biological replicates.

The CAT data shown in Figure 3.11 confirmed that O2, or part of it, is a key regulatory sequence for the functionality of the LAC system. The deletion of O2, alone or in combination with O1, led to a severe reduction of gene expression and no differences were observable when lactose was present or absent. The same expression profile was found when *bgaR* was deleted,

indicating that the system requires BgaR for the activation of gene expression. In contrast, deletion of the hypothetical operator sequence O1 did not result in a significant difference to gene expression when compared to the native system in the presence of the inducer (unpaired two-tailed Student's *t-test*; $p \leq 0.0332$). Interestingly, this was not the case in the absence of lactose, evidenced by a significant difference in background expression between the construct lacking O1 and the native system. According to previous results, O1 is located 29 bp upstream of the TSS (distance relative to the 10^{th} base of the 20 bp of O1) and overlaps the determined putative -10 regulatory element. It is possible that the deletion of O1 affects the promoter region of *bgaR*, which by pure coincidence, results in an increased expression of *bgaR*, and as a consequence, of the background level of expression. Alternatively, it could be possible that O1 includes an operator sequence that serves to negatively regulate the expression of BgaR; as such, increased levels of BgaR would lead to an increased activation of gene expression driven by P_{bgaL}. In all other cases, background expression was slightly higher than that of the promoter-less reporter vector but lower if compared to the native system. Since this also occurs in the absence of *bgaR*, where background expression was reduced by 50%, it suggests that this basal expression is the result of the leaky transcription from P_{bgaL}. In the same context, promoter leakage of the native system is likely the result of the combinational leakiness derived from both, P_{bgaR} and P_{bgaL}.

Having established that O2 might be involved in P_{bgaL} activation, further investigations of the importance of the DNA sequence around this motif were pursued via mutational analysis. Several pMTL-HZ1-derived mutant plasmids with truncated or mutated sequences located upstream of the TSS were therefore

generated employing a previously reported mutagenesis strategy[92]; P_{bgaL} activity was quantified by determining levels of CAT activity in lysates of plasmid-carrying *C. sporogenes* strains. A total of eleven truncated variants were constructed, termed pMTL-ICT1 to -T11, by sequentially deleting five nucleotides between positions -89 and -34 (where the -35 element resides) (Figure 3.12a). Additionally, a total of eleven mutated variants were constructed, named pMTL-ICM1 to -M11, whereby blocks of five nucleotides were substituted by their complementary base pairs, between positions -89 and -34 (Figure 3.12b).

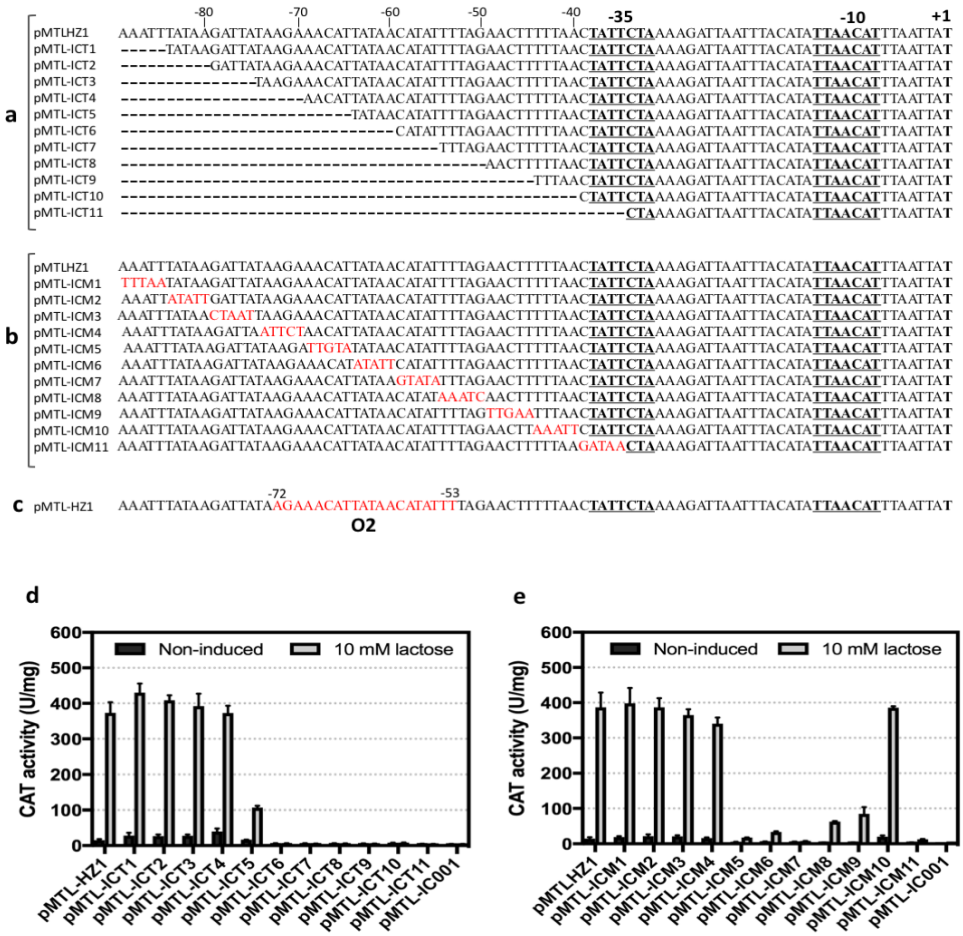


Figure 3.12 LAC stepwise mutations and truncations. Analysis of the intergenic region of the LAC system by CAT measurements of *C. sporogenes* strains carrying pMTL-HZ1-derived mutant plasmids. The nucleotide sequences of the WT (pMTL-HZ1) and the mutants are shown. **(a)** DNA sequence of the LAC system variants with truncated intergenic region. **(b)** DNA sequence of the LAC system variants with blocks of five nucleotide substitutions (highlighted in red). **(c)** Sequence of pMTL-HZ1 intergenic region with the consensus motif O2 highlighted in red. **(d)** CAT results of variants with truncated 5´ UTR (sequences from (a)). **(e)** CAT results of variants with five nucleotide substitutions (sequences from (b)). In both cases, the promoter-less reporter plasmid, pMTL-IC001, was used as negative control. Cells containing the designated plasmids were cultivated in TYG medium in the presence and absence of 10 mM lactose; induction took place when cultures had reached early exponential growth phase ($OD_{600} \approx 0.5$). CAT activity was determined on cell lysates derived from stationary phase cultures. Error bars represent the standard deviation of three biological replicates.

Truncations downstream of position -69 (pMTL-ICT5), within the 5´ sequence of P_{bgaL} caused a strong decrease in promoter activity from position -69

(pMTl-ICT5). In particular, deletions downstream of position -64 (pMTL-ICT6) resulted in the same level of gene expression previously observed when deleting the consensus motif O2 or the transcriptional factor *bgaR*, which corresponds to the background expression of P_{bgaL}. Additionally, substitutions of 5 bp length within the same region provided extra information. Mutations between positions -89 and -74 did not have a significant effect on P_{bgaL} activity, relative to the native system. Substitutions between positions -74 and -69 slightly decreased gene expression (pMTL-ICM4), whereas those occurring between positions -69 and -44 had a more severe effect on P_{bgaL} expression. In some cases, such as in the substitutions between positions -69 to -64 and -59 to -54, corresponding to vectors pMTL-ICM5 and pMTL-ICM7, respectively, gene expression was almost abolished. Interestingly, substitutions between the positions -44 and -39 did not have any effect on P_{bgaL} expression (pMTL-ICM10), whereas the five nucleotide substitution immediately after this (between positions -39 and -34, pMTL-ICM11) resulted in a level of gene expression comparable to the background. This result supports the validity of the previously determined position of the putative -35 element, which is partially deleted in pMTL-ICM11.

Accordingly, it is proposed that the regulatory elements that facilitate expression from P_{bgaL} are located within the 30 nucleotides located between positions -74 and -54. However, expression is only severely reduced when less than 70 bases from the original sequence are present. These elements, or some of them, are located within the consensus motif O2, previously found by multiple sequence alignment and motif search. Additionally, CAT data from this study support that BgaR is the TF that activates P_{bgaL} expression and that BgaR binding also occurs in the absence of lactose. Unfortunately, experiments

required to further confirm these findings (i.e., Electrophoretic Mobility-Shift Assays, EMSA[162]), have not yet been performed. Since AraC/XylS type regulators involved in carbon metabolism act as dimers, it is expected that the LAC system, whose regulator belongs to the same family, functions as a dimer. An N-terminal AraC-like binding and dimerization domain could also be found in the amino acid sequence of BgaR (BLAST search – data not shown). In fact, the 5´ region of P_{bgaL} has two perfect and two similar repeats located within the 30 nucleotides involved in P_{bgaL} activation. This sequence is: AGAA-N_6-TAACA-N_6-AGAA-N_5-TAACT, and it is located between positions -72 and -37 (Figure 3.13). Accordingly, the first base of the first repeat at position -72 coincides with the position of the hypothetical operator sequence O2. The severe decrease in gene expression observed when less than 70 bases from the original sequence are present, coincides with the deletion of the first perfect repeat of this hypothetical BgaR binding site. These results are also in accordance with most of the substitutions made within the 5' sequence of P_{bgaL}. An exception is the case of the five nucleotide substitution located between positions -44 and -39, corresponding to the vector pMTL-ICM10, where P_{bgaL} expression was not affected. Here, the three last residues of the five nucleotide substitution (TTTAA) correspond with the first three nucleotides of the second imperfect repeat of the hypothetical BgaR binding site (TAACT). One plausible explanation is that since it is not a deletion, but rather the substitution of AAATT for TTTAA, where the last three nucleotides (ATT) were substituted by TAA in the hypothetical imperfect repeat, this still forms a functional BgaR binding site. However, this possibility can only be speculative without further analysis. Demonstration of the presence of one or more operator sequences could be

achieved via DNase footprinting[163], a common method for both identifying and characterizing DNA-protein interactions as well as via EMSA. Such assays could also be employed to test if O1 includes an operator sequence that serves to regulate the expression of BgaR. Many TFs regulate their own transcription rate[164]. In fact, more than 50% of the TFs in *E. coli* are known to auto-regulate via negative or positive regulation or in combination with other TFs[165]. Although further research of the LAC system from *C. perfringens* strain 13 is taken into consideration, this work has provided a general outline of the mechanisms behind its functionality, which was considered adequate for the specific purpose of this study: to generate a synthetic system with both, reduced background expression and large dynamic range.

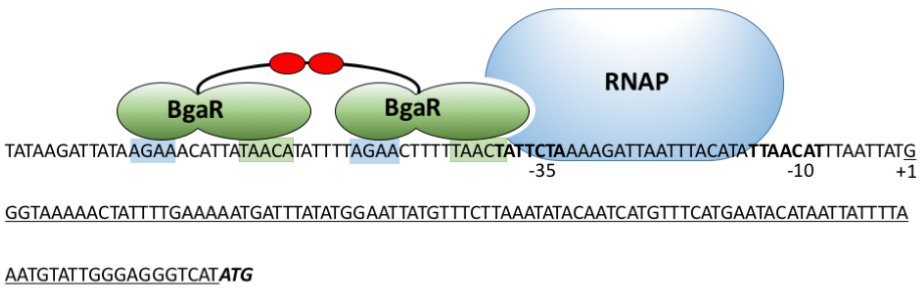


Figure 3.13 Hypothetical BgaR binding sites proposed in this study and hypothetical regulation of the LAC system. In the presence of lactose (red circles), BgaR dimerizes and each monomer binds to a 15-bp half-site containing similar (but not identical) motifs. Perfect repeats of the hypothetical BgaR binding site AGAA-N_6-TAACA-N_6-AGAA-N_5-TAACT are highlighted in blue. Imperfect repeats are highlighted in green. The 5´UTR sequence is underlined. The translational start site is in bold and italicised. According to the position of the putative -10 and -35 elements (in quite proximity to RNAP), activation of P_{bgaL} could occur via direct interaction with the RNAP or by creating a conformational change in the DNA that increases the affinity between the RNAP and the DNA.

3.4 Key outcomes

1. Successful demonstration of the functionality of the *C. perfringens* LAC system in *C. sporogenes*, characterization of its dose-dependency and kinetics as well as the establishment of lactose as an effective, non-metabolizable inducer in *C. sporogenes*.
2. Characterization of the intergenic region of the LAC system through: determination of the TSS and putative -35 and -10 elements of both the P_{bgaR} and P_{bgaL} promoters; confinement of the regulatory sequences involved in P_{bgaL} activation to thirty nucleotides between positions -74 and -54, relative to the TSS downstream of P_{bgaL}; establishment of BgaR as the TF responsible of P_{bgaL} stimulation, and; in silico identification of a hypothetical BgaR binding site that fits the experimental results obtained.
3. Mutational analysis of the LAC system revealed that promoter leakage can be significantly reduced in the absence of BgaR, suggesting that the tight control of *bgaR* expression could be a means of reducing the background leaky expression of the system.

Chapter 4
/ Theophylline responsive riboswitches in *C. sporogenes*

"In the beginning, there was simplicity"
– Richard Dawkins, The Selfish Gene

4.1 Background

4.1.1 Riboswitches: RNA aptamers to control gene expression

RNA molecules have long been known to be involved in an extensive range of regulatory functions in bacteria and archaea[166]. One of the most studied classes of regulatory RNAs are small regulatory RNAs (sRNA). sRNAs are non-coding *trans*-acting elements of between 50 and 500 nucleotides. Generally, they regulate translation or stability of target mRNAs via exposing or sequestering *cis*-acting sites by modifying their secondary structure[167]. Another extensively studied class of regulatory RNAs are riboswitches. Contrary to sRNAs, riboswitches are *cis*-acting elements located in the 5′ UTR region of mRNAs, which regulate gene expression through small ligand-induced conformational switching[166]. The increased understanding of these regulatory RNAs has revealed the sophisticated nature of gene regulation, which, in turn, inspired synthetic biologists and biotechnologists, who are now using RNA as a tool to reprogram cellular functions.

The attractiveness of riboswitches for synthetic biology applications is due to several reasons: (i) they consist solely of RNA. Therefore, they are easy to implement, since only the transfer of a single genetic element, generally smaller than 100 bp, is needed[168]; (ii) they are modular, so that different expression platforms can be combined with different aptamers domains[169]; (iii) very importantly, they allow dose-dependent and temporal control over target gene expression[168], and; (iv) they have been already successfully used in a wide range of microorganisms for several applications (recently reviewed in[167]).

Natural riboswitches control transcription, translation, ribozyme activity and splicing[168]. Generally, they consist of an aptamer domain, which binds a

specific small ligand, and an expression platform, that undergoes a structural change which serves to regulate gene expression. The most widespread mechanisms of riboswitch regulation are transcription termination and translation inhibition. The first type comprises regulators of transcriptional elongation, which operate by impeding the continuation of the RNAP during RNA synthesis[170]. These elements can regulate the production of both coding and noncoding RNAs and act on operons formed by multiple genes. The second type comprises regulators of translational initiation, which act by means of altering the access to parts of the mRNA that are relevant for translation (i.e., to the RBS)[170]. In contrast to regulators of transcriptional termination, these type of regulators act on individual genes and cannot control the production of mRNA.

Riboswitches were discovered by the combination of two contemporaneous advances. Firstly, *in vitro* evolution of random short RNA libraries showed that these small RNA molecules – the so-called aptamers – were capable of selectively binding to a variety of small molecules, termed ligands[171]. Secondly, there was genetic evidence for regulatory sequences within the 5´ UTR sequence of mRNA that were responsible for feedback inhibition by the metabolite product. This includes the particular cases of the now well-characterized pyrophosphate (TPP)[172,173] and flavin mononucleotide (FMN) riboswitches[173]. It is not surprising, therefore, that since their discovery, several naturally occurring riboswitches have been used for metabolite biosensor or metabolic engineering applications. Examples of the former include the c-di-GMP riboswitch reporter system, which enabled the screening of c-di-GMP-metabolizing enzymes *C. difficile* in heterologous hosts[174] and the B12-

responsive riboswitch, which allowed the study of membrane transporters involved in maintaining vitamin B_{12} homeostasis in *E. coli*[175]. An example of the latter application includes the improvement of lysine yield in the fermenting bacterium *Corynebacterium glutamicum*[176].

While naturally occurring riboswitches seemed to satisfy several synthetic biology applications, it is not surprising that this type of regulation rapidly expanded towards the development of synthetic aptamer-metabolite combinations. In fact, synthetic riboswitches based on artificial aptamers are valuable tools because they can bind non-native ligands, orthogonal to the host organism's metabolites, without affecting vital functions of the host[98]. Additionally, aptamers against small molecules of interest can – in principle – be selected via systematic evolution of ligands by exponential enrichment (SELEX)[177]. During SELEX, a huge library of small RNA molecules are subjected to affinity of a small RNA chromatography against the ligand/s of choice. Bound aptamer-ligand combinations are selected, amplified and subjected to further rounds of selection. As a result, RNA molecules with high affinity and specificity properties for the ligand of choice – the so-called aptamers – can be obtained. However, the small number of aptamers successfully employed in riboswitch-mediated regulation thus far indicates that high affinity and specificity are not the only features required for ligand-dependent control of gene expression. Such is the case of the SELEX-derived aptamer R23, which despite showing high affinity for its ligand, neomycin, remains incapable of mediating ligand-dependent regulation when cloned upstream of a reporter gene[178]. The understanding of the functioning of aptamers that have successfully demonstrated control of gene expression *in vivo*, such as

the theophylline- and tetracycline-binding aptamers, on the other hand, indicate that formation of a highly defined and robust ligand-bound aptamer structure, may be a prerequisite for its riboswitch compatibility[179,180]. It is worth mentioning that most of the synthetic riboswitches developed so far act at the level of translation, possibly because the design of riboswitches that control transcription elongation rely on a moving polymerase, requiring the consideration of kinetic and structural factors that are yet poorly defined[97]. On the other hand, translational-based riboswitches are attractive because RBS-based interactions are increasingly easier to design using predictive thermodynamic models[181]. Of particular interest for synthetic riboswitches is the theophylline-binding aptamer, which will be the focus of the next section.

4.1.2 The theophylline responsive riboswitch

The theophylline-binding aptamer, and in particular the so-named TCT8-4 aptamer, isolated by Jenison and colleagues in 1994[182], is the best characterized SELEX-derived, and most frequently used, aptamer in synthetic riboswitches. This aptamer has high affinity for the purine theophylline. Its discriminatory capacity against structurally closely related purines is also high, with a binding affinity 10000-fold greater than that for caffeine, which only differs from theophylline in a methyl group located at nitrogen atom N-7. Its successful application to create synthetic riboswitches is further favoured by the stability, cell permeability and bio-orthogonality of theophylline, despite being toxic at high concentrations[98,183].

Making use of the TCT8-4 aptamer, Gallivan and coworkers developed several screening methods to identify riboswitches with engineered expression

platforms[86,184,185]. In all cases, thousands of riboswitches with random nucleotide variations within the sequence located between the aptamer and the RBS were screened via a reporter gene (Figure 4.1). As a result, riboswitches capable of displaying low background expression levels in the absence of theophylline, and robust increases in expression in its presence, were identified. In particular, riboswitch 8.1 and its derivative 8.1* exhibited the highest dynamic range and, as such, they were subjected to rational modifications and implemented in diverse bacterial species[101]. In this last study, Gallivan and colleagues generated a library of theophylline responsive variants, named A to E, which established the basis for the implementation of this riboregulator in other organisms. The universality of these synthetic riboswitches has now been demonstrated in a number of both Gram-positive and Gram-negative bacteria, including *E. coli*[101], *Mycobacterium smegmatis*[101], *Streptococcus pyogenes*[101], *B. subtilis*[101,186], *Bdellovibrio bacteriovorus*[187], *Streptomyces coelicolor*[188] and the cyanobacteria *Synechococcus elongates*[189] and *Synechocystis sp.*[190]. However, the performance of these riboswitches, including their activities and induction efficiencies, varies greatly according to the host organism, especially as a consequence of the different strength of the promoters placed upstream. A given riboswitch, therefore, needs to be characterized and well-suited to the specific host.

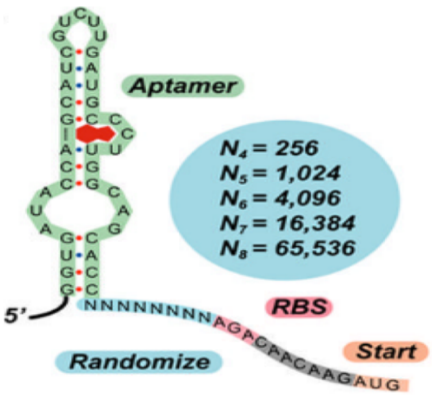


Figure 4.1 The theophylline responsive riboswitch. Diagram of the 5′ region and the randomization strategy followed by Gallivan and coworkers. The aptamer is shown in green, the randomized regions in blue, the ribosome binding site is pink and the AUG start codon in peach. Theoretical library sizes are shown in the circle. Adapted from [86].

4.2 Chapter overview

The widespread acceptance of engineered riboswitches as artificial genetic control elements over the past decade has established an elegant alternative to protein-based inducible systems for synthetic biology and biotechnological applications. Yet, synthetic riboswitches had not been utilised in clostridia. In this chapter, it is shown that previously existing, as well as newly designed theophylline-dependent riboswitches, function in *C. sporogenes* with a response that is dependent on inducer concentration. In particular, the novel riboswitch-G outperformed previously published riboswitches, exhibiting the largest dynamic range and a very low basal expression. By replacing the promoter and modifying the 5′ UTR sequence located upstream of the riboswitch, a library of inducible switches with a full set of dynamic ranges was generated and characterized. This study consolidates the theophylline responsive riboswitch as a novel device to regulate gene expression in clostridia.

4.3 Results and Discussion

4.3.1 Theophylline responsive riboswitches for Clostridium

As previous work had shown that many diverse theophylline-dependent riboswitches were functional in both Gram-positive and Gram-negative bacteria[101], it was decided to evaluate the performance of this inducible translational regulatory system in *C. sporogenes* NCIMB 10696. Accordingly, three riboswitch sequences, riboswitch-D, -E and -E*, which were expected to exhibit higher translation efficiencies after induction, out of the six theophylline-responsive riboswitches developed by Gallivan and coworkers[86,101] were selected. In previous publications, riboswitch-E* was shown to be the best theophylline-dependant riboregulator in Gram-positive bacteria, thanks to the high dynamic range and very low basal expression of the regulated gene[187–190]. In the present study, 4 new riboswitches (named -F to -I) were constructed by rationally modifying the space between the SD sequence and the translational start site of the original riboswitch-E* (Table 4.1).

Table 4.1 Nucleotide sequences of the riboswitches used in this study. The aptamer sequence is underlined. The translational start is bold and italicised. Modifications of the parent riboswitch-E* are indicated in red. Rational modifications were based on known strong RBS sequences in the genus *Clostridium*.

Riboswitch	Sequence (5' ➔ 3')	Modification
-D	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGUAAC AACAAG***AUG***	Riboswitches developed by Gallivan and coworkers[101]
-E	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGU AACAACAAG***AUG***	
-E*	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGC AACAAC***AUG***	
-F	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGU AACAAC***AUG***	This study
-G	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGU AACUUA***AUG***	
-H	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGU GUGUUA***AUG***	
-I	GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCCUGCUAAGGAGGU CAACAAG***AUG***	

A reporter plasmid, pMTL-IC101, derived from the vector pMTL82251[109], was used as a chassis for all of the subsequently created plasmids (Figure 4.2). This backbone contains the native promoter of the *C. sporogenes* ferredoxin gene (P_{fdx}, associated with the protein coding gene *Clspo_c0087*) upstream of the reporter *catP* and serves as a reference to compare CAT expression measurements. pMTL-IC101 also contains the same elements as the reporter vectors used in Chapter 3, pMTL-HZ1 and pMTL-IC001, including the pBP1 Gram-positive origin of replication, the Gram-negative ColE1 origin of replication, the erythromycin resistance gene (*ermB*) and the T_{CD0164} and T_{fdx} terminators, respectively derived from the *C. difficile* CD0164 and *C. sporogenes* ferredoxin genes, here flanking the promoter-gene sequence. Due to this compatibility, the promoter-less backbone previously generated, pMTL-IC001, was also used to detect background expression in the experiments carried out in this chapter.

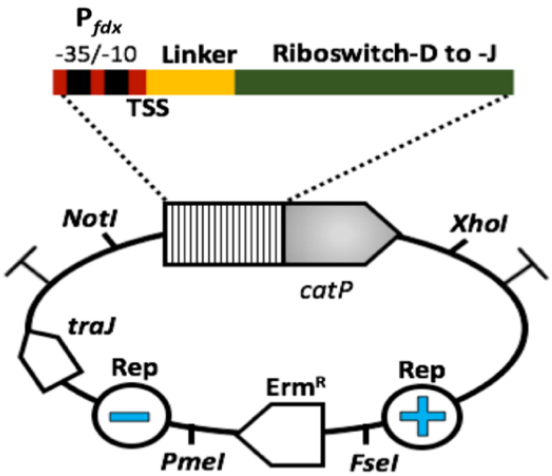


Figure 4.2 Schematic illustration of the riboswitch-reporter plasmid. *E.coli/Clostridium* shuttle plasmid containing the *catP* reporter under the control of P_{fdx}. Riboswitches -D to -J were placed downstream of the TSS. (-) is the Gram-negative ColE1 origin of replication that allows replication of the shuttle plasmid in *E. coli. traJ* encodes the TraJ protein, needed for conjugative plasmid transfer. (+) is the Gram-positive replicon derived from the *C. botulinum* plasmid pBP1.

As a default, the riboswitches, preceded by a linker sequence, were genetically fused to the core region (-35 and -10) of the strong P_{fdx}, just downstream of the TSS and excluding the native 5′ UTR sequence, as previously described[188]. The sequence comprising the promoter-riboswitch-reporter was then cloned into the aforementioned reporter chassis.

In order to place the riboswitch sequences downstream of the TSS of P_{fdx}, the TSS of P_{fdx} had to be determined. To this end, the 5′ RACE procedure described in Chapter 3 was used. Total RNA was extracted from stationary cultures of *C. sporogenes* harbouring the reporter plasmid pMTL-IC101 and 5′ RACE was performed using the same *catP* specific primers described in Chapter 3. As a result, the amplification of the 5′ UTR sequence downstream of P_{fdx} was successfully obtained and the TSS determined (Figure 4.3).

Plasmid constructs, designated pMTL-IC111-D to -J, were conjugated into *C. sporogenes* and cultures of the resultant transconjugant cells exposed to 2 mM theophylline inducer when they had reached the early exponential phase of growth ($OD_{600} \approx 0.5$). CAT activity was determined in cell lysates derived from stationary phase cultures cultivated in the presence or absence of the inducer (Figure 4.4). A detailed list of plasmids used in this study is provided in the Appendix, Table S5 in the methods section.

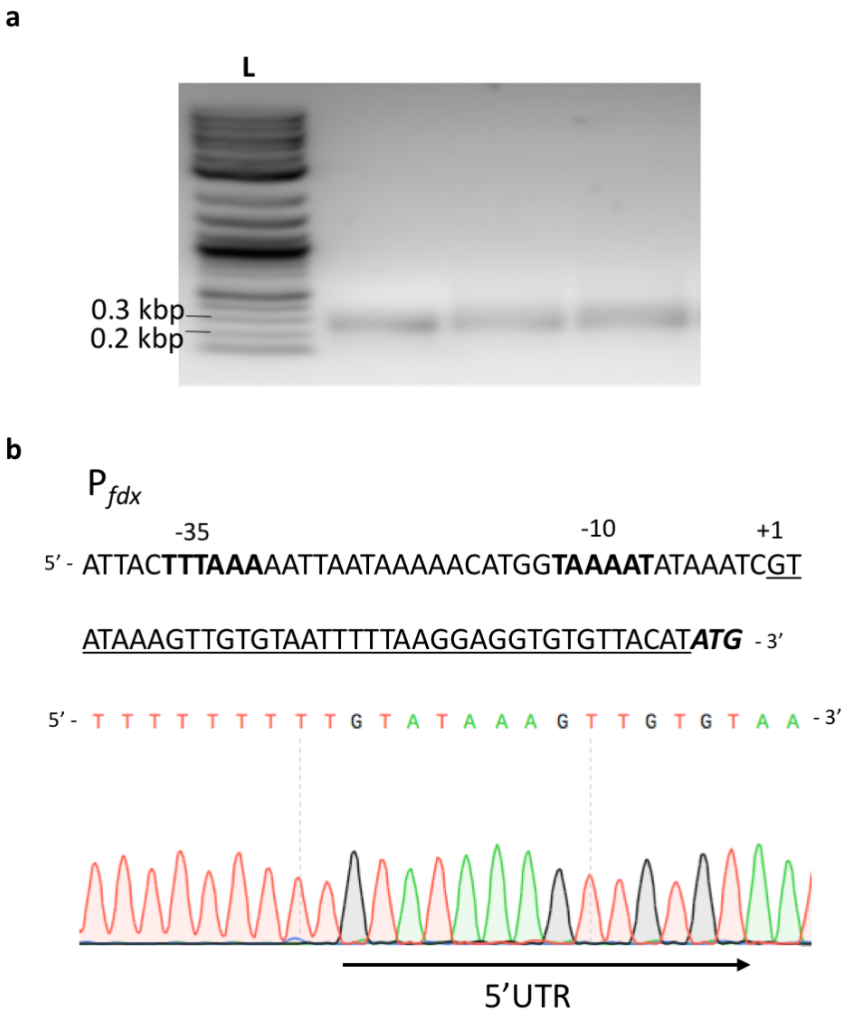


Figure 4.3 5′ RACE experiments on P_{fdx}. **(a)** The agarose gel of the PCR amplification products of three independent DNA samples that were reverse-transcribed from mRNA template according to the 5′ RACE protocol (section 2.5.9). cDNA transcription was performed with a primer aligning 250 bp downstream of the translational start site. Extension PCR was performed with a primer aligning to the adaptor at the 3′ end of the cDNA (corresponding with the 5′ UTR of the mRNA) and a primer aligning 190 bp downstream of the translational start site, producing a product of at least 0.2 kbp. L is 2-log ladder. **(b)** Sequence of P_{fdx} and 5′ RACE sequencing results. The putative -10 and -35 boxes were annotated manually in agreement with the determined TSS (+1). The translational start site is in bold and italicised. The 5′ UTR is underlined.

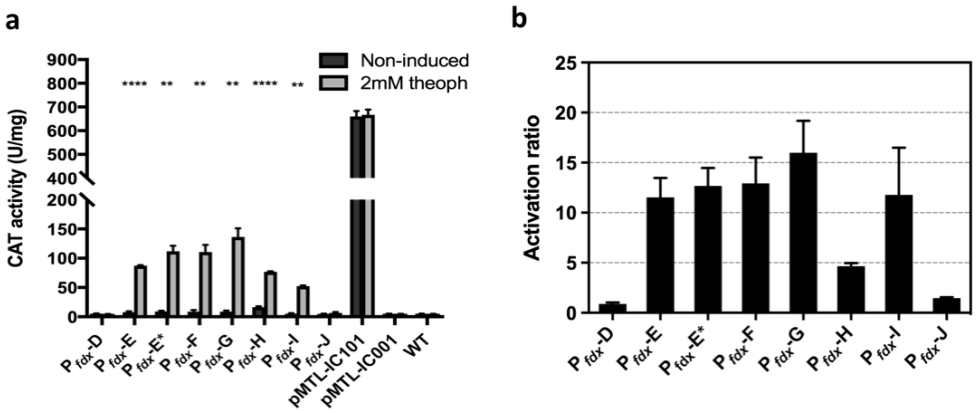


Figure 4.4 Riboswitch library CAT results **(a)** CAT activity and its ligand-dependant induction in each pMTL-111- reporter plasmid. Cells harbouring the different plasmids were grown in TYG medium and induction with 2 mM theophylline took place when they had reached the early exponential growth phase ($OD_{600} \approx 0.5$). The reporter plasmids pMTL-IC101 (P_{fdx}-*catP*), pMTL-IC001 (promoterless *catP*) as well as the WT strain were used as controls. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant induction values for *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001 (paired two-tailed Student's *t*-test). **(b)** Activation ratio of riboswitches –D to –J. The activation ratio in each riboswitch-based reporter plasmid was calculated by dividing the value of CAT activity measured in the presence of the inducer by the value of that in the absence of inducer.

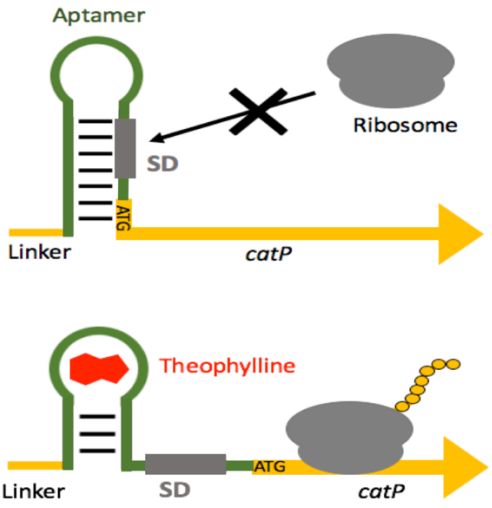


Figure 4.5 Schematic representation of a functional model of the theophylline responsive riboswitch. In the absence of theophylline, the RBS is sequestered via a stem-loop formed in the mRNA. After theophylline binding, the riboswitch conformation changes, releasing the RBS and allowing the translation of the gene of interest (*catP*).

The synthetic theophylline responsive riboswitch is composed of an aptamer and a synthetic SD (Figure 4.5). Theoretically, transcription of the riboswitch under the control of P_{fdx} occurs in a constitutive manner during cell growth. However, the synthetic SD sequence located downstream of the aptamer is sequestered via pairing with the stem of the riboswitch, resulting in translational block. The binding of theophylline releases the SD by altering the downstream base pairing. As a consequence, gene translation occurs when the ribosome binds to the SD[86]. As shown in Figure 4.4a, strains harbouring riboswitches -E, -E*, -F, -G, -H and -I exhibited statistically significant induction in *C. sporogenes* after addition of 2 mM theophylline to the culture. In particular, riboswitch-G outperformed previous theophylline-dependant riboswitches (i.e., riboswitches -E and -E*), demonstrating higher levels of CAT activity, low basal expression and the strongest activation ratio (Figure 4.4b). In all cases, the incorporation of any riboswitch in the 5′ UTR led to a strong reduction in CAT activity; this agrees with previously published studies, indicating that secondary structures near the RBS play a major role in the translation of the downstream mRNA[86,187,188]. In agreement with the assays performed in Chapter 3, CAT activity of cells harbouring the control pMTL-IC001 was similar to that from cultures lacking the reporter backbone (WT), indicating no detectable transcriptional read-through and demonstrating the suitability of the selected backbone as an insulated chassis for the construction of switchable systems.

4.3.2 Adjustability of the theophylline responsive riboswitches

Due to the importance of having inducible systems where gene expression can be controlled within a large regulatory window, it was decided to compare the expression pattern of genes regulated with the riboswitches -E, -G and -H when placed downstream of the core region of a synthetically weakened P_{fdx}, named P_{fdx4}, creating vectors pMTL-IC211-E, -G and -H. The promoter P_{fdx4} was generated by replacing the core elements -10 and -35 of P_{fdx} with the same regulatory elements of the constitutive promoter from the *C. acetobutylicum* ATCC 824 *ptb* gene (P_{ptb}, associated with the protein coding gene *Ca_3076*) (Figure 4.6a). To further examine the regulatory response of the theophylline responsive riboregulators in clostridia, riboswitches -E, -G and -H were additionally fused to the promoters P_{fdx} and P_{fdx4}, retaining the bases downstream of the TSS but excluding their native SD sequences, as previously described[186]. These sequences, named P_{fdx}* and P_{fdx4}*, maintain the full upstream region, including the core region (-35 and -10), the TSS and the space between the TSS and the native SD. Constructs, named pMTL-IC121-E, -G and -H (for P_{fdx}*) and pMTL-IC221-E, -G and -H (for P_{fdx4}*), were designed to express the reporter gene *catP*, conjugated into *C. sporogenes* and the growing cultures of the transconjugants obtained exposed to 2 mM theophylline inducer or grown without induction. Since the linker sequence located between P_{fdx} and

the riboswitch slightly differs from that placed between P_{fdx}* and the riboswitch,

P_{fdx}_Riboswitch X

5′
TACAGTTATATAAAAATTACTTTAAAAATTAATAAAAACATGGTAAAAT
+1
ATAAATCGGTACCAATACGACTCACTATAGGTTCCGGTGATACCAGCA

TCGTCTTGATGCCCTTGGCAGCACCCTGCTAAGGAGGXXXXXXX***ATG***
3′

P_{fdx}*_Riboswitch X

5′
TACAGTTATATAAAAATTACTTTAAAAATTAATAAAAACATGGTAAAAT
+1
ATAAATCGTATAAAGTTGTGTAATTTTTCAATACGACTCACTATAGGTG

ATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAAGGAGGX

XXXXXX***ATG***
3′

Figure 4.7 shows the exact nucleotide residues of both linker sequences.

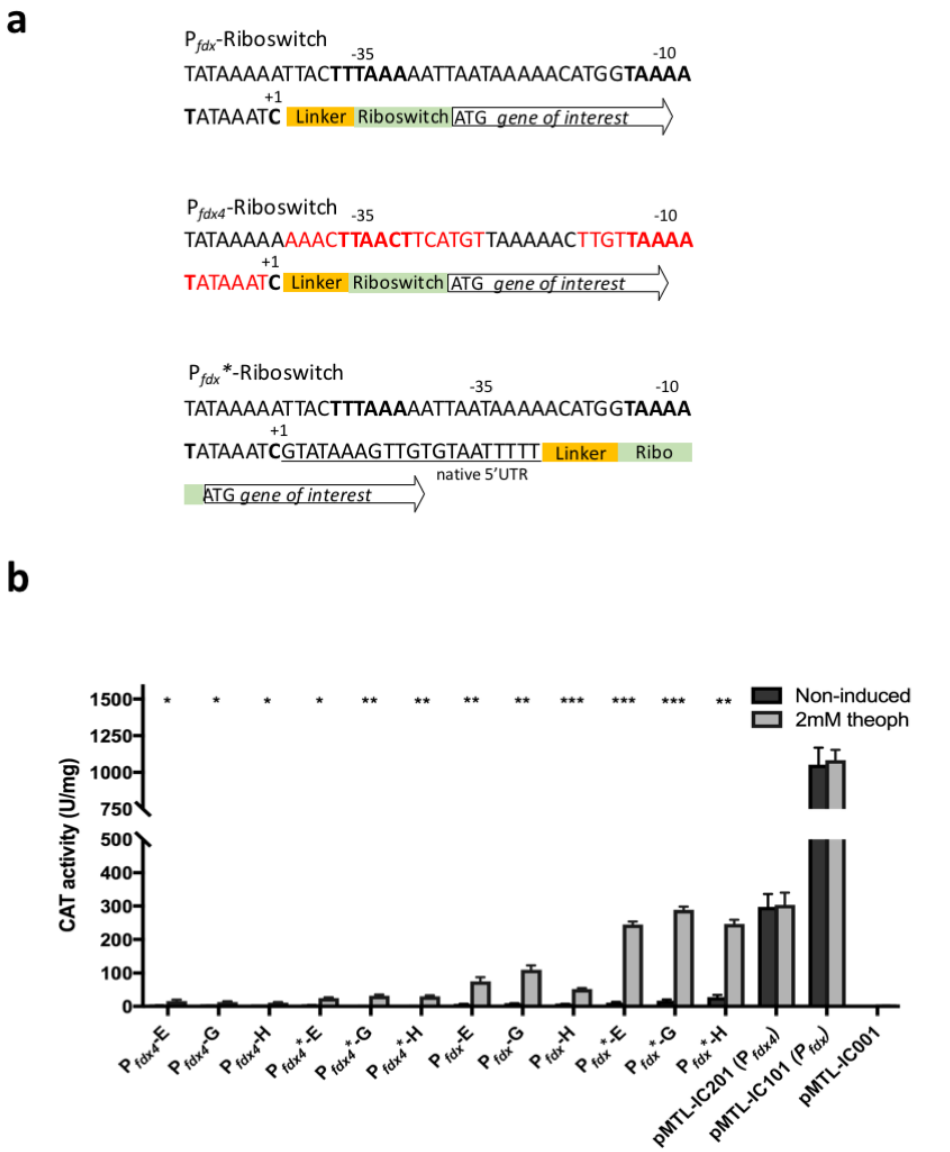


Figure 4.6 CAT results of theophylline responsive riboswitches with different 5′ UTRs **(a)** The sequences of the constructed theophylline-responsive switches. The putative -35 and -10 sequences are in bold. The TSS is indicated as +1. The 5′ UTR sequences downstream of the TSS are underlined. Core elements replaced to create the synthetic hybrid promoter P_{fdx4} are in red. **(b)** Library of theophylline-dependent riboswitches tested in *C. sporogenes* in the presence and absence of 2 mM theophylline. Induction took placed at 4 hours, when cells had reached the early exponential growth phase ($OD_{600} \approx 0.5$) and CAT activity was determined on cell lysates derived from stationary cultures. The reporter plasmids pMTL-IC201 (P_{fdx4}-*catP*), pMTL-IC101 (P_{fdx}-*catP*) and pMTL-IC001 (promoterless *catP*) were used as controls. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant induction values (paired two-tailed Student's *t*-test).



Figure 4.7 Nucleotide sequences of the theophylline responsive riboswitch located downstream of P_{fdx} and P_{fdx}*. The putative promoter -35 and -10 sequences are shaded in grey. The TSS is indicated with +1. The linker sequence is underlined. The aptamer sequence and the RBS are highlighted in red and blue respectively. Xs represent riboswitch-specific nucleotides. The translational start site is in bold and italicized. The linker sequences from P_{fdx} and P_{fdx}* were obtained from[101] and[186], respectively.

As shown in Figure 4.6b, the combination of different riboswitches with various promoters and two different 5′ UTRs allowed the expansion of the regulatory range, providing a library of theophylline-inducible switches suitable for applications where protein yield is crucial as well as for the expression of detrimental or toxic proteins. The combination of any riboswitch with the weak promoter P_{fdx4} led to the lowest expression levels. This was also the case when the native 5′ UTR sequence was retained, establishing a direct correlation between detected CAT activity and promoter strength (P_{fdx}> P_{fdx4}). However,

basal expression appears to be compromised when high expression levels are achieved, resulting in increased leakiness in those constructs that exhibited higher expression levels (i.e., with P_{fdx}*-G). Expression of *catP* was always higher when the native sequence downstream of the TSS was maintained. Since riboswitch-G showed the highest induction level of the analysed riboregulators, it was subjected to further characterization.

4.3.3 Further characterization of the theophylline responsive riboswitches

Although the primary aim of this study was to determine whether the theophylline responsive riboswitch works in clostridia, it was also decided to further characterize certain aspects of riboregulator function. It is well known that the RNA structure plays an important role in controlling both mRNA transcription[191] and protein synthesis[192,193]. It follows that the response of the theophylline responsive riboswitch can be altered, or even abolished, by modifying the 5′ UTR[187]. This also implies that the performance of a given riboswitch will inevitably be dependent on the genetic context. Therefore, it was decided to determine whether the differences in inducible CAT expression observed when using different 5′ UTRs were solely due to the abundance of mRNA transcripts. To this end, mRNA transcribed under the control of riboswitch G located downstream of either P_{fdx} or P_{fdx}* (P_{fdx} along with the native 5′ UTR sequence) was quantified via RT-qPCR (see section 2.5.10). This technique allows the comparative quantification of specific mRNA transcripts relative to the levels of other mRNA, termed "reference genes"[194]. In this study, the genes *16Srrn* (the reference gene most commonly used in bacterial

studies[195]) and *gyrA* were used as references because they had demonstrated to be the most stable candidate reference genes in a study that evaluated various reference genes for RT-qPCR analysis in the closely related organism *C. botulinum* ATCC 3502[196]. Relative expression analyses using either single or both reference genes resulted in the same expression pattern (considering mean errors), suggesting that the selected reference genes were appropriate for qPCR studies in *C. sporogenes*, since they seem to be stably expressed and their expression not affected by the experimental conditions[197] (Figure 4.8). However, since the invariant expression of these reference genes under the experimental conditions of this study was not determined *per se*, both reference genes were considered for data analysis, providing a more robust normalization of the gene expression levels[198]. Results showed a 2.5-fold increase in the mRNA transcribed from P_{fdx}* compared with the mRNA transcribed from P_{fdx}, in cultures induced with 2 mM theophylline when they had reached early exponential growth phase ($OD_{600} \approx 0.5$) (Figure 4.8b). It should be noted that this increase in mRNA levels was accompanied by a 2.6-fold increase in CAT activity (Figure 4.6b), which suggests a direct link between the relative abundance of *catP* transcripts and *catP* expression. For ease of comparison, these data are further summarized in Table 4.2. The levels of mRNA when P_{fdx} is placed upstream of the riboswitch without its native 5′ UTR were similar in the presence or absence of the inducer. Interestingly, there was a significant difference in the amount of transcript between the on/off state in the case of P_{fdx}*, abundance being higher when the inducer was added. These differences suggest that the sequence downstream of the TSS plays an important role in mRNA abundance, possibly as a result of a complex interplay between the

processes of transcription, translation and mRNA degradation. Plausible explanations for these findings could be that the additional nucleotides downstream of the TSS create an RNA conformation that is less susceptible to degradation or that the amount of transcript is indirectly influenced by the conformation of the mRNA (i.e., with or without theophylline) and its associated translation efficiency[192]. No published information regarding riboswitch function and mRNA levels could be found in the literature.

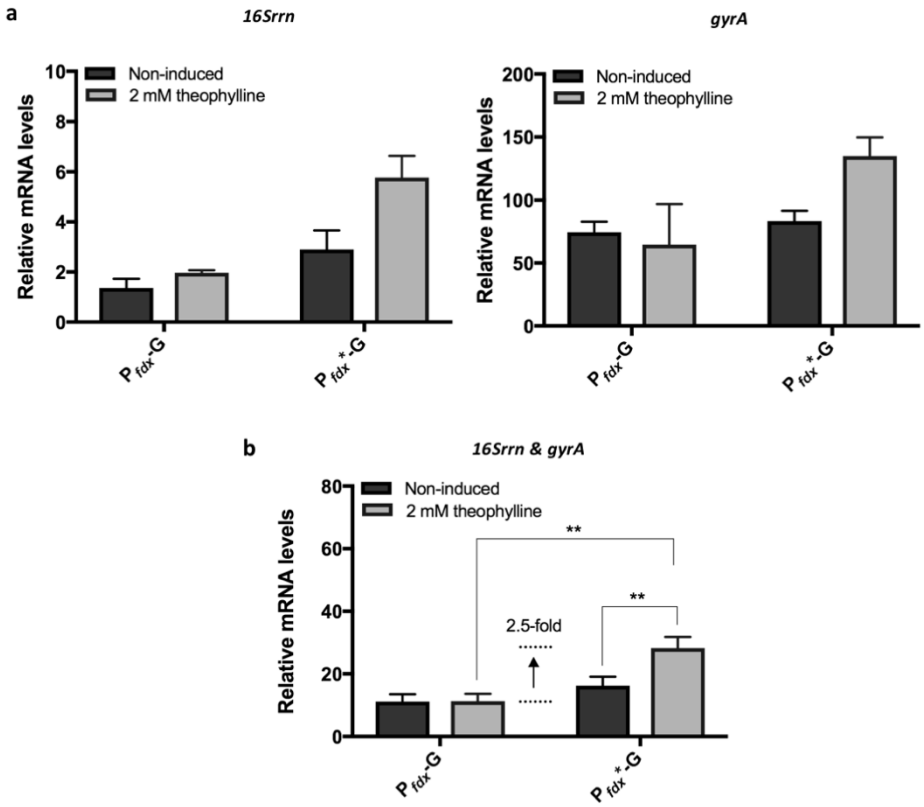


Figure 4.8 RT-qPCR results. *catP* transcript abundance under the control of either P_{fdx}-riboswitch-G or P_{fdx}*-riboswitch-G relative to **(a)** single reference genes 16*Srrn* or *gyrA*; and **(b)** both reference genes 16*Srrn* and *gyrA*. Total RNA was isolated from late exponential cultures grown in the presence and absence of 2 mM theophylline. Error bars represent standard deviations of three biological replicates. Statistically significant differences are indicated with asterisks (paired/unpaired two-tailed Student's *t*-test).

Table 4.2 Comparison of CAT activity results from cells carrying the different 5′ UTR versions upstream of riboswitch G in the presence or absence of theophylline. The mean values and standard deviations represent the CAT results from three biological replicates 4 hours after induction. P_{fdx}*/P_{fdx} ratio is calculated by dividing CAT activity of cells harbouring P_{fdx}* by that of cells harbouring P_{fdx}. The induction factor represents CAT activity of induced cells to CAT activity of non-induced cells.

	Uninduced		Induced		
	CAT activity (U/mg)	**P_{fdx}*/P_{fdx} *ratio***	**CAT activity (U/mg)**	**P_{fdx}*/P_{fdx} *ratio***	**Induction factor**
P_{fdx}	7.733±1.35	2.159	109.927±13.03	2.623	14.21
P_{fdx}*	16.69 ±3.45		288.412±9.80		17.21

Finally, the riboswitches were analysed using the Riboswitch Calculator, which applies a statistical thermodynamic model to predict riboswitch function[181]. This model calculates the activation fold of the riboswitch by considering both the free energy change from a folded mRNA in the absence of the ligand and the ribosome and the ligand-bound mRNA in complex with the ribosome. The model appeared to predict the main aspects of the observed riboswitch performance, including the relative difference between P_{fdx}* and P_{fdx} and the low expression levels and tightness of P_{fdx}-G (Figure 4.9). Altogether, these data indicate that the studied riboswitch responds to theophylline regardless of the genetic context and that the appropriate combination of a riboswitch with a certain promoter and the characterized 5′ UTRs would, therefore, allow for the adjustment of a desired regulatory window.

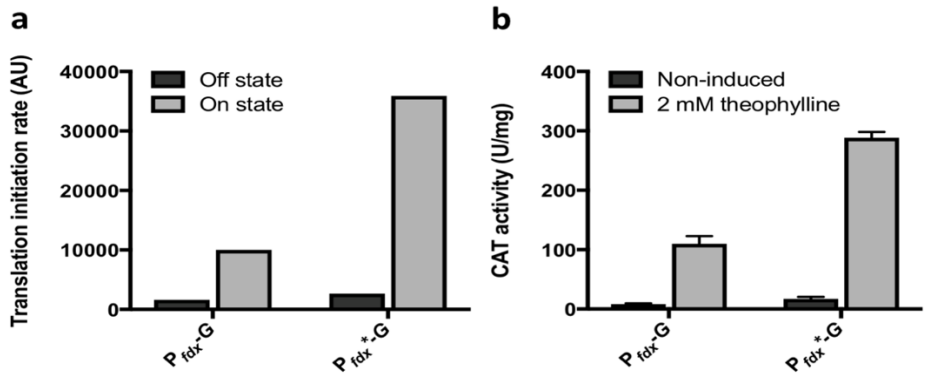


Figure 4.9 Predicted and observed behaviours for riboswitch-G downstream P_{fdx} or P_{fdx}* in *C. sporogenes*. **(a)** Predicted riboswitch responses using the Riboswitch Calculator from the Salis Laboratory [181]. **(b)** Observed riboswitch responses (same data as in Figure 4.6).

4.3.4 Functional characterization of the theophylline responsive riboswitch in C. sporogenes NCIMB 10696

To assess the dose dependency of riboswitch regulation, strains harbouring plasmids that carried the riboswitch-G downstream of P_{fdx} or P_{fdx}* were cultivated with increasing concentrations of theophylline, ranging from 0.1 to 10 mM (Figure 4.10a), and the optical density (OD_{600}) of the cultures monitored over time (Figure 4.10b). Increasing doses of theophylline would be expected to result in higher levels of gene expression. Concentrations higher than 5 mM, however, resulted in a reduction of growth, in agreement with previously reported results in other Gram-positive and Gram-negative bacteria[199]. The translational switches were also analysed for CAT expression over time in the presence and absence of 2 mM theophylline. As shown in Figure 4.10c, CAT expression above the level of the uninduced culture was already detectable 30 minutes after induction (4.5 hours of growth). The maximum level of reporter gene expression was reached at 8 and 10 hours when using P_{fdx} and P_{fdx}*, respectively. Although both have a similar profile of induction, after 24

hours, CAT expression was reduced by 75% in P_{fdx} whereas only by 40% in P_{fdx}*. Given that in both cases the translated product is the same (CAT), these results suggest that the reversal of induction is strongly dependent on the abundance or stability of the mRNA.

As with lactose in Chapter 3, it was decided to establish whether theophylline is metabolized by *C. sporogenes*. *C. sporogenes* cells were, therefore, cultivated in TYG medium supplemented with 2 mM theophylline and the concentration of theophylline was monitored by HPLC/MS analysis of cell-free supernatant samples over time. Results showed that the concentration of theophylline remained constant over the course of the experiment, with a slight increase in the culture supernatant 20 hours after induction, possibly due to the release of the inducer from the cells after lysis (Figure 4.10d).

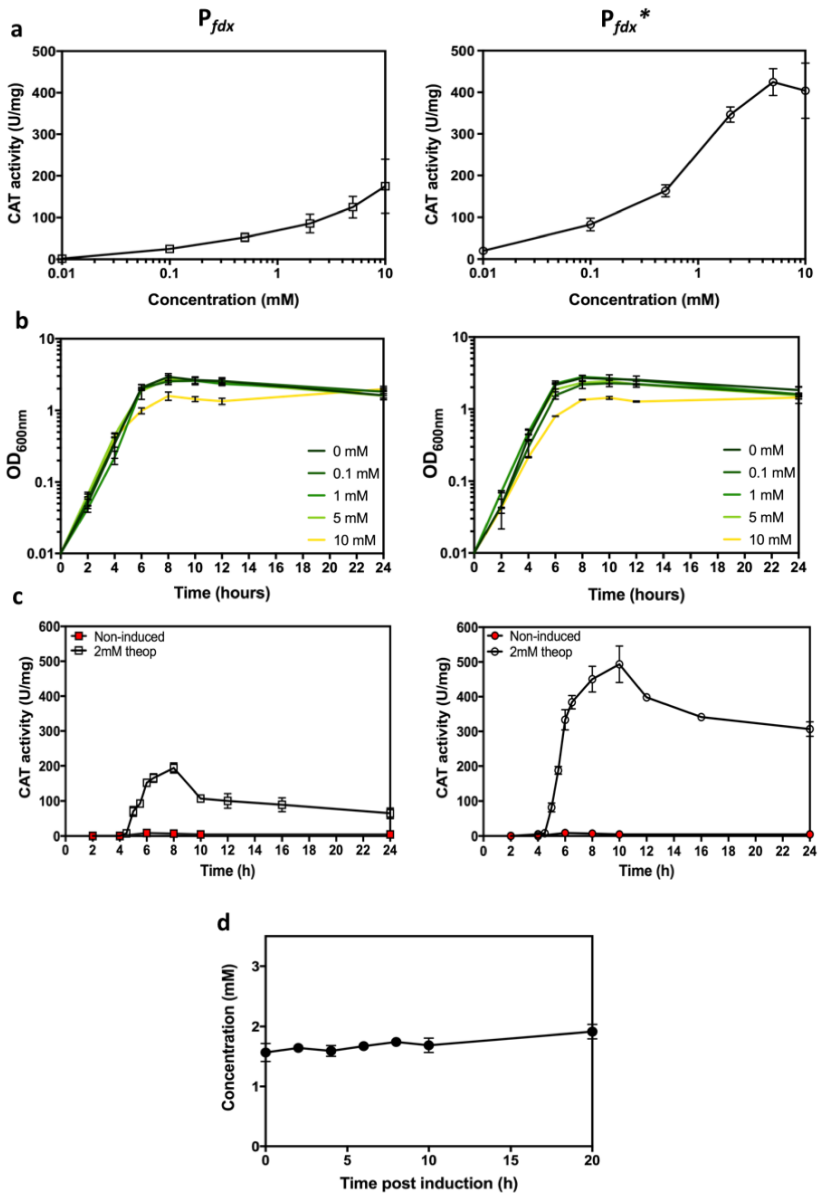


Figure 4.10 Dynamic and kinetic profiles of the theophylline responsive riboswitch located downstream of P_{fdx} or P_{fdx}*. **(a)** Response to different concentrations of the inducer theophylline. Cells containing the reporter plasmid with riboswitch G downstream of either P_{fdx} or P_{fdx}* were cultivated in TYG medium supplemented with various concentrations of theophylline (0, 0.1, 0.5, 2, 5 and 10 mM) at early exponential growth phase (4 hours of growth, $OD_{600} \approx 0.5$); CAT activity was measured in cell lysates derived from stationary phase cultures. **(b)** Optical density (OD_{600}) of *C. sporogenes* harbouring the same reporter plasmids in response to 0.1-10 mM of theophylline supplemented at time zero. **(c)** CAT expression profiles over time. Cells harbouring the reporter plasmids were cultivated in TYG medium in the presence or absence of 2 mM theophylline; CAT was measured on cell lysates from stationary phase cultures. **(d)** Stability profile of theophylline in *C. sporogenes* cultures over time. In all cases, error bars represent the standard deviations of three biological replicates.

Altogether, these data demonstrate that the theophylline responsive riboswitches function in *C. sporogenes* with a response that is dependent to inducer concentration. In particular, the rationally designed riboswitch-G outperformed previously published riboswitches[101], exhibiting very low basal expression and the largest dynamic range. The regulatory window of the theophylline responsive riboswitches could be adjusted by modifying the 5′ UTR sequence located upstream of the riboswitch or by using promoters with different strengths. However, the incorporation of any riboswitch in the 5′ UTR led to a strong reduction in reporter activity in all cases, with higher levels of gene expression linked to increased leakiness in the absence of the inducer. This, in fact, compromised the dynamic range of the riboregulators. In all cases, the riboswitch was able to respond to the inducer theophylline, ensuring its performance independent of the genetic context. Additionally, because *C. sporogenes* does not degrade theophylline, this is confirmed, thus far, as an appropriate inducer for further applications.

4.4 Key outcomes

1. First implementation of synthetic riboswitches in the genus *Clostridium*, including the generation of a riboswitch (riboswitch-G) that outperforms previously published riboswitches, showing very low basal expression and a larger dynamic range.
2. Generation of a library of various promoter-riboswitch combinations that exhibited several dynamic ranges, suitable for applications where a tight control of gene expression is needed as well as for the expression of large amount of protein products.
3. Further demonstration of the functioning of the theophylline responsive riboswitch independent of the genetic context.
4. Characterization of the dynamics and kinetics of the theophylline responsive riboswitch when placed downstream of different 5′ UTRs.

Chapter 5
/ A dual mechanism to tightly control gene expression – RiboLac

"Progress in science depends on new techniques, new discoveries and new ideas, probably in that order"

– Sydney Brenner

5.1 Background

While there has been some progress in improving the dynamic range of inducible systems by controlling a single gene expression process, only a handful of studies have explored the idea of regulating multiple regulatory processes to tightly control gene expression[97,200–202]. Specifically, Liu and colleagues described a method to convert regulators of translational initiation into regulators of transcriptional elongation to achieve tight control of gene expression by controlling both, transcription and translation[97]. Morra *et al*., combined the LAC promoter system from *E. coli* and the orthogonal Pyrimido[4,5-d]pyrimidine-2,4-diamine (PPDA) riboswitch to achieve dual transcriptional-translational control over both, T7 RNAP expression and the recombinant gene of interest[200]. With this system, termed *RiboTite*, more than 800-fold induction was achieved; however, maximum expression was reduced by 40% in the presence of both inducers, IPTG (allolactose analogue) and PPDA, if compared with the expression obtained when the reporter was regulated by only the LAC system. Horbal and Luzhetsky also combined the cumate inducible system with the theophylline responsive riboswitch to control pamamycin production in *Streptomyces albus*[201]. However, reduction of leaky expression in the OFF state was achieved by an increased control of the expression of the target gene, which is regulated both at the level of transcription and translation. As such, and similarly to the *RiboTite*, maximum expression was reduced by 50% if compared to the same gene regulated by the cumate system alone. Finally, Westbrook and Lucks used RNA-mediated transcriptional regulators to regulate transcription and translation in *E. coli*[202]. The dual control improved expression from 10-fold to 923-fold due to both a higher ON level and

a lower OFF level, demonstrating this to be one of the most tightly regulated inducible systems thus far. In order to achieve such control, a transcriptional antisense and a small transcription activating RNAs (STAR), which is designed to contain an anti-terminator sequence complementary to the 5′ half of the terminator stem in the transcriptional antisense, are needed. Since the terminator is placed in close proximity to the RBS, addition of a STAR prevents termination and allows transcription and translation to proceed. The activity of the STAR is therefore the key function of the system, which makes use of multiple plasmids – one plasmid encoding the transcriptional antisense in combination (or not) with a second plasmid encoding the STAR. A second alternative to regulate the presence and absence of the STAR could be via an inducible system. However, the functionality of the system would rely on the performance of the inducible system and its associated leakiness, which can rarely result in the level of regulation observed when using several plasmids.

5.2 Chapter overview

Having established the limitations of the existing transcriptional inducible systems to tightly control gene expression in *Clostridium*, it was decided to engineer a system with improved characteristics so that high levels of gene expression do not compromise the dynamic range. By combining the theophylline responsive riboswitch with the transcriptionally regulated LAC system from *C. perfringens* strain 13, a dual translational-transcriptional ON device that controls the expression of both, *bgaR* and the gene of interest, was generated. In this system, termed RiboLac, optimal expression of the target gene requires induction with both theophylline and lactose. Functionality of the

system was also examined via the study of its kinetics and dynamics. Additionally, to determine whether the degree of repression and dynamic range would provide suitable levels of gene expression once integrated in the chromosome, the system was further incorporated into the clostridial host genome. The enhanced performance of the RiboLac, which regulates the translation of *bgaR*, which, in turn, regulates the transcription of the target gene, was further demonstrated by direct comparison with an engineered system where the dual control is only exerted over the target gene.

5.3 Results and Discussion

5.3.1 RiboLac, a dual mechanism with improved dynamic range

In Chapter 3, it was demonstrated that the LAC system requires BgaR to activate gene expression. It was also shown that background expression was reduced by 50% if the system lacks BgaR, providing a clue as to how leaky expression could potentially be reduced in this inducible system. Additionally, some of the riboswitches characterized in Chapter 4, in particular riboswitches -E, -G and -H, demonstrated tight control of gene expression in the OFF state, exhibiting diverse expression profiles upon activation. By combining these two types of regulation, it was hypothesized that a system with translational control over *bgaR* and transcriptional control over the target gene (*catP*) could be generated, in which expression was subject to the addition of two inducers, lactose and theophylline. Since the amount of BgaR required to activate gene expression in the native inducible system was unknown, it was decided to generate several systems that would lead to various levels of BgaR in both the ON and OFF states. It was expected that one of the systems would allow an adequate expression of BgaR in the presence and absence of the inducer, successfully facilitating controlled activation of P_{bgaL}. Accordingly, three riboswitch sequences, riboswitch-E, -G and -H, were genetically fused to the core region (-35 and -10) of the *bgaR* promoter, P_{bgaR}, just downstream of the TSS (determined in Chapter 3) and excluding the native 5´UTR, generating vectors pMTL-ICE1, pMTL-ICG1 and pMTL-ICH1. The rest of the system remained as it was in the parent LAC reporter vector pMTL-HZ1. Additionally, the same riboswitches were placed downstream of P_{bgaR}*, which retains the nucleotides downstream of the TSS but excludes the native SD sequence, as

described in Chapter 4, generating the vectors pMTL-ICE1*, pMTL-ICG1* and pMTL-ICH1*. Constructs, which contained the reporter gene *catP*, were conjugated into *C. sporogenes* and growing cultures of the transconjugants obtained were exposed to either one of the inducers – theophylline or lactose – to both inducers, or grown without induction.

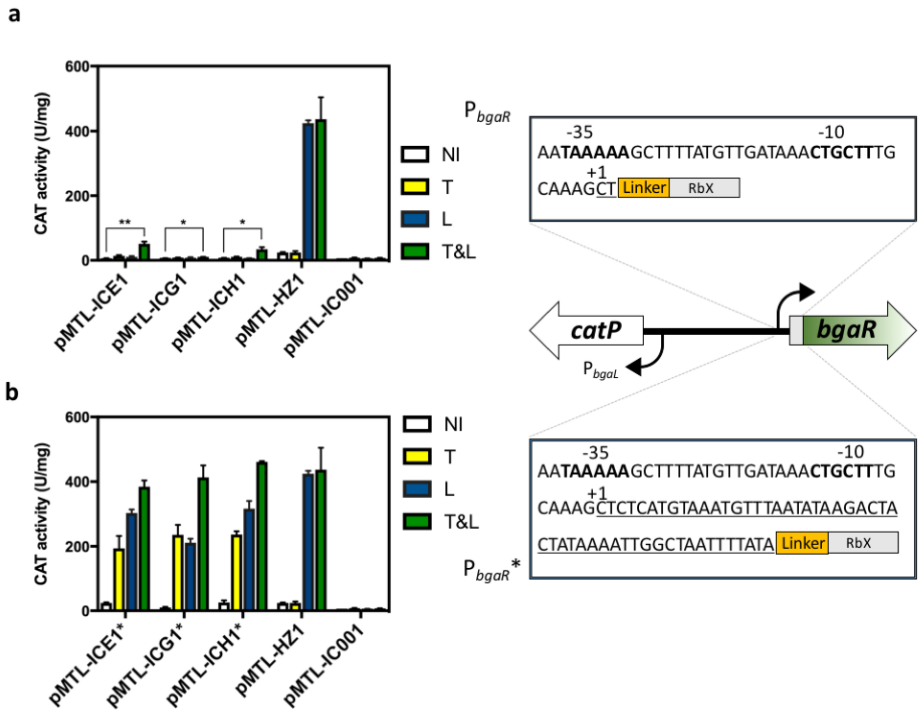


Figure 5.1 Response of the dual theophylline-LAC systems (RiboLacs) in *C. sporogenes* to different concentrations of theophylline and lactose. **(a)** CAT results of the dual systems when the riboswitches are placed downstream of P_{bgaR} and nucleotide sequence of P_{bgaR}. **(b)** CAT results of the dual systems when the riboswitches are placed downstream of P_{bgaR}* and nucleotide sequence of P_{bgaR}*. In both cases, cells harbouring the different plasmids were cultivated in TYG medium supplemented with either 2 mM theophylline (T) or 10 mM lactose (L), with both 2 mM theophylline and 10 mM lactose (T&L) or in the absence of inducers (NI). The vector bearing the LAC system, pMTL-HZ1, and the promoter-less vector, pMTL-IC001, were used as positive and negative controls, respectively. Induction took place at 4 hours, when cells had reached the early exponential growth phase ($OD_{600} \approx 0.5$); CAT activity was determined on cell lysates derived from stationary cultures. Error bars represent the standard deviation of three biological replicates. Asterisks indicate statistically significant induction values for *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001 (paired two-tailed Student's *t*-test). RbX refers to any of the three riboswitches -E, -G and -H.

The CAT data shown in Figure 5.1 demonstrates that the dual theophylline-lactose system, termed RiboLac, responds to both inducers and generates a response that is dependent on BgaR expression. In RiboLacs containing the riboswitch downstream of P_{bgaR}, statistically significant expression of *catP* above the level of the uninduced culture was achieved only in the presence of both inducers, indicating that these dual systems act as Boolean AND logic gates (Figure 5.1a). Specifically, the RiboLac harbouring riboswitch-E (pMTL-ICE1) showed the highest dynamic range (approximately 13-fold), out of the three systems containing P_{bgaR}. In all three cases, expression after induction was severely reduced if compared with the native system, which despite exhibiting considerable background expression, still afforded a higher dynamic range (approximately 15-fold). Conversely, significantly increased expression of *catP* was observed in RiboLacs with riboswitches located downstream of P_{bgaR}* when cultures were supplemented with only one of the inducers (Figure 5.1b, asterisks not shown). However, expression after induction with one inducer was lower than that of the LAC system in the presence of lactose, indicating that the system is capable of the dual regulation of P_{bgaR} and P_{bgaL} at the translational and transcriptional levels, respectively. Interestingly, the system containing riboswitch-G downstream of P_{bgaR}* exhibited very low basal expression in relation to the native system, and maximum expression was not compromised after induction with both inducers. As a consequence, the dynamic range increased from approximately 15-fold in the native system to approximately 41-fold in this specific RiboLac. As expected, theophylline did not have any effect on the expression of the LAC system. Data from this study indicate that expression of BgaR under the control of a riboswitch located

downstream of P_{bgaR} is lower than the expression of the same protein when the riboswitch is placed downstream of P_{bgaR}*, both in the presence and absence of the inducer. Repression of the transcriptional activator in the OFF state is very tight in systems containing P_{bgaR}; induction, however, did not result in the high level of expression observed in the native system. By contrast, systems containing P_{bgaR}* achieve the same levels of expression as the native system in the presence of inducers, subject to higher basal levels of gene expression. These results also agree with the data obtained when characterizing the theophylline responsive riboswitches in Chapter 4, where the same pattern of expression between P_{fdx} and P_{fdx}* (or P_{fdx4} and P_{fdx}4*) was observed. Unexpectedly, expression of *catP* under the control of P_{bgaR}*-based RiboLacs in the presence of only theophylline was higher than that of the native inducible system in the absence of lactose (i.e., background expression level). This suggests that expression of the transcriptional activator when located downstream of the riboswitch could be higher than the expression obtained when regulated by its native promoter. One plausible explanation could be that the introduction of the riboswitch downstream of P_{bgaR}* renders a 5´UTR less susceptible to RNA degradation, and hence, to a higher translation of BgaR (subjected to the addition of theophylline). Since the dual system containing P_{fdx}*-G showed the highest induction level out of the analysed systems, it was decided to characterize it further. Henceforth, the term RiboLac will be used to refer to this particular system in this and subsequent chapters.

5.3.2 Functional characterization of the RiboLac

Having established the functionality of the RiboLac, further experiments were undertaken to determine the dynamics of induction; specifically, the activation of P_{bgaL} when *bgaR* expression is induced with various concentrations of theophylline. Accordingly, strains harbouring the plasmids pMTL-HZ1 and pMTL-ICG1* were cultivated with increasing concentrations of theophylline, ranging from 0 to 2 mM, in the presence and absence of 10 mM lactose (Figure 5.2). Increasing doses of theophylline would be expected to result in higher levels of BgaR, and therefore, increased levels of CAT upon P_{bgaL} activation. This was the case when the system was induced with theophylline alone. Although little change in induction was observed between 0.5 and 1 mM theophylline, addition of 2 mM resulted in a higher P_{bgaL} activation. In no case did induction with one inducer achieve maximum expression levels. Interestingly, the response of the RiboLac seemed to be peak when induced with 0.5 mM theophylline in the presence of lactose, with levels of gene expression similar to those obtained at maximum induction (2 mM theophylline and 10 mM lactose). The expression of RiboLac-dependent *catP* in the presence of lactose alone, which is lower than that of the LAC system under the same conditions, seems to be the result of P_{bgaL} activation when the levels of BgaR are below the saturation capacity of the system.

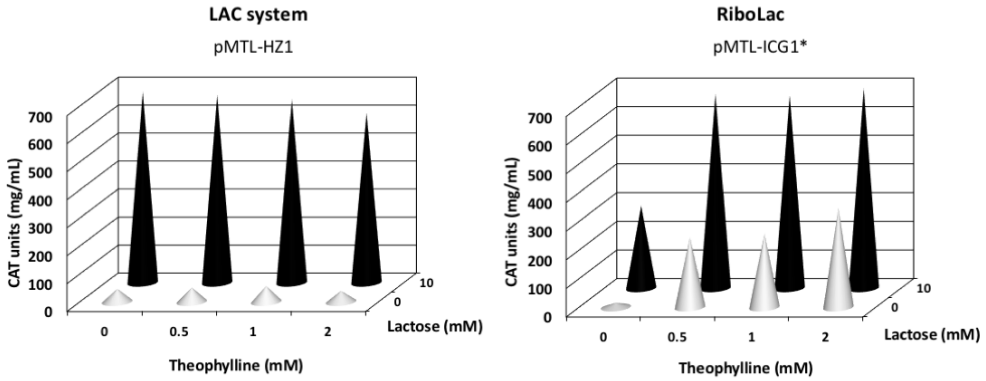


Figure 5.2 Dose-response comparison. Response of the RiboLac and LAC system to various concentrations of theophylline in the presence and absence of 10 mM lactose. Cells containing plasmids pMTL-HZ1 and pMTL-ICG1* were cultivated in TYG medium supplemented with various concentrations of the inducers theophylline and lactose when they had reached early exponential growth phase ($OD_{600} \approx 0.5$). CAT activity was determined on cell lysates from stationary cultures. Data represent the mean of three biological replicates.

Additionally, the RiboLac was also analysed for CAT expression over time in the absence of inducers and at the highest concentrations of inducers tested. In the previous theophylline-dependency assay, maximum induction was obtained with 0.5 mM theophylline and 10 mM lactose. However, in Chapter 4, maximum induction of the theophylline responsive riboswitches, without affecting cell growth, was obtained with 2 mM theophylline. As a matter of consistency, 2 mM theophylline was used to obtain maximum levels of expression in the RiboLac as well. Accordingly, CAT expression of cultures grown in the presence and absence of 2 mM theophylline and 10 mM lactose, as well as their optical density (OD_{600}), were monitored over time (Figure 5.3a and Figure 5.3b).

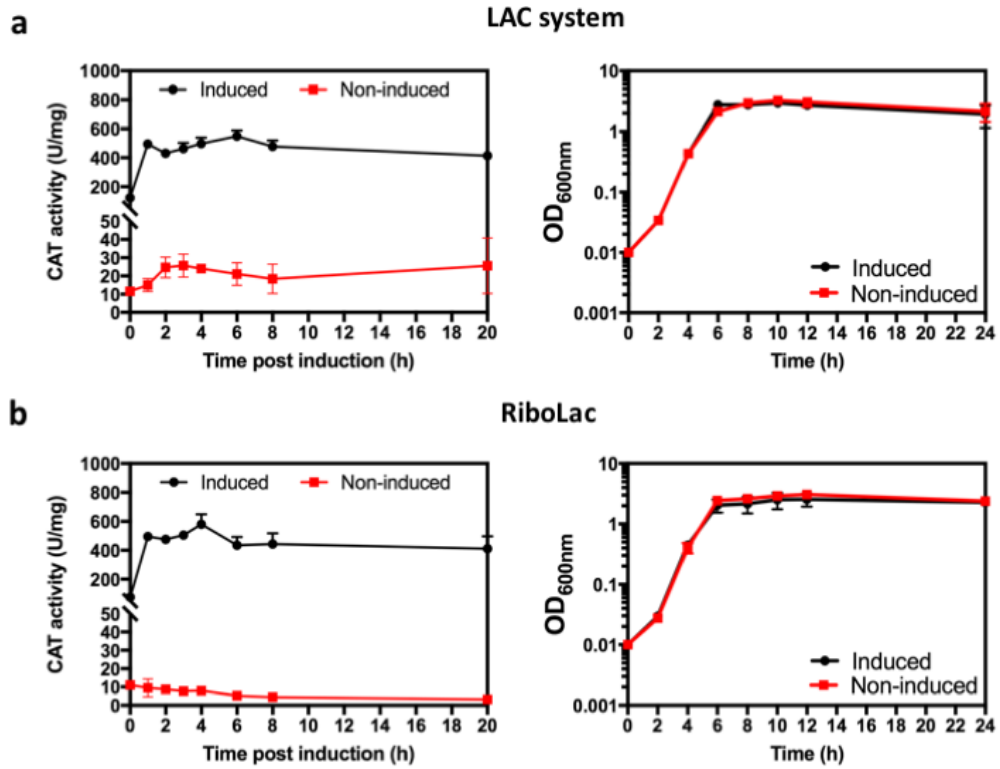


Figure 5.3 Time-course induction assays. CAT expression profiles and growth monitored over time of *C. sporogenes* harbouring **(a)** the LAC system (pMTL-HZ1) and **(b)** the RiboLac (pMTL-ICG1*) in the presence and absence of inducers. In both cases, cells were cultivated in TYG medium lacking lactose as well as cultures to which 2 mM theophylline and 10 mM lactose were added when they had reached early exponential growth phase (at 4 hours of growth, $OD_{600} \approx 0.5$). Samples at 4 hours were collected immediately after induction (time 0 post-induction). CAT activity was determined on cell lysates derived from stationary phase cultures. Error bars represent the standard deviation of three biological replicates.

In both systems, expression of *catP* above the level of the uninduced culture was detectable immediately upon addition of the inducers (at 4 hours), indicating that expression of BgaR under the control of the riboswitch occurs within a few minutes of theophylline being added. No significant differences in CAT expression were observed between the LAC system and the RiboLac in the presence of inducers. However, whilst CAT expression in the absence of inducer slightly increased with time in cultures harbouring the LAC system, from six hours onwards, it decreased in cultures harbouring the RiboLac. Maximum

expression under the control of the theophylline responsive riboswitch studied in Chapter 4 was reached between 4 and 6 hours post induction (8-10 hours of growth), with expression after 20 hours reduced by 75% and 40% in P_{fdx} and P_{fdx}*, respectively. Accordingly, reversal of background expression in the RiboLac over time might be associated with the abundance or stability of the mRNA encoding BgaR, which is strongly reduced between 4 and 6 hours after induction. No differences in growth rates were observed between the cells grown in the presence and absence of the inducers.

5.3.3 The RiboLac in the chromosome

The engineering of plasmid-free stable strains requires that the biological parts conferring the desired functionality are integrated into the host genome. To determine whether the degree of repression and dynamic range would provide suitable levels of gene expression under these conditions, the RiboLac was integrated into the clostridial host chromosome. When this study initiated, chromosomal integration into clostridial genomes was achieved via recombination-based allelic exchange[125,203,204]. Specifically, integration of a particular cargo into the same strain of *C. sporogenes* used in this study had previously been accomplished in this group via ACE, relying on the interconversion of the *pyrE* locus between a functional and mutant allele[125]. In ACE, the coupling of a plasmid-specified allele with a chromosomal allele during allelic exchange results in the formation of a new allele that has a selectable phenotype. The *pyrE* gene encodes the enzyme orotate phosphoribosyltransferase which is involved in *de novo* synthesis of the pyrimidine base, uracil[205]. Cells carrying a mutant, chromosomally located *pyrE*

allele, lacking the 3′ end of the gene, can no longer make uracil, and are unable to grow in defined media without the addition of exogenous uracil. ACE plasmids carrying a *pyrE* allele comprising all but the first 8 codons of the gene can be used to restore the mutant, chromosomal allele to WT through selection for uracil prototrophy. Moreover, cargo DNA inserted into the ACE plasmid immediately downstream of the *pyr*E allele becomes integrated into the chromosome concomitant with restoration of prototrophy[125].

To integrate RiboLac in the *pyrE* locus, the vector pMTL-JH29[125], which contains two unequally sized regions of homology to the *C. sporogenes* chromosome, was modified to include a double terminator immediately downstream of the *pyrE* gene, intended to insulate the integrated expression unit. The double terminator, composed of the *E. coli rrnB* T1 terminator and the phage T7 early transcription terminator, was synthesised based on the existing sequence in the registry of standard biological parts (BBa_B0015)[206]. The terminator *rrnM* T1 from *E. coli* had been demonstrated to be functional in *C. acetobutylicum*[207], which, combined with the phage T7 early transcription terminator, is the most frequently used double terminator from the registry of standard biological parts. The integration vector containing the double terminator was renamed pMTL-JH29T. In this vector, one of the homologous regions, corresponding to the longer 1200 bp region immediately downstream of *pyrE*, directs the first recombination event resulting in single crossover integration of the plasmid. Subsequent plasmid excision via the second, smaller region of homology, corresponding to a *pyrE* allele lacking its 5′ end, results in the integration of the double terminator concomitant with restoration of uracil prototrophy. The vector includes the restriction sites NotI and XhoI between the

double terminator and the longer homologous region so that a specific cargo can be easily inserted in the vector for its consequent integration. Accordingly, two vectors, pMTL-JH29T-L and pMTL-JH29T-R, were created by inserting the LAC system-*catP* and the RiboLac-*catP* coding sequences, respectively. These sequences were directly taken from the reporter vectors pMTL-HZ1 and pMTL-ICG1*, respectively. As such, the LAC and RiboLac coding sequences also include the native terminator downstream of *bgaR*. The designated plasmid constructs were transferred into *C. sporogenes* (in triplicates) from their respective *E. coli* donor strains. Integrants were isolated by plating cultures of the resultant transconjugants onto defined medium lacking uracil (see section 2.5.12). Integration of both the LAC system and the RiboLac reporter systems into the *pyrE* locus were confirmed via PCR, using external primers targeting the *C. sporogenes* chromosome (Figure 5.4), and subsequent DNA sequencing of the amplified DNA fragments. A second PCR product was obtained in all cases, including the WT strain, indicative of non-specific amplification using the selected pair of primers.

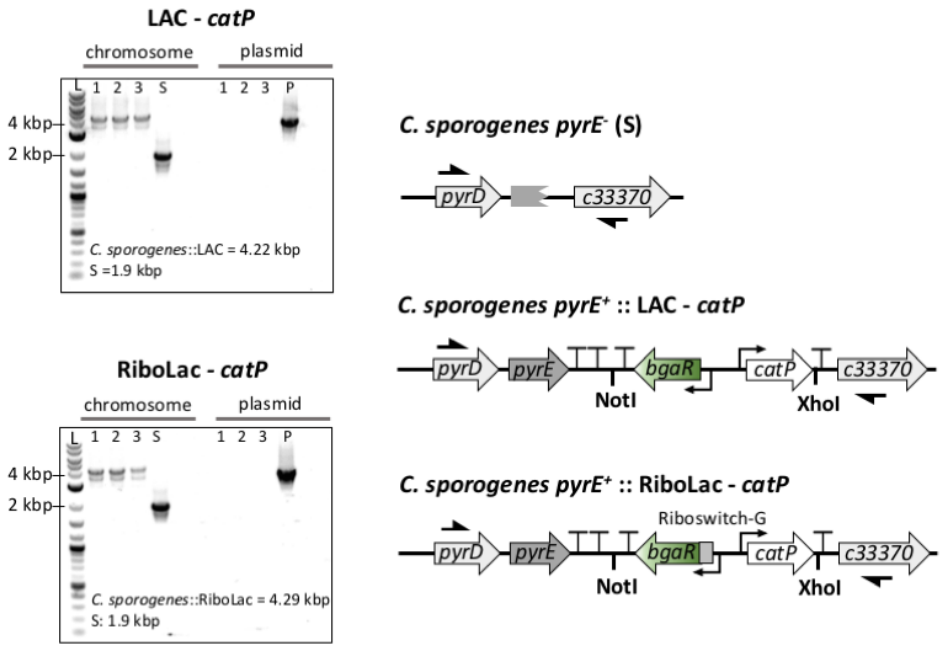


Figure 5.4 Agarose gel electrophoresis of PCR products of the *pyrE* locus following integration of the Lac-*catP* and RiboLac-*catP* sequences. PCRs were conducted using chromosomal external primers IC001-f and IC002-r. Three separate clones, derived from three independent conjugations, were examined. S lane shows PCR of insertion site in *C. sporogenes pyrE*⁻ strain using both external primers. Plasmid loss was verified in the same samples by PCR, using screening primers IC007-f and IC008-r, which align to the integration vector (see Appendix, Fig. S5). P lane shows PCR of the pMTL-JH29T-L and pMTL-JH29T-R plasmids. L is 2-log ladder.

In order to assess expression, cultures of the integrants obtained, named *C. sporogenes::*LAC and *C. sporogenes::*RiboLac, were cultivated in TYG medium and induced with only one of the inducers, theophylline (2 mM) or lactose (10 mM), or with both inducers (2 mM theophylline and 10 mM lactose) as well as in the absence of inducers. Strains harbouring the reporter plasmids pMTL-HZ1 and pMTL-ICG1* were used as positive controls as well as to determine the differences in expression encountered between the plasmid-based systems and the same systems integrated into the clostridial chromosome.

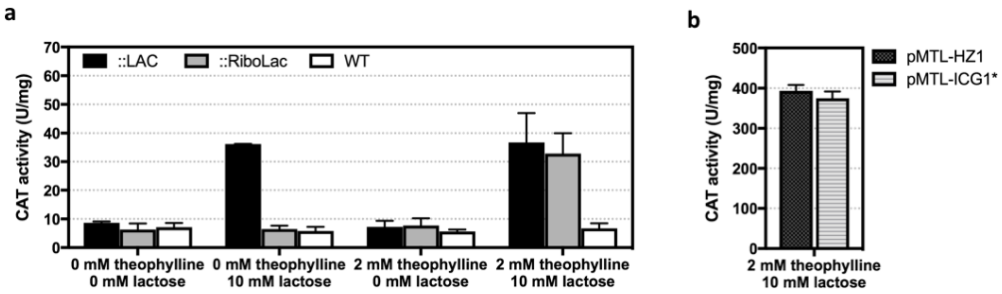


Figure 5.5 CAT assay of systems integrated into the chromosome. **(a)** CAT assay of *C. sporogenes::*LAC and *C. sporogenes::*RiboLac integrants in the presence or absence of the inducers theophylline and lactose. The WT strain was used as negative control. **(b)** CAT assay of cells harbouring the vectors pMTL-HZ1 and pMTL-ICG1*. In both cases, cells were cultivated in TYG medium lacking theophylline and lactose, as well as in the presence of either one of the inducers or in the presence of both inducers. Induction took place when cultures had reached the early exponential growth phase ($OD_{600} \approx 0.5$). CAT activity was determined on cell lysates derived from stationary phase cultures. Error bars represent the standard deviation of three biological replicates.

Expression of *catP* driven by the integrated RiboLac was only detectable in the presence of both inducers, indicating that the dual mechanism is functional also in the chromosome of *C. sporogenes* (Figure 5.5). Expression in the absence of inducers as well as in the presence of only one inducer were comparable to that of the WT strain. The level of CAT observed when *C. sporogenes*::RiboLac was induced with both, theophylline and lactose, was similar to that of *C. sporogenes::*LAC in the presence of lactose. As expected, maximum expression was markedly reduced if compared with cells harbouring the plasmids pMTL-HZ1 and pMTL-ICG1*, being reduced by approximately 10-fold when the systems were integrated into the single copy clostridial chromosome. No differences in background *catP* expression were shown between *C. sporogenes*::LAC and *C. sporogenes*::RiboLac. Considering that differences were observable when the systems were tested in a plasmid (Figure 5.2 and Figure 5.3), it might be possible that the CAT assay used in this study is

not sensitive enough to detect the low levels of expression obtained from the single copy chromosomal integration. It is expected, however, that the integrated RiboLac is likely to be providing tighter control of gene expression than the integrated native LAC system.

5.3.4 The Lactoswitch, dual control over the target gene

As mentioned in the introduction of this chapter, previously reported dual systems were created to regulate the expression of the target gene, both at the transcriptional and translational level[200,201]. Although the dynamic range was considerably improved, maximum expression was greatly reduced when the riboswitch was placed upstream of the target gene. To further demonstrate the outperformance of the RiboLac, where the transcriptional activator is regulated at the translational level, it was decided to create a system where the expression of the reporter gene, and not that of *bgaR*, was regulated at the translational level. To facilitate comparisons, the system was termed Lactoswitch. Accordingly, two theophylline responsive riboswitches, -G and -H, were placed downstream of P_{bgaL}*, generating the vectors pMTL-ICLG1* and pMTL-ICLH1*, respectively. According to previous results, riboswitches located downstream of a promoter that retains the 5′ UTR sequence (i.e., P_{fdx}*) exhibit higher levels of expression than those obtained when the same riboswitch is placed immediately downstream of the TSS (i.e., P_{fdx}). Since the aim of generating a dual system with improved regulation is to reduce background expression while maintaining a large dynamic range, it was considered that P_{bgaL}*, and not P_{bgaL}, would be better suited to the purpose.

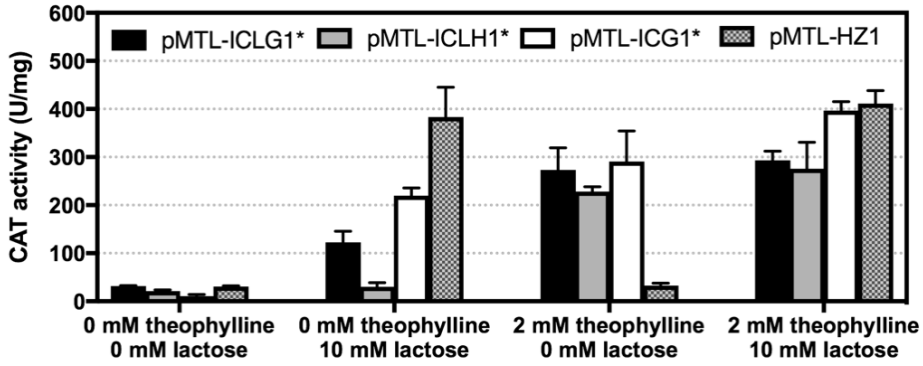


Figure 5.6 Lactoswitch evaluation. Plasmid-based CAT assay of *C. sporogenes* strains expressing *catP* under the control of the Lactoswitch (pMTL-ICLG1* if contains riboswitch-G and pMTL-ICLH1* if contains riboswitch-H). Strains harbouring the RiboLac (pMTL-ICG1*) and the native LAC system (pMTL-HZ1) were also included in the assay. Cells containing the designated plasmids were cultivated in TYG medium lacking inducers as well as cultures to which only one inducer (2 mM theophylline or 10 mM lactose) or both inducers (2 mM theophylline and 2 mM lactose) have been added. Error bars represent the standard deviation of three biological replicates.

The CAT data shown in Figure 5.6. illustrates the enhanced performance of the RiboLac over a system where the target gene is dually regulated – the Lactoswitch. In the Lactoswitch containing riboswitch-G, background expression was not reduced in the absence of inducers and expression after induction did not reach the levels of the native LAC system. The Lactoswitch containing riboswitch-H, on the other hand, exhibited reduced background expression in the absence of inducers, with expression higher than the background of the LAC system only in those cases where theophylline had been added. This indicates that riboswitch-H tightly represses P_{bgaL} in the absence of inducers, with expression being maximally compromised after induction with both theophylline and lactose. Owing to the maximum levels of expression being lower than those obtained with the native LAC system or with the RiboLac, the Lactoswitch (referring to the one containing riboswitch-H) afforded a modest

13-fold induction, which was slightly lower than the 14-fold dynamic range observed here in the LAC system. Expression changes for both remained far from the 40-fold activation obtained with the RiboLac.

Altogether, these data indicate that the dual system is functional in *C. sporogenes,* both on a plasmid and after being integrated into the host chromosome. Maximum expression levels were comparable to those obtained with the native LAC system after induction, with an improvement of the dynamic range from approximately 15-fold to approximately 40-fold. It is worth noting that the levels of induction achieved are still far below from those reported in other inducible systems. However, comparisons cannot be easily made, since in other studies the predominant reporter systems employed are based on fluorescent proteins, which cannot be used in anaerobic organisms such as *Clostridium*. The differences in induction levels might be due to the sensitivity of the assay rather than to the response of the systems themselves. A reporter with higher sensitivity than CAT which may overcome this problem has been increasingly used in clostridial organisms; this is the GUS reporter system, which utilizes ß-glucuronidase (*E. coli gusA*) determining activity via a fluorometric assay in cell lysates[108,129]. This reporter system benefits from the commercially available range of fluorimetric and spectrophotometric substrates and kits, which represents an improvement over the CAT assay[84]. Characterization of the systems developed in this study with a different and more sensitive reporter gene would support the data obtained, facilitating comparisons with other systems reported in the literature.

5.4 Key outcomes

1. Successful demonstration of the functionality of a dual system that regulates translation of a transcriptional activator, which regulates the transcription of a target gene.
2. Generation of a library of dual systems where expression of the transcriptional activator is regulated by various riboswitches placed downstream of different 5′ UTRs. The best system, termed RiboLac, improved the dynamic range from 15-fold to 40-fold, relative to the native LAC system.
3. Characterization of the dynamics and kinetics of the dual system RiboLac.
4. Demonstration of the functionality of the RiboLac integrated into the clostridial chromosome.
5. Demonstration of the enhanced performance of the RiboLac *vs* a dual system controlling expression of the target gene both at the transcriptional and translational level, the Lactoswitch.

Chapter 6
/ Development of a conditionally sporulating *C. sporogenes* strain

¨Any sufficient advanced technology is indistinguishable from magic¨

– Arthur C. Clarke

6.1 Background

6.1.1 A conditionally sporulating strain

In the main introduction of this thesis, it was hypothesized that a conditionally sporulating strain could be generated by tuning the sporulation cascade, so that the pathway is activated/completed in the presence or absence of a ligand. To do it, it was necessary to have systems capable of finely regulating such pathway. In chapters 3, 4 and 5, a wide range of inducible systems were generated, exhibiting various levels of repression, activation and dynamic range. As such, it was expected that (at least) one of the characterized systems allowed the construction of a conditionally sporulating strain. With such range of inducible gene expression systems for *Clostridium* to hand, the next step was to identify which gene/s should be regulated.

6.1.2 Controlling spo0A to regulate sporulation

As discussed in the Introduction of this thesis, one common aspect of both *Bacillus* and *Clostridium*, is that activation of spore formation is governed by the key master regulator Spo0A[63]. Although the full initiation pathway has been identified only in *Bacillus* species, it seems that phosphorylation of this master regulator by one or more histidine kinases leads to the initiation of spore formation in both genera[66,208]. In *Clostridium*, such phosphorylation seems to involve a simple two-component system(s), in contrast to the complex phosphorelay that regulates Spo0A in *B. subtilis*[209]. In *Bacillus*, phosphorylation of Spo0A is further regulated via a positive feedback loop in which the regulatory protein activates the transcription of the sigma factor σ^H, which, in turn, leads to an augmented transcription of *spo0A*[208]. Once activated by

phosphorylation, Spo0A binds to a specific DNA motif, termed the 0A box, and initiates the sporulation process by turning on and off the transcription of spore and vegetative cell-related genes, respectively.

In *Bacillus*, not only is phosphorylation necessary to initiate the process of sporulation, but a gradual increase in the level of Spo0A is also required for its successful activation[68]. This is supported by the fact that many genes in the Spo0A regulon respond differentially to high and low levels of Spo0A, with some genes repressed or activated at a low dose of the regulator whereas others require a certain threshold to be turned on and off[210,211]. In *B. subtillis*, for example, the Spo0A regulon comprises approximately 120 genes[212]; genes that are involved in spore formation require higher doses of Spo0A than genes that do play a direct role in sporulation[210].

The mechanisms underlying the Spo0A regulon in *Clostridium* have yet to be elucidated. Several lines of evidence suggest a more ancestral mechanism of sporulation in *Clostridium* if compared to that of *B. subtilis*. These include a less stringent temporal segregation of sigma factor activities[213]; the lack of phosphorelay, and; a smaller core set of sporulation genes under the control of cell type-specific sigma factors[213,214]. It seems very plausible, however, that a transcriptional regulator that "makes" such an important decision – to sporulate or not – is highly regulated in *Clostridium* as well, and that a certain threshold of Spo0A might be required to initiate spore formation. In this sense, the few publications regarding *spo0A* expression profiles in clostridial organisms are subject to some contradictions. Dürre and coworkers showed that in contrast to *B. subtilis*, *spo0A* is constitutively expressed in *C. acetobutylicum* ATCC 824, and that expression remains constant throughout the growth cycle[215]. However,

three transcription start points were identified upstream of *spo0A*, from which two σ^H-dependant promoters were deduced. In addition, no significant temporal differences have been observed in the expression pattern of Spo0A in *C. perfringens*[216]. On the other hand, the overexpression of *spo0A* in *C. acetobutylicum* ATCC 824 resulted in a bimodal pattern of *spo0A* transcription, similar to that of *B. subtilis*[217]. Variable Spo0A expression has also been found during the growth cycle of *C. botulinum* ATCC 3502[196] and some clinically relevant strains of *C. difficile*[218]. In these cases, a massive decrease in the levels of *spo0A* transcripts occurred when cells entered stationary phase. Overall, these data suggest that *spo0A* could represent a suitable candidate to conditionally control sporulation. If a certain amount of Spo0A is required to initiate sporulation, an inducible system could mimic the levels of Spo0A required to be "above" and "below" the threshold from which sporulation is triggered by repressing or inducing the expression of *spo0A*. On the other hand, if *spo0A* is expressed in a constitutive manner, its absence and presence would impose an external synthetic control over the sporulation process. Additionally, controlling the sporulation process via Spo0A would allow regulation of spore formation before cells commit to sporulation[219], without compromising cell and/or spore viability. Such would be the case if the gene to be regulated in an inducible manner is required at a point after which the course of sporulation can no longer be reversed, and the physiological changes that affect cell survival have already taken place.

While this work was being carried out, a study describing the inducible expression of *spo0A* in *C. difficile* was published[220]. There, the inducible *spo0A* strain generated was able to form spores only in the presence of the inducer by

placing *spo0A* under the control of a tetracycline responsive system. This showed that a conditionally sporulating strain can be created by controlling the expression of the master regulator Spo0A. It was also demonstrated that the amount of *spo0A* transcript was directly related to the amount of spore forming cells within the population, but evidence that a threshold level of Spo0A is required for sporulation was not presented. Nevertheless, high doses of the inducer anhydrotetracycline, a tetracycline-analogue, were necessary to produce the same amount of spores as the WT strain. The levels required, however, have significant inhibitory effects on cell growth[104].

6.1.3 Other targets to generate conditionally sporulating strains

At least 60 genes within the *B. subtilis* genome are apparently essential for sporulation[221], which could all be potential targets to generate a conditionally sporulating strain. However, understanding the genetic and molecular basis of the *Clostridium* sporulation program, and specifically that of *C. sporogenes*, seemed to be crucial to correctly choose the "best" target gene to generate a strain with conditional control of spore formation. Several methods for direct gene inactivation are now available[125,222,223], allowing reverse genetic studies, in which the role of a gene with a putative function could have been experimentally determined. Additionally, the generation of random mutant libraries using transposon mutagenesis and high-throughput sequencing methods[224] could have served to identify genes involved in the sporulation program of *C. sporogenes*. Nevertheless, this was impractical in the time available, and was also the focus of other members of this research group.

Although little was known about the sporulation program of *C. sporogenes* by the time this work was being carried out, one gene, *spoIIR*, had already been identified by members within the group as being essential for sporulation (data not shown). For this reason, and because this gene is involved at the onset of sporulation (stages I-II), before cells commit to sporulate, it was selected as another potential candidate to regulate the sporulation program in *C. sporogenes*. The transcription of *spoIIR* is initiated in the pre-spore exclusively by RNA polymerase bound to σ^F soon after the septum is formed and it encodes a secretory protein that leads to the activation of the transcription program directed by σ^E in the mother cell[63,225]. Interestingly, and at least in *Bacillus*, *spoIIR* is subject to "positional regulation", meaning that its position within the chromosome is important for its expression and function[62].

6.2 Chapter overview

In this section, the inducible systems described in previous chapters are utilized to generate conditionally sporulating strains by controlling the expression of the master regulator Spo0A. After demonstrating that plasmid-based conditionally sporulating strains could be created, stable strains with inducible control of spore formation integrated into the bacterial chromosome were generated. Specifically, by addition of the inducers lactose and theophylline, the strain *C. sporogenes*::RiboLac-*spo0A* can be artificially induced to sporulate, with a sporulation capacity similar to that of the WT strain. This strain is completely asporogeneous in the absence of inducers. Additionally, a similar strategy was applied to a second target gene, *spoIIR*; however, inducible systems regulating *spoIIR* expression were directly

evaluated in the chromosome. The reasoning behind this approach is further discussed in the relevant section.

Insights into some aspects of the sporulation pathway in *C. sporogenes* gained by this study are also discussed in this chapter.

6.3 Results and Discussion

6.3.1 Generation of an asporogeneous host to test conditional sporulation systems

As discussed in the Introduction, *spo0A* represents one suitable target for the creation of a conditionally sporulating strain. As such, a strategy wherein the inducible systems from Chapters 3, 4 and 5 are used to regulate the expression of *spo0A* was considered. Since the systems exhibited various dynamic ranges, it was expected that one of them would result in a strain incapable of sporulating in the absence of induction but able to restore spore formation – ideally to WT levels – in the presence of inducer/s.

In order to generate a conditionally sporulating strain via tightly regulating the expression of the *spo0A* gene, a strain deficient in this gene had to be generated. Deletion of the *spo0A* gene (*Clspo_c18650*) was achieved via creating an in-frame deletion by *pyrE*-based allelic exchange[203]. As with the integrations by *pyrE*-based ACE carried out in Chapter 5, *pyrE*-based allelic exchange also requires a *pyrE*-deficient background. With the PyrE enzyme truncated, the organism is uracil auxotrophic and resistant to the compound 5-FOA. The selection of mutants is achieved by using a heterologous *pyrE* gene as a negative counter-selection marker in the presence of 5-FOA. Additionally, since the in-frame deletion is achieved in a *pyrE*-deficient background, complemented strains can be easily obtained using ACE whilst the *pyrE* allele is corrected.

The isolation of the single cross-over plasmid integrants is facilitated by "pseudo-suicide" vectors[203]. These are plasmids that carry a selection marker (i.e., *catP*, encoding resistance to thiamphenicol) which are sufficiently

defective in its replication so that cells carrying the non-integrated plasmid exhibit a growth disadvantage relative to cells in which the vector, containing the gene that confers resistance to the antibiotic (i.e., *catP*), has been integrated. As such, clones with the plasmid integrated grow faster on agar medium supplemented with thiamphenicol than cells carrying *catP* on a non-integrated, replication defective plasmid because if integrated growth is not limited by the rate at which the plasmid is segregated among the progeny. A vector derived from the construct pMTL-CH14, which had been previously generated by a member of this group, was used to delete the *spo0A* gene from the *C. sporogenes* *pyrE*$^{-}$ strain. The plasmid pMTL-CH14 incorporates a heterologous, functional *pyrE* gene from *C. acetobutylicum* and the *C. difficile* pCD6 replicon. Since the origin of replication from the *Bacillus* plasmid pIM13[226] was shown to be more appropriate for the isolation of faster-growing single-colonies in the same strain of *C. sporogenes*[125], the pCD6 replicon was replaced by the pIM13 replicon in pMTL-CH14, generating the deletion vector pMTL-ICD. The *spo0A* knockout cluster, comprising two equal 1000 bp homologous regions upstream and downstream of *spo0A*, respectively, was generated by SOEing PCR and inserted into the deletion vector. The deletion vector, termed pMTL-ICD-*spo0A*, was transferred into *C. sporogenes* (in triplicates) from its respective *E. coli* donor by conjugative plasmid transfer; transconjugant cells were screened for plasmid integration by colony PCR (section 2.5.11) (Figure 6.1).

Successful single-cross mutants derived from three independent conjugations – all obtained by homologous recombination through the left homologous regions (LHR) – were streaked onto TYG medium supplemented with 5-FOA; deletion of the *spo0A* gene was confirmed via colony PCR from

5-FOA-resistant colonies, using primers aligning outside of the crossover region (Figure 6.2), and subsequent Sanger sequencing. The mutant strain, named *C. sporogenes* Δ*spo0A pyrE*$^{-}$, was subject to complementation and *pyrE* restoration for its appropriate characterization.

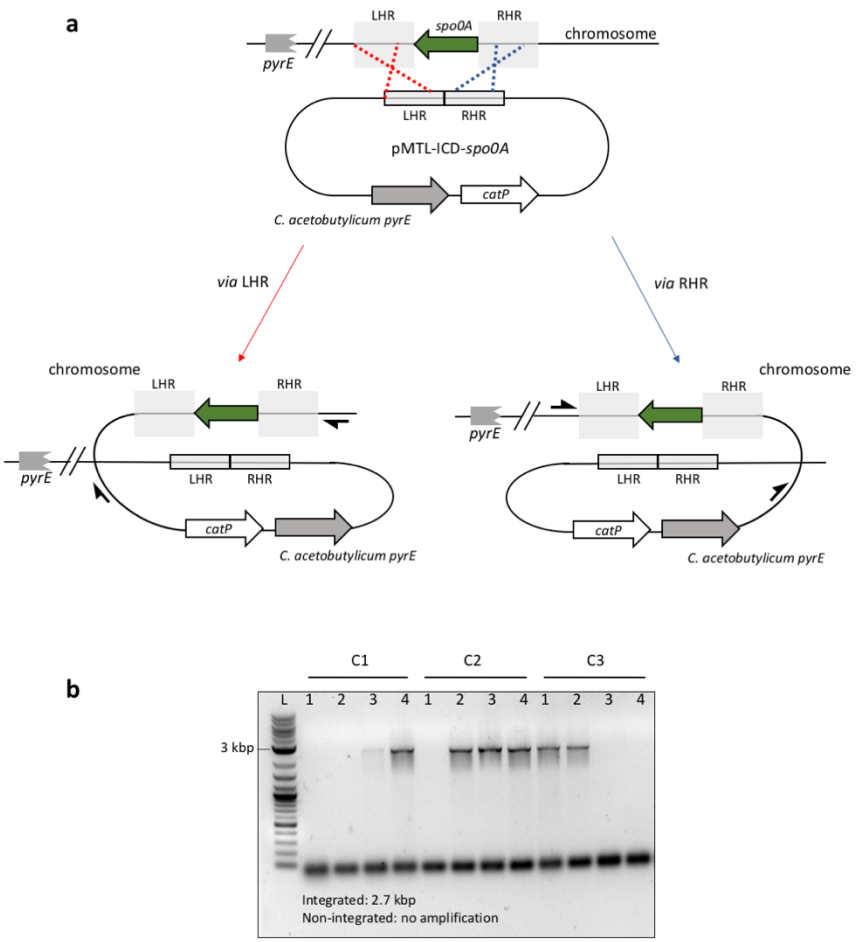


Figure 6.1 Allelic exchange single crossovers **(a)** Schematic illustration of single-cross events via left (LHR) and right homologous regions (RHR). **(b)** Agarose gel electrophoresis of PCR products of single-crossover *spo0A* mutants. Colonies from three independent conjugations (C1, C2 and C3) were analysed by colony PCR using the pairs of primers IC003-f/IC006-r and IC005-f/IC004-r, to detect plasmid integration within the chromosome of *C. sporogenes* through the LHR and RHR, respectively. The pairs of primers are specific combinations of one primer aligning to the chromosome and a second primer aligning to the deletion vector (see Appendix, Fig S4). L is 2-log ladder. The gel only shows PCR products from amplifications via the LHR, since no plasmid integrations were detected through the RHR.

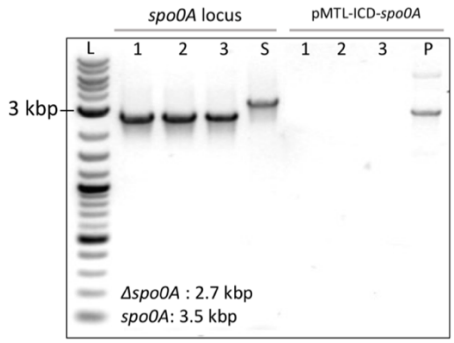


Figure 6.2 Allelic exchange double crossovers**.** Agarose gel electrophoresis of PCR products of the *spo0A* locus following deletion of the *spo0A* gene from the *C. sporogenes pyrE*⁻ strain. PCRs were conducted using chromosomal external primers IC003-f and IC004-r. Three separate clones, derived from three independent single-cross mutants were examined. Loss of the plasmid was verified in the same samples by colony PCR, using screening primers IC005-f/IC006-r. L is 2-log ladder. S is *C. sporogenes pyrE*⁻ strain. P lane shows PCR of the pMTL-ICD-*spo0A* plasmid**.**

6.3.2 spo0A complementation and pyrE repair

In order to confirm that the phenotype observed in the mutant strain is directly attributable to the deletion of *spo0A*, a *spo0A* complemented strain was generated via ACE, incorporating *spo0A* under the control of its own promoter downstream of the repaired *pyrE* locus. The integration plasmid pMTL-JH29-*spo0A*, carrying *spo0A* flanked by its native promoter and terminator, was conjugated into *C. sporogenes* Δ*spo0A pyrE*⁻ and integrants were isolated by plating cultures of the resultant transconjugants onto defined medium lacking uracil (see section 2.5.12). The complemented strain was named *C. sporogenes* Δ*spo0A*^{COMP}. Additionally, in order to fairly compare the *spo0A* mutant with the complemented strain, which contains a functional *pyrE* gene after being complemented with *spo0A* in the *pyrE* locus, a strain lacking *spo0A* but with a functional *pyrE* gene was also created. Accordingly, the "empty" pMTL-JH29

vector (with no integration cargo) was used in a similar fashion to restore the full-length of the *pyrE* gene, generating *C. sporogenes* Δ*spo0A*.

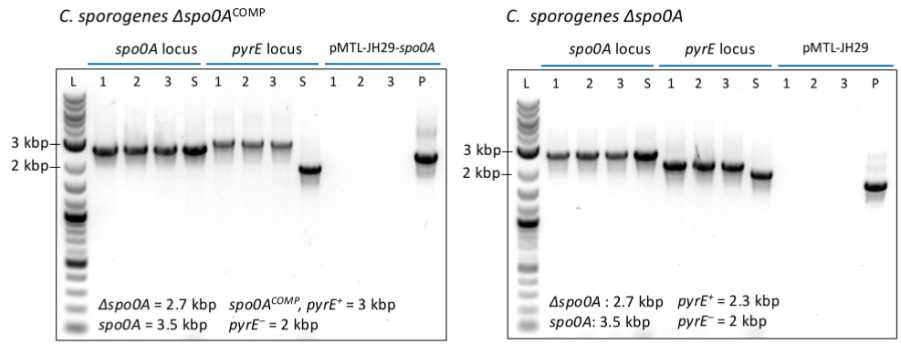


Figure 6.3 ACE double crossovers**.** Agarose gel electrophoresis of PCR products of the *spo0A* locus, the insertion site in the *pyrE* locus and the integration plasmids pMTL-JH29-*spo0A* and pMTL-JH29 for *spo0A* complementation and repair of *pyrE*, respectively. Three independent clones, derived from three independent conjugations, were examined. PCRs were conducted using chromosomal external primers IC003-f/IC004-r, IC001-f/IC002-r for and IC007-f/IC008-r for the *spo0A* locus, the *pyrE* locus and the integration plasmids, respectively. PCR products were further verified by DNA sequencing. L is 2-log ladder; S is *C. sporogenes pyrE*$^{-}$ parent strain; P is pMTL-JH29-*spo0A* or pMTL-JH29 vectors for *spo0A* complementation and *pyrE* repair, respectively.

6.3.3 Characterization of *C. sporogenes* Δ*spo0A pyrE*$^{-}$

C. sporogenes Δ*spo0A pyrE*$^{-}$*,* the strain which would serve as a host to create a conditionally sporulating strain, was characterized for both its sporulation capacity and growth profile. Accordingly, the *spo0A* complemented strain, *C. sporogenes* Δ*spo0A*^{COMP} was used to evaluate if the deletion of *spo0A* is the cause of the expected asporogeneous phenotype. The *spo0A* mutant with regained uracil autotrophy, *C. sporogenes* Δ*spo0A*, was included in the assay to account for any effect *pyrE* could have on spore formation and growth. *C. sporogenes WT* and *C. sporogenes pyrE*$^{-}$ were used as positive controls. Detection of heat resistant spores follows a simple procedure of heat-shocking cell suspensions at 80°C for 20 minutes; CFUs are obtained from plating serial-

dilutions of such suspensions on appropriate nutrient agar plates (see section 2.7.3). Accordingly, synchronized cultures of the designated strains were cultivated in TYG medium and total and heat-resistant CFU were enumerated at 24 hour intervals for five days (Figure 6.4a). To assess growth of the different strains, the optical density (OD_{600}) of the same cultures was monitored during the first 24 hours of cultivation (Figure 6.4b).

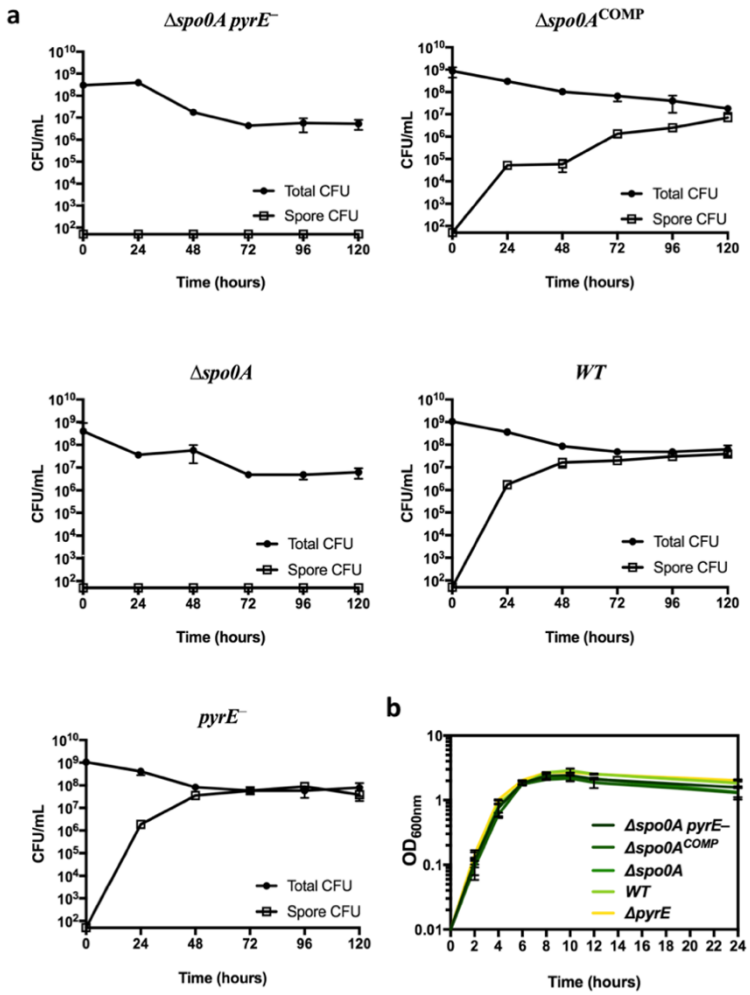


Figure 6.4 **(a)** Sporulation and **(b)** growth profiles of *C. sporogenes* Δ*spo0A pyrE*$^{-}$, *C. sporogenes* Δ*spo0A*^{COMP}, *C. sporogenes* Δ*spo0A*, *C. sporogenes WT* and *C. sporogenes pyrE*$^{-}$. After culture synchronization (see section 2.7.3), samples were removed at the indicated times points, serially diluted in and spotted onto TYG agar. For spore counting, samples were heat-treated for 20 minutes at 80°C prior to serial dilution. The optical density of the cultures (OD_{600}) was monitored over the first 24 hours of growth. Error bars represent the standard deviation of three biological replicates from two independent assays. Y axis set at the limit of detection (50 spores/mL).

As anticipated, none of the *spo0A* mutants, *pyrE*$^{-}$ and *pyrE*-repaired, were able to form spores over the course of the experiment; in both cases, total CFU was significantly lower than in the strains containing *spo0A* (approximately 10-fold) (unpaired two-tailed Student's *t-test*; *$p \leq 0.0332$ in both cases), indicative of the role Spo0A plays in cell viability. Moreover, a slight decrease in cell density was observed after 24 hours of growth in strains lacking *spo0A*. This would be in line with the notion that a mechanism (i.e., phosphorelay in *B. subtilis*), at the centre of which is Spo0A, integrates nutritional, cell density and other signals to initiate the sporulation process [227]. The complemented strain, *C. sporogenes* Δ*spo0A*^{COMP}, was able to sporulate; however, its sporulation capacity was reduced by approximately 10-fold at all time points when compared to the *spo0A*-positive strains. Spore formation after 24 hours in *spo0A*-positive strains was comparable to that of the complemented strain only after 72 hours of growth, indicating a delay in spore formation. Since sporulation might be triggered by a certain threshold of Spo0A-P[210], it is possible that *spo0A* expression in the *pyrE* locus, despite being driven by its own promoter, is lower than in its native site. This would have limited the amount of Spo0A that could have been phosphorylated, and hence, restricted the entry into the sporulation program. Conversely, the reduced sporulation capacity of the complemented strain could be the result of the accelerated accumulation of Spo0A in the cell, which has been demonstrated to have detrimental effects in *B. subtilis* sporulation[211]. In line with this hypothesis is the fact that *spo0A* is natively located near the terminus of the chromosomal replication, whereas *pyrE* – where *spo0A* was complemented – is placed near the origin of replication; in bacterial chromosomes, highly expressed genes are generally located close to

the origin[228]. The optical density of *C. sporogenes* Δ*spo0A*^{COMP} cultures was slightly lower than that of the WT and *pyrE*$^{-}$ strains after 24 hours of growth, reinforcing the link between *spo0A* levels, cell density and sporulation.

In order to assess if the lack of heat-resistant CFUs was due to the inability of the *spo0A*-deficient strain to form spores or rather to other underlying processes (i.e., inability to germinate), *C. sporogenes* Δ*spo0A* cultures were examined with a phase-contrast microscope (see section 2.7.5). No spores were observed in three biological replicates of *C. sporogenes* Δ*spo0A* cultures under the microscope (Figure 6.5), which, regardless of any limit of detection associated with this technique, strongly suggests that the lack of CFU after heat-treatment is linked to a deficiency in spore formation.

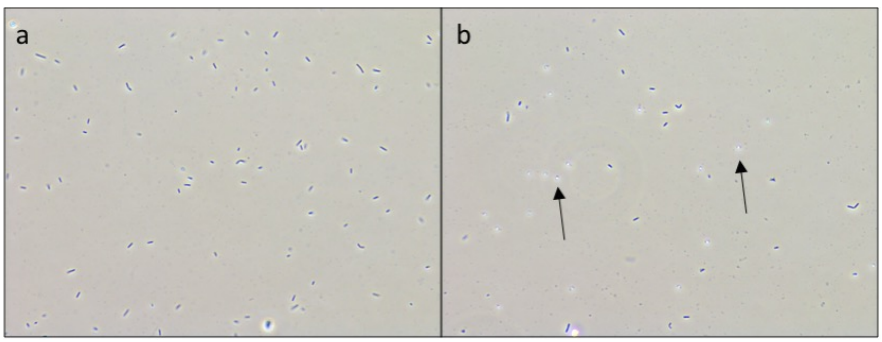


Figure 6.5 Microscopic view of **(a)** *C. sporogenes* Δ*spo0A* and **(b)** *C. sporogenes* WT cultures after 120 hours. Spores appear phase bright (arrows) and vegetative cells and debris as purple phase dark elements.

6.3.4 Plasmid-based conditionally sporulating strains

Having generated the strain of *C. sporogenes* required to test if the inducible expression of *spo0A* would result in the conditional control of spore formation, it was decided, initially, to evaluate the sporulation capacity of *C. sporogenes* Δ*spo0A* strains harbouring plasmids where *spo0A* expression was

regulated by the inducible systems developed in the previous chapters; these are: theophylline responsive riboswitches, the LAC system from *C. perfringens* strain 13 and the RiboLac.

To facilitate comparisons of the results obtained in the sporulation assays and to estimate how the reported strengths and dynamic ranges of the inducible systems relate to spore formation when controlling *spo0A* expression, CAT assays to monitor inducible promoter activities using all the systems generated in the previous chapters were carried out (Figure 6.6a). Considering that the incorporation of any riboswitch in the 5′ UTR downstream of a promoter led to a strong reduction of gene expression (Figure 4.6, Chapter 4), and that expression would be further reduced when generating stable strains with *spo0A* incorporated into the single copy bacterial chromosome, the very strong constitutive promoter of the *Clostridium acetobutylicum araE* gene ($P_{araE,}$ associated with the protein coding gene Ca_1339) was included in the repertoire of theophylline responsive riboswitches. Since gene expression levels are higher when the native 5′ UTR is retained upstream of the riboswitch, and due to the easy construction of riboswitches located downstream of a SD-deficient promoter, only P_{araE}* (see Appendix, Fig. S7), and not P_{araE}, was considered as a candidate to generate a conditionally sporulating strain. The reporter vectors containing P_{araE}* upstream of riboswitches -E, -G and -H were created similarly to the constructs described in Chapter 4, generating vectors pMTL-IC321-E, -G and -H (see section 2.6.4) and strains harbouring these plasmids were included in the CAT assay.

For assessing sporulation, the backbone pMTL82251[126] used in the previous chapters was also considered as a backbone for all the *spo0A*

expressing plasmids. All vectors were assembled in an identical fashion to the CAT reporter plasmids, but with *spo0A* as the target gene instead of *catP* (section 2.6.4). The series of *spo0A* expressing plasmids, termed pMTL-ICSPO followed by the name of the regulatory system (e.g., pMTL-ICSPO-P_{fdx4}-E for the vector expressing *spo0A* under the control of P_{fdx4} located upstream of riboswitch-E), were conjugated into *C. sporogenes* Δ*spo0A* from their respective *E. coli* donors. Cultures of the resultant transconjugants were analysed for heat-resistant spore production after 120 hours of nutrient starvation in liquid culture, in the presence and absence of the inducer/s. This is normally considered as the time after which the sporulation process in most of the cells of a *C. sporogenes* culture has been completed[229]. *C. sporogenes WT* and *C. sporogenes* Δ*spo0A* were used as positive and negative controls, respectively.

Results from Figure 6.6b demonstrate that sporulation can be regulated in an inducible manner in *C. sporogenes*. However, only strains harbouring plasmids with *spo0A* driven by P_{fdx4} and P_{fdx4}*, both located upstream of riboswitch-E, produced no spores in the absence of theophylline. Nevertheless, the spore count in the presence of the inducer was reduced by approximately 10^3-fold and 10^5-fold in P_{fdx4}-E and P_{fdx4}*-E, respectively, when compared to the WT strain. As P_{fdx4}-E and P_{fdx4}*-E are among the systems that exhibited lower levels of *catP* expression in the uninduced state (see Figure 6.6a), these results indicate that a tight control of *spo0A* expression is required to abolish spore formation. Sporulation under the control of any other riboswitch downstream of P_{fdx4} and P_{fdx4}* resulted in spore counts close to the detection limit. This agreed with the low expression levels of CAT observed when these promoters, combined with any riboswitch, regulate *catP* expression. Figure 6.7 shows the

plates onto which heat-treated cultures of the conditionally sporulating strain carrying P_{fdx4}*-E were plated, and how growth compares to the WT strain both, in the presence and absence of induction. Strains harbouring any other system were capable of sporulating in the absence of inducer, with the amount of spores higher in those systems exhibiting higher CAT background expression levels. Strains containing *spo0A* regulated by any riboswitch located downstream of P_{fdx} or P_{fdx}*, as well as in the cases of the LAC system and the RiboLac, resulted in spore counts after induction comparable to those of the WT strain. Considering that the expression of *catP* under the control of riboswitches located downstream of P_{fdx} in the presence of the inducer is, for example, 10-fold lower than that under the control of the induced LAC system, these data indicate that at a certain level of Spo0A sporulation efficiency reaches saturation. Higher expression of *spo0A* (i.e., P_{fdx}*-, P_{araE}*-, LAC and RiboLac) did not further increase spore counts after 120 hours, although more experiments would be needed in order to further examine the sporulation profile at early data points.

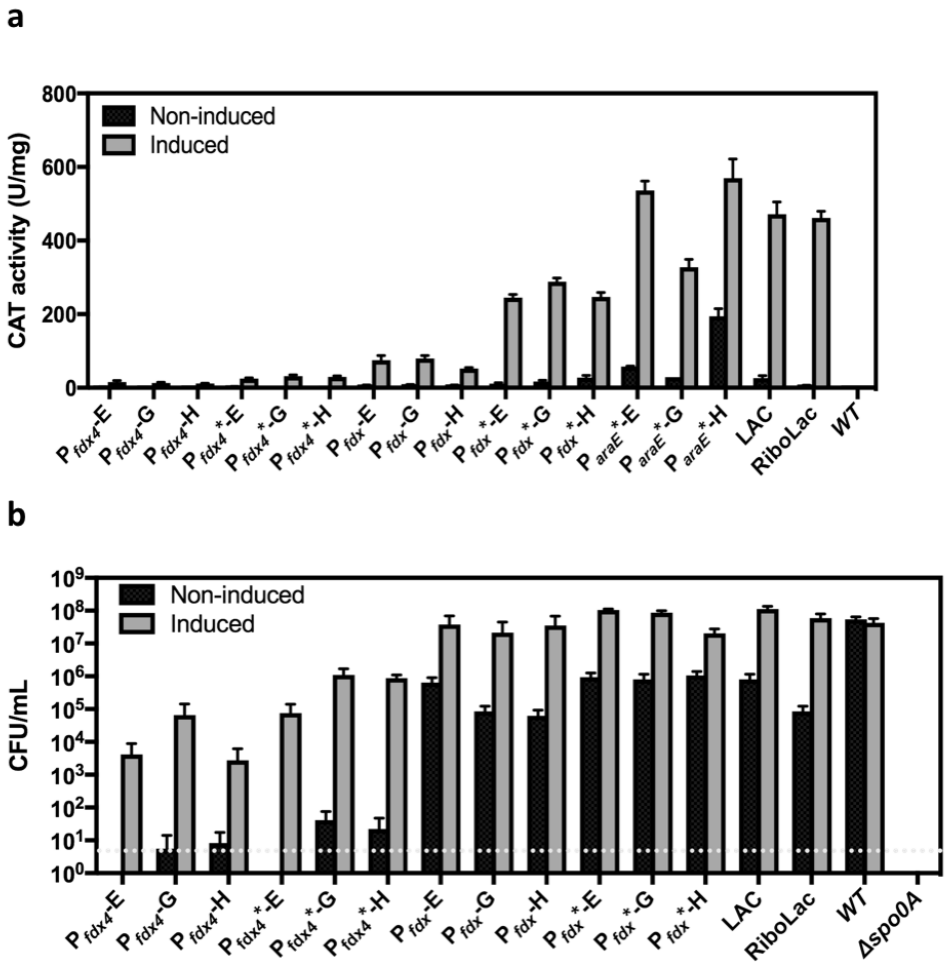


Figure 6.6 Assay of plasmid-based conditionally sporulating strains **(a)** Plasmid-based CAT assays of *C. sporogenes* strains expressing *catP* under the control of various inducible systems generated in this study. Cell cultivations and assay were performed as described in previous chapters. Error bars represent the standard deviation of three biological replicates. **(b)** Sporulation profile of *C. sporogenes* Δ*spo0A* strains harbouring the different pMTL-ICSPO plasmids. Cultures were grown synchronized (see section 2.7.3) and used to inoculate the starting culture of the assay. Samples were removed at time 0 to verify the lack of spores in the starting cultures (data not shown). At early exponential phase ($OD_{600} \approx 0.5$) cultures were split into two equal volumes and the following inducers added to one of them: 2 mM theophylline (if regulated by riboswitches), 10 mM lactose (if regulated by the LAC system) and 2 mM theophylline and 10 mM lactose (if regulated by the RiboLac). After 120 hours, cell suspensions were heat-treated, serially diluted and spotted onto TYG agar. Colonies were enumerated after 48 hours of incubation under anaerobic conditions. Error bars represent the standard deviation of three biological replicates from two independent assays. Dotted line represents the limit of detection (50 spores/mL).

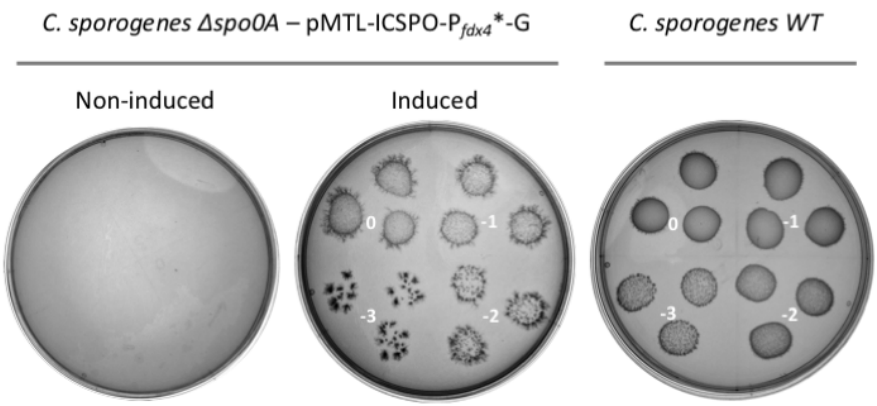


Figure 6.7 Plates of heat-treated plasmid-based conditionally sporulating strain. Agar plates of serially diluted heat-treated induced and non-induced cultures of the plasmid-based conditionally sporulating strain *C. sporogenes* harbouring the *spo0A* expressing plasmid with *spo0A* driven by P_{fdx4}*-G and *C. sporogenes WT*. Numbers indicate 10^x dilution. Plates were examined after 48 hours of incubation under anaerobic conditions.

6.3.5 Chromosomally stable conditionally sporulating strains

Progression towards a stable strain capable of conditionally forming spores required the integration of *spo0A,* driven by an inducible system, into the chromosome of the clostridial host. Given that the level of repression and dynamic range of the different inducible systems regulating *spo0A* once integrated into the chromosome was unknown, it was decided to integrate several systems so that a wide range of expression profiles was covered. These included the two systems that were able to repress spore formation in the plasmid-based sporulating strains (i.e., P_{fdx4}-E and P_{fdx4}*-E); the promoters P_{fdx}, P_{fdx}* and P_{araE}* located upstream of the riboswitches -E and -G; the LAC system, and; the RiboLac. Although riboswitches -E and -G resulted in similar levels of plasmid-based spore counts in the presence and absence of the inducer, it was expected that expression from the single copy chromosome would result

in observable differences, which could translate into a different profile of spore formation.

All inducible system-*spo0A* sequences were ligated into pMTL-JH29T and integrated into the genome of *C. sporogenes* at the *pyrE* locus, as previously described (see section 2.5.12), and the sporulation capacity of the resultant mutants analysed. Accordingly, synchronized cultures of the designated strains were cultivated in TGY medium, in the presence and absence of induction, and total and heat-resistant CFU were enumerated after 120 hours of growth (Figure 6.8).

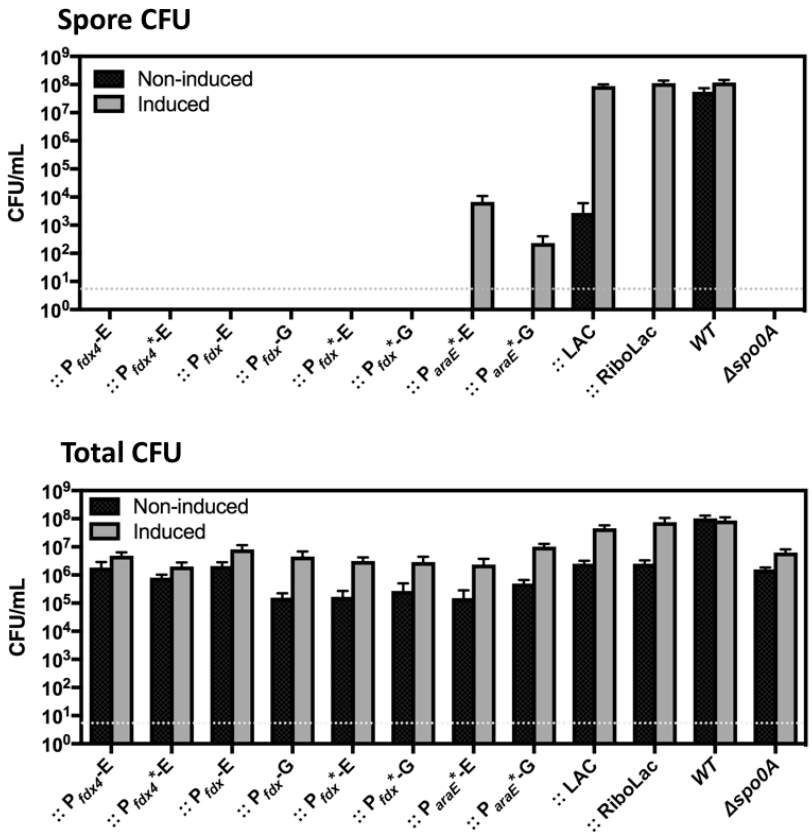


Figure 6.8 Sporulation capacity of stable conditionally sporulating strains. Error bars represent the standard deviation of three biological replicates from two independent assays. Dotted line represents the limit of detection (50 spores/mL).

Analysis of the sporulation capacity of strains containing *spo0A* driven by an inducible system integrated into the chromosome revealed one conditionally sporulating strain with a tight regulation of spore formation, resulting in no spores in the absence of the inducer and spore counts comparable to those of the WT strain after induction. This system was the RiboLac. Surprisingly, only the strong P_{araE}* promoter, out of all the integrated theophylline responsive riboswitches, was able to express *spo0A*, after induction, to an extent that cells were capable of activating the sporulation program; however, sporulation capacity was reduced by 10^4-fold and 10^6-fold relative to the WT strain, when regulated by riboswitches -E and -G, respectively. The rest of the strains with *spo0A* controlled by a riboswitch did not sporulate, either in the absence or in the presence of induction. Since these systems exhibited various levels of *catP* expression and induction profiles in previous assays (Figure 6.6), these data support the idea that a certain threshold of Spo0A might be needed to activate the sporulation programme in *C. sporogenes*, and that this level is not reached when some of these systems are integrated into the chromosome. Both, the LAC system and the RiboLac restored sporulation capacity of the respective strains to WT levels after induction; however, background *spo0A* expression from the LAC system was enough to trigger the formation of spores, indicating the superior ability of the dual system RiboLac to tightly regulate gene expression (Figure 6.9 and Figure 6.8). In all cases, total CFU was higher in the presence of induction, reaching comparable counts to WT in ::LAC-*spo0A* and ::RiboLac-*spo0A* strains.

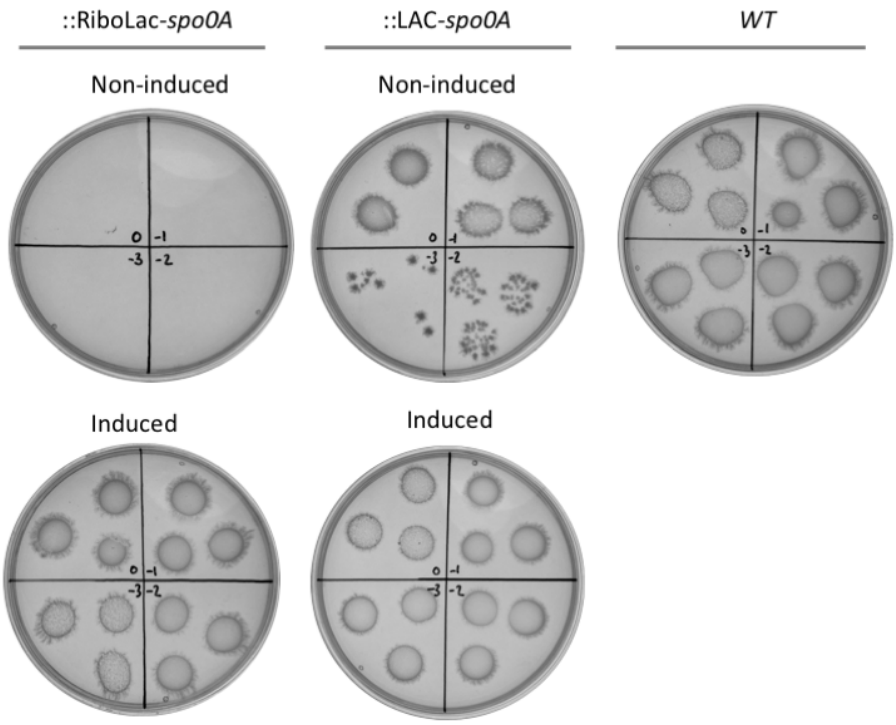


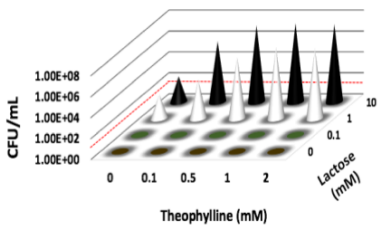
Figure 6.9 Plates of heat-treated stable conditionally sporulating strains. Agar plates of serially diluted heat-treated cultures of *C. sporogenes*::RiboLac-*spo0A*, *C. sporogenes::*LAC-*spo0A* and *C. sporogenes WT* in the presence and absence of induction. Numbers indicate 10^{x} dilutions. Plates were examined after 48 hours of incubation under anaerobic conditions.

6.3.6 Characterization of the stable conditionally sporulating C. sporogenes::RiboLac-spo0A strain

The data obtained had demonstrated that the RiboLac is sufficiently tightly regulated to generate, after 120 hours, an asporogeneous strain when grown in the absence of inducers, and capable of forming spores in a WT fashion when cultivated at maximum induction (2 mM theophylline and 10 mM lactose). It was, therefore, of interest to investigate how changes in Spo0A levels, obtained via induction with various concentrations of inducers, would affect the overall sporulation efficiency (Figure 6.10). As expected, a minimum concentration of inducers was required to activate the sporulation program.

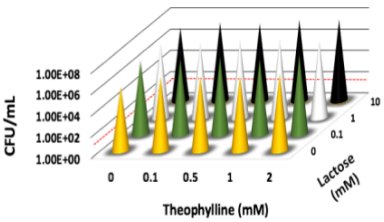
Above that level, increasing concentrations of theophylline and lactose showed a dose-dependent increase in heat-resistant CFUs, reaching the maximum levels (>10^7 CFU/mL) from 0.5 mM theophylline and 10 mM lactose. In agreement with previous experiments, cell viability was slightly reduced in the absence of, or at very low concentrations of inducers, and increased with higher concentrations and therefore with increased expressions of *spo0A*.

Spore CFU/mL



Lactose (mM) \ Theophylline (mM)	0	0.1	0.5	1	2
0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
0.1	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
1	$2.06 \times 10^2 \pm 1.62 \times 10^2$	$5.17 \times 10^3 \pm 3.10 \times 10^3$	$8.00 \times 10^5 \pm 4.37 \times 10^5$	$6.79 \times 10^6 \pm 4.08 \times 10^6$	$5.22 \times 10^6 \pm 4.19 \times 10^6$
10	$2.83 \times 10^2 \pm 4.14 \times 10^1$	$6.63 \times 10^5 \pm 5.63 \times 10^5$	$2.18 \times 10^7 \pm 4.69 \times 10^6$	$3.56 \times 10^7 \pm 5.09 \times 10^6$	$2.83 \times 10^7 \pm 1.42 \times 10^7$

Total CFU/mL



Lactose (mM) \ Theophylline (mM)	0	0.1	0.5	1	2
0	$1.06 \times 10^6 \pm 1.40 \times 10^5$	$8.78 \times 10^6 \pm 2.44 \times 10^6$	$1.03 \times 10^7 \pm 1.74 \times 10^6$	$1.11 \times 10^7 \pm 2.72 \times 10^6$	$1.58 \times 10^7 \pm 6.01 \times 10^5$
0.1	$8.44 \times 10^6 \pm 5.09 \times 10^5$	$1.12 \times 10^7 \pm 2.77 \times 10^6$	$1.76 \times 10^7 \pm 6.27 \times 10^6$	$1.07 \times 10^7 \pm 8.21 \times 10^6$	$2.71 \times 10^7 \pm 3.33 \times 10^6$
1	$9.22 \times 10^6 \pm 2.43 \times 10^6$	$1.09 \times 10^7 \pm 6.81 \times 10^6$	$1.99 \times 10^7 \pm 4.95 \times 10^6$	$2.35 \times 10^7 \pm 3.53 \times 10^6$	$2.12 \times 10^7 \pm 7.52 \times 10^5$
10	$8.56 \times 10^6 \pm 6.74 \times 10^5$	$2.78 \times 10^7 \pm 1.80 \times 10^7$	$2.16 \times 10^7 \pm 1.83 \times 10^6$	$4.72 \times 10^7 \pm 1.17 \times 10^7$	$5.22 \times 10^7 \pm 6.74 \times 10^7$

Figure 6.10 Dose-dependent increase in sporulation efficiency in *C. sporogenes*::RiboLac-*spo0A*, represented as expression matrices and numerically. Starting cultures were split into 20 equal volumes and each of them cultured under different inducer concentrations; total and heat-resistant CFU were enumerated after 120 hours of growth. The mean values and standard deviations represent the CFUs/mL of biological triplicates in two independent assays. Dotted line in the matrices represents the limit of detection (50 spores/mL).

Once the dose-dependent sporulation profile of *C. sporogenes* ::RiboLac-*spo0A* was determined, it was decided to investigate if the induction

of *spo0A* had any effect on the dynamics of the sporulation process. Accordingly, spore production was monitored over time with three different concentrations of inducers, chosen on the basis of the range of spore counts obtained in the dose-dependency assay: 0 mM theophylline and 0 mM lactose (no sporulation); 0.1 mM theophylline and 10 mM lactose (medium induction), and; 2 mM theophylline and 10 mM lactose (maximum induction) (Figure 6.11a). In agreement with previous assays, no spores could be detected based on heat-resistant CFU counts in the absence of inducers. Maximum spore titres were observed at 120 and 72 hours, corresponding to $6.7 \times 10^5 \pm 3.17 \times 10^5$ CFU/mL and $4.8 \times 10^7 \pm 9.5 \times 10^6$ CFU/mL in cultures supplemented to generate medium and maximum induction, respectively. Taking into account the total CFU observed at time 0, which is similar in all conditions, these values correspond to a sporulation efficiency of 0.19% (0.1 mM theophylline and 10 mM lactose) and 13.9% (2 mM theophylline and 10 mM lactose). This indicates a reduced activation of the sporulation process, independent of cell viability (see total CFU at 24 and 48 hours in both cases), which correlates with the level of expression of *spo0A* and which had been previously observed when analyzing the sporulation efficiency of the complemented strain, *C. sporogenes* Δ*spo0A*^{COMP} (Figure 6.4). The reduction in spore formation between WT and the conditionally sporulating strain was only significant during the first 48 hours of cultivation (Figure 6.11c). After 120 hours, total CFU was comparable to the WT strain only if the ::RiboLac-*spo0A* strain had been exposed to maximum induction. The inducing agents neither had an effect on the final spore counts observed in WT cultures nor changed the dynamics on the sporulation process (Figure 6.11b). The optical density of the cultures (OD_{600}), which had been

monitored over the first 24 hours of growth (Figure 6.11d), showed a slight decrease in cell density during the first 12 hours; from this time onwards, it reached optical density values comparable to the WT strain and remained constant over the remainder of the experiment. An explanation of why this is the case can only be speculative without further information about the *spo0A* regulon in *C. sporogenes*; it may be possible that both, high and low levels of *spo0A* affect growth. In *B. subtillis*, the phosphorelay regulates the gradual build-up of Spo0A[210], implying differential gene regulation when Spo0A is present at low and high levels; differences in the gradual accumulation of Spo0A could be the reason for the small differences observed both in spore formation and growth. The reduction of the sporulation capacity of *B. subtilis* upon artificial induction of Spo0A has been proposed to be related to the premature expression of DivVA, a protein involved in chromosome segregation that is downregulated by Spo0A in the later stages of spore formation[211]. The results of this study, however, contrast with studies performed in *C. difficile*, where accelerated accumulation of Spo0A resulted in more rapid sporulation[220], indicating that a gradual build-up of the master regulator is not important to the sporulation process itself. Based on these studies, it is hypothesized that a gradual build-up of Spo0A seems to be important in *C. sporogenes.*

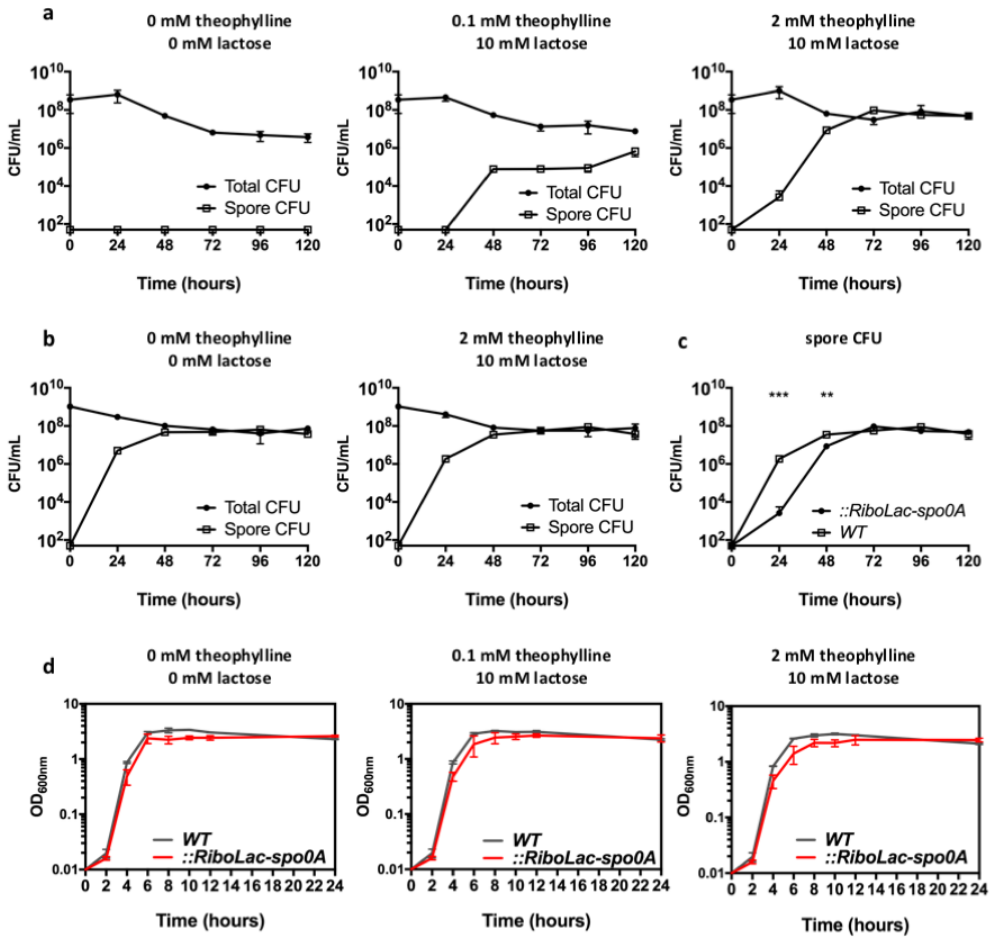


Figure 6.11 Sporulation profiles of **(a)** *C. sporogenes*::RiboLac-*spo0A* and **(b)** *C. sporogenes WT* with various concentrations of inducers. **(c)** Comparison of the sporulation profiles of both strains. Asterisks indicate statistically significant induction values for *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001 (unpaired two-tailed Student's *t*-test). Y axis set at the limit of detection (50 spores/mL). **(d)** Optical density (OD_{600}) of *C. sporogenes*::RiboLac-*spo0A* and *C. sporogenes* WT in response to various concentrations of theophylline and lactose. Error bars represent the standard deviation of three biological replicates from two independent assays.

6.3.7 Other targets to conditionally sporulate – spoIIR

The second target gene to generate a conditionally sporulating strain considered in this study was *spoIIR*. As with *spo0A*, a strain deficient in *spoIIR* (*Clspo_c01510*) had to be generated. However, this work was performed after the development of a novel CRISPR tool for the genus *Clostridium*, which is the

focus of the next chapter of this thesis. The generation of the *spoIIR* mutant, hence, was achieved via this method, whose specific details are discussed in Chapter 7. Due to the easy complementation of mutants in a *pyrE*$^{-}$ background (via ACE), *spoIIR* mutants using the CRISPR method were generated in both *C. sporogenes* WT and *C. sporogenes pyrE*$^{-}$.

The *spoIIR* complemented strain was generated via ACE, as previously described, incorporating *spoIIR* under the control of its own promoter downstream of the repaired *pyrE* locus. The integration plasmid pMTL-JH29-*spoIIR*, carrying *spoIIR* flanked by its native promoter and terminator, was used to generate *C. sporogenes* Δ*spoIIR*^{COMP}.

As expected, the *C. sporogenes spoIIR* mutant strain showed an asporogeneous phenotype (Figure 6.12). The complemented strain, *C. sporogenes* Δ*spoIIR*^{COMP}, however, formed spores but its sporulation capacity was reduced to 0.1% of the level obtained when *spoIIR* was located at its original location (WT). In line with the work done in *B. subtilis*, whose sporulation capacity was impaired when *spoIIR* was moved away from the origin of replication, this data indicate a possible "positional regulation" of *spoIIR* also occurs in *C. sporogenes*[62]. The *spoIIR* locus in *C. sporogenes* is located at 346° near the origin of replication (0/360°). The *pyrE* locus, where the complementation took place, is located at 42°. In *B. subtilis*, however, the movement of *spoIIR* to regions in the proximity of the origin of replication of the chromosome (i.e., from 324° to 28°) had little effect on sporulation, but sporulation was reduced by 90% if *spoIIR* was moved to 190°.

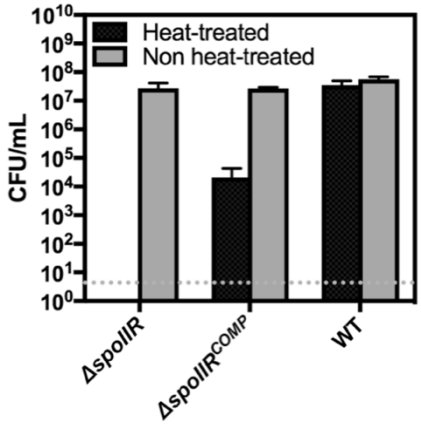


Figure 6.12 Sporulation capacity of *C. sporogenes* Δ*spoIIR and C. sporogenes* Δ*spoIIR*^{COMP} at 120 hours. Error bars represent the standard deviation of three biological replicates from two independent assays. Dotted line represents the limit of detection (50 spores/mL).

Considering the possibility of the "positional regulation" of *spoIIR* in *C. sporogenes*, it was decided to evaluate if a *spoIIR*-inducible strain would sporulate in a conditional fashion by integrating the system directly into the clostridial chromosome, and specifically, in the *spoIIR* locus. Due to *spoIIR* transcription is only σ^F-dependent, the *spoIIR* promoter – or the elements responsive to the specific sigma factor σ^F – should remain in the regulatory system employed to generate the conditionally sporulating strain. This implies that systems such as the LAC system or the RiboLac were not suitable for this purpose unless they were further engineered (i.e., by modifying P_{bgaL} to contain σ^F-responsive elements). Riboswitches, on the other hand, could be easily included into the *spoIIR* promoter, especially if the promoter retains the native 5′ UTR (P_{spoIIR}*, see Appendix, Fig. S7). Riboswitch G, which had demonstrated the highest dynamic range when expressing *catP* (Chapter 4), was chosen to regulate the expression of *spoIIR*.

Integration of the regulatory system was achieved in a similar fashion to the deletion of *spo0A* but introducing a cargo; this is, via "insertion" by *pyrE*-based allelic exchange. *C. sporogenes* Δ*spoIIR pyrE*$^{-}$, and not the *spoIIR*-positive strain, was selected as the target host to facilitate screening, since the size of the inserted riboswitch is only 74 nucleotides. Accordingly, the sequence P_{spoIIR}*-G-*spoIIR* was ligated into the pMTL-ICD deletion vector (now use for integration) and integrated into the *spoIIR* locus of *C. sporogenes* generating the strain *C. sporogenes*::P_{spoIIR}-*G-spoIIR*.

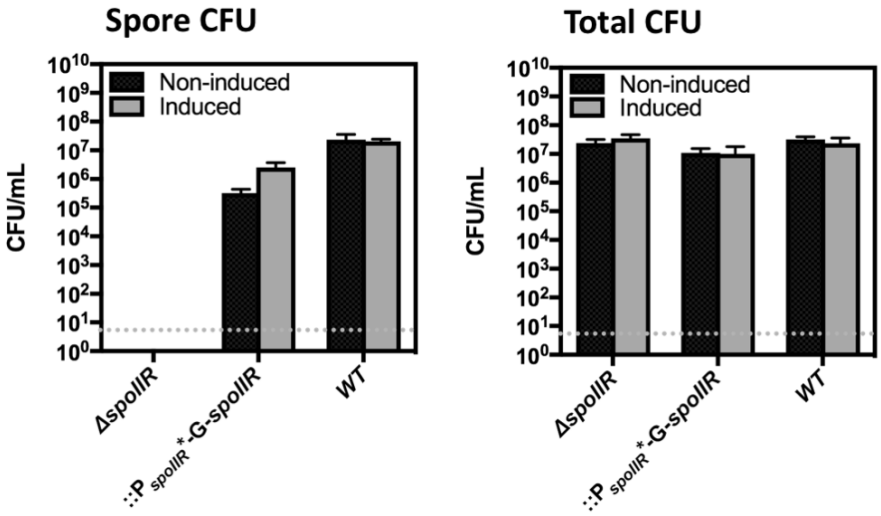


Figure 6.13 Sporulation capacity of *C. sporogenes*::P_{spoIIR}-*G-spoIIR*. Cultures were grown to late exponential phase (OD_{600} ≈ 2; 6 hours of growth) and sub-cultured 1:100 in fresh TYG medium. Samples were removed at time 0 to verify the lack of spores in the starting cultures (data not shown). At early exponential phase (OD_{600} ≈ 0.5) cultures were split into two equal volumes, with one of them induced with 2 mM theophylline. Total and heat-resistant CFU were enumerated after 120 hours of growth. *C. sporogenes* Δ*spoIIR* and *C. sporogenes* WT were used as negative and positive control, respectively. Error bars represent the standard deviation of three biological replicates from two independent assays. Dotted line represents the limit of detection (50 spores/mL).

Regulation of *spoIIR* with a riboswitch resulted in a strain capable of sporulating in the absence of theophylline (Figure 6.13). In addition, its

sporulation capacity in the presence of induction was reduced by approximately 10-fold, relative to WT. This is perhaps indicative of the ultrasensitive regulation of *spoIIR* in the cell, with slight variations in its level of expression markedly affecting spore production. It could be possible as well that the combination Riboswitch-G-*spoIIR*, although inducible (differences between the induced and uninduced states were observed), does not repress *spoIIR* translation very efficiently. However, since sporulation was not fully restored after induction, independently of a tight or leaky repression, *spoIIR* regulation is less permissive than that of *spo0A*, in the sense of regulating spore formation. Considering that expression of *spoIIR* is probably highly reduced when it is placed downstream of a riboswitch, and that sporulation occurs even in the absence of the inducer, it seems very likely that the amount of SpoIIR needed to successfully sporulate is independent of a certain threshold, as it seems to be the case with Spo0A. However, further experiments will be necessary to confirm such assumptions. One simple solution could be to determine transcript levels, via RT-qPCR, at different concentrations of inducers and determine how they correlate with spore formation.

6.4 Key outcomes

1. Generation and characterization of *C. sporogenes* Δ*spo0A* and *C. sporogenes* Δ*spoIIR* strains, useful backgrounds for generating conditionally sporulating strains.
2. Demonstration that induction of Spo0A drives sporulation in *C. sporogenes* via generating plasmid-based conditionally sporulating strains.
3. Generation of stable conditionally sporulating strains by means of regulating the master regulator Spo0A. Specifically, the strain *C. sporogenes*::RiboLac-*spo0A* can be artificially induced to sporulate in a WT manner by addition of the inducers lactose and theophylline. This strain is completely asporogeneous in the absence of inducers. The inducible regulation of Spo0A with systems that had exhibited different levels of CAT expression indicate that a certain threshold of Spo0A might be needed to activate the sporulation program in *C. sporogenes*.
4. Characterization of *C. sporogenes*::RiboLac-*spo0A* through the analysis of the dose-dependent sporulation profile and the dynamics of spore formation over time.
5. Generation of a strain where the gene encoding the stage II sporulation protein, SpoIIR, was missing and its subsequent complementation revealed that *spoIIR* might have a "positional regulation" in *C. sporogenes*.

Chapter 7
/ A universal Cas9-based editing tool for the genus *Clostridium*: RiboCas

Nothing in life is to be feared, it is only to be understood.
Now is time to understand more, so that we may fear less.

—Maria Skłodowska-Curie

7.1 Background

One important limitation encountered in the application of synthetic biology to *Clostridium spp.* is the lack of rapid and reliable methods for genome editing. Traditionally, chromosomal modifications have been primarily achieved via intron retargeting using ClosTron[222,230] or via recombination-based allelic exchange[125,203,204]. Both approaches are far from ideal. The ClosTron, for example, is an insertional mutagen based on a group II intron, the insertion of which into the target gene can lead to polar effects on downstream genes. Procedures based on allelic exchange, such as those described in Chapters 5 and 6, allow more precise modification of the genome, but are reliant on the availability of counter selection markers (i.e., *pyrE* or *codA*) and invariably involve several time-consuming steps.

Even though CRISPR/Cas9-based systems have been previously used in several *Clostridium* species, including *C. pasteurianum*[231,232], *C. acetobutylicum*[231,233,234], *C. beijerinkii*[234–236] and *C. difficile*[237,238], this technology is still hindered by low transformation efficiencies, possibly related to the large size of the plasmids used, and the inherent lethality of Cas9 activity. The use of a Cas9 nickase (nCas9), a Cas9 mutant that introduces a breakage on only one strand of the chromosome[239], has been also demonstrated in *Clostridium spp.*[234,240]; however, because nCas9 has a less powerful selection capacity than Cas9, the isolation of mutants becomes more time consuming, requiring several passages onto fresh medium[240].

Inducible control of CRISPR-associated components offers tailored solutions to increase the efficiency of Cas9-based genome editing. In particular, the transient expression of the CRISPR components could circumvent the

problem associated with the low transformation efficiency of *cas9*-carrying plasmids, providing, additionally, the time required for homologous recombination events to occur before the nuclease mediates the site specific DSB. The few inducible systems employed in bacteria for CRISPR applications, all regulating *cas9* or sgRNA expression at the level of transcription, seem to function more efficiently than those that constitutively express the heterologous Cas9 nuclease[233]. However, these strategies are limited by the cargo capacity of the plasmid, which already harbours the 4.1-kbp *cas9* gene. One way of circumventing this problem is to use a two-plasmid system[233]; a disadvantage of this strategy, however, is that two transformation steps are required to temporarily regulate Cas9 or sgRNA expression. Inducible strategies that control Cas9 expression at the level of translation have not been, as yet, considered, and this approach will be the focus of this section of the thesis.

7.2 Chapter overview

Synthetic biology applications in *Clostridium spp.* are still hampered by the lack of reliable methods for genome editing. In this section, the development of a novel CRISPR system, named RiboCas, which regulates the expression of Cas9 at the level of translation, is discussed. RiboCas allows very low basal expression of Cas9 in the absence of inducer and an adequate level of gene expression after induction. As such, the tool results in both high transformation efficiency and selection capacity, overcoming the previously reported obstacles associated with CRISPR-based technologies in clostridial organisms. The universality of RiboCas was further demonstrated by obtaining chromosomal modifications in four different clostridial species.

7.3 Results and Discussion

7.3.1 Riboswitches to create an efficient Cas9-based tool

With a tightly inducible gene expression system for *Clostridium* available, the possibility of generating an efficient clostridial genome editing tool based on CRISPR/Cas9 was explored. In this sense, it was hypothesized that a riboswitch-based CRISPR/Cas9 editing tool (RiboCas) would circumvent the obstacles previously seen for several reasons. Firstly, the theophylline responsive riboswitch can be designed to have the desired regulatory window, whereby there should be very low basal expression of Cas9 in the absence of inducer, allowing homologous recombination to occur before Cas9 mediates the site-specific DSB, and after induction, to be expressed at a level that is sufficient for effective genome DNA cleavage but which minimizes toxicity that can be associated with Cas9 overexpression. Secondly, riboswitches are smaller structures compared to systems that require a TF; in a riboswitch-based system, 84 nucleotides are enough to tightly control the expression of Cas9, minimizing the size of the plasmid employed and enabling higher transformation efficiencies. Thirdly, because repression occurs at the level of translation, the RiboCas system allows the use of high-copy origins of replication, avoiding the undesired effects linked to read-through transcription on the plasmid backbone and facilitating the processes of cloning, screening and sequencing. Finally, and due to the aforementioned characteristics, RiboCas permits the confinement of all the essential components of a functional genome editing tool to the same plasmid. Figure 7.1 illustrates the assembly of the RiboCas vector series, generically named pRXCasN vectors, where X and N refer to the riboswitch and

the promoter driving the sgRNA respectively, and the general RiboCas-based editing process.

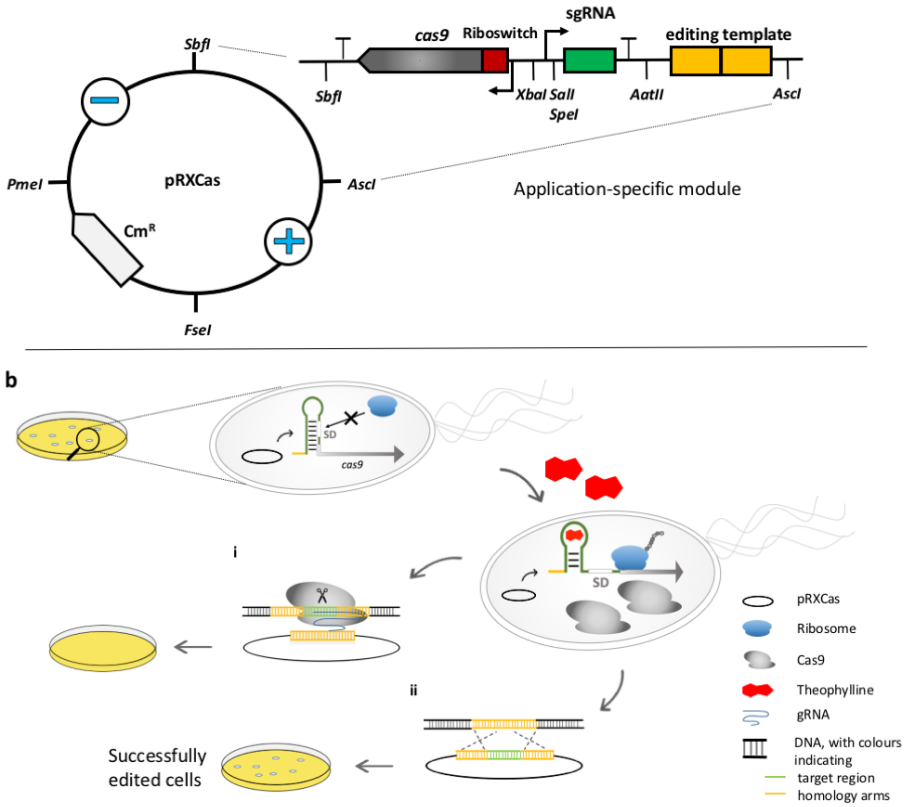


Figure 7.1 RiboCas **(a)** Schematic illustration of the RiboCas vector containing unique restriction sites used for modular assembly. (-) is the Gram-negative origin of replication that allows replication of the shuttle plasmid in *E. coli*. (+) is the Gram-positive replicon. The application-specific module harbours the components of the editing tool, including the gene encoding Cas9, the sgRNA and the homologous DNA template needed for homologous recombination. The terminators CD0164 (derived from *C. difficile*) and T_{fdx} (derived from the *C. pasteurianum* ferredoxin gene) are placed downstream of *cas9* and the sgRNA, respectively. The system was designed to be compatible with our previous pMTL80000 vector series enabling rapid exchange of selection markers and origins of replication[109]. **(b)** Illustration of the RiboCas-mediated genome editing. Transformed cells survive on selective medium in the absence of theophylline due to the tight repression exerted by the riboswitch, which impedes the translation of the nuclease Cas9. After induction, the translated Cas9 forms a complex with the sgRNA (Cas9-sgRNA) inducing DSB on the target DNA. (i) The Cas9-sgRNA complex kills the unedited cells. (ii) Cells are able to survive only if homologous recombination occurs between the gene-targeting plasmid and the genome.

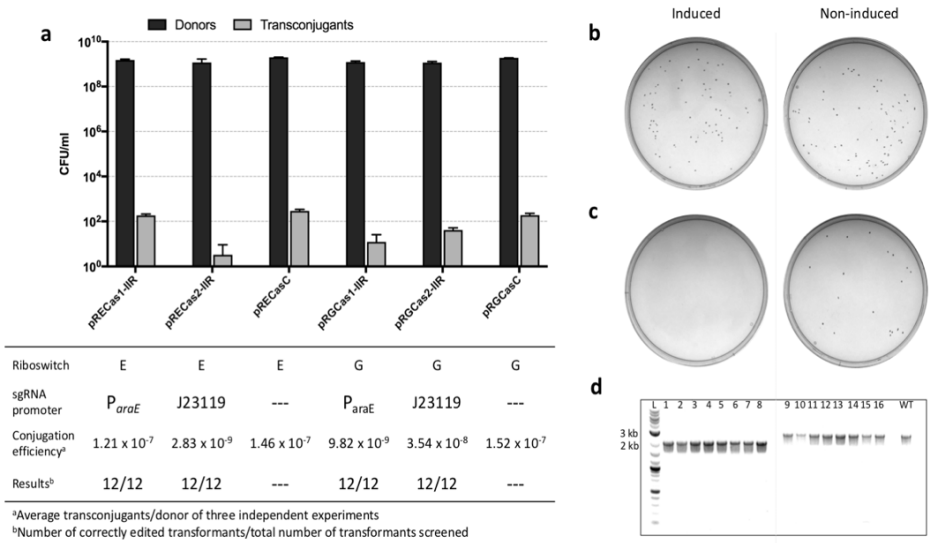
7.3.2 Construction of RiboCas

In order to have an appropriate control over *cas9* expression, both in the uninduced state and after induction, we selected two riboswitches, riboswitches -E and -G, from the riboswitch library generated in Chapter 4, both located downstream of the promoter P_{fdx}. Riboswitch-E had shown the tightest repression in the non-induced state amongst all tested riboswitches (see Figure 4.4, Chapter 4); on the other hand, riboswitch-G showed higher background levels but increased expression when the inducer was added. Two promoters, the constitutive promoter of the *C. acetobutylicum araE* gene (P_{araE},) and the synthetic promoter J23119, were chosen to bring about the expression of the sgRNA. Both had been previously used for CRISPR strategies with *Clostridium* species[127,234]. In all cases, the editing template comprised two homology arms composed of the regions of the chromosome needed to replace the target region. In line with the wider aim of this project to develop a conditionally sporulating strain, it was decided to target the *spoIIR* gene (*Clspo_c01510,* encoding the stage II sporulation protein R) as this gene had been previously disrupted in *C. sporogenes* via insertional mutagenesis using ClosTron (data not shown). The DNA modules encoding either riboswitch-E or -G, *cas9*, the sgRNA downstream of either P_{araE} or J23113 and the DNA editing template were inserted into the pCB102-based modular vector pMTL83151[109], yielding the following pRXCasN vector series: pRECas1-IIR, pRECas2-IIR, pRGCas1-IIR and pRGCas2-IIR (see section 2.6.5). The same vectors lacking the editing homology arms were assembled to determine the killing capacity of Cas9, and were named pRECas1G-IIR, pRECas2G-IIR, pRGCas1G-IIR and pRGCas2G-IIR. Two vectors, pRECasC and pRGCasC, lacking both the homology arms

and sgRNA, were also created as controls for conjugation efficiency comparisons.

7.3.3 Analysis of RiboCas in C. sporogenes

Transfer of the resulting vectors into *C. sporogenes* was performed via conjugative plasmid transfer from an *E. coli* donor, plating the conjugated cultures onto media supplemented with thiamphenicol in the presence and absence of 5 mM theophylline. As expected, the different plasmid constructs were transferred with different conjugation efficiencies (Figure 7.2a). The combination of riboswitch-E and the promoter P_{araE} upstream of the sgRNA (pRECas1) provided the highest conjugation efficiency (1.21×10^{-7} CFU/donor; CFU: colony forming units), comparable to the 1.46×10^{-7} CFU/donor obtained with pRECasC (non-targeting Cas9). Conjugative transfer of the control plasmids, pRECasC and pRGCasC, resulted in similar colony counts on plates with and without theophylline (Figure 7.2b), indicating a very low off-target activity in the absence of the targeting module. In all cases the conjugation efficiency was lower if riboswitch-G was used to regulate the expression of Cas9, which implies a higher background expression level in the absence of the inducer and agrees with the CAT results obtained in Chapter 4. No colonies were observed on plates supplemented with theophylline if the sgRNA was expressed, whether or not the editing template was included, indicating highly efficient Cas9-mediating killing of host cells (Figure 7.2c). These results also suggest that it could be possible to select mutant cells just after transformation if the homologous recombination event had occurred during the recovery time.



Riboswitch	E	E	E	G	G	G
sgRNA promoter	P_{araE}	J23119	---	P_{araE}	J23119	---
Conjugation efficiency[a]	1.21×10^{-7}	2.83×10^{-9}	1.46×10^{-7}	9.82×10^{-9}	3.54×10^{-8}	1.52×10^{-7}
Results[b]	12/12	12/12	---	12/12	12/12	---

[a]Average transconjugants/donor of three independent experiments
[b]Number of correctly edited transformants/total number of transformants screened

Figure 7.2 Analysis of various RiboCas systems. **(a)** Conjugation efficiency of different RiboCas plasmids involved in CRISPR-mediated gene editing. Error bars represent standard deviations of three biological replicates on three independent experiments. The data summarize the obtained editing results from each RiboCas plasmid. **(b)** CFU obtained from conjugation of pRECas1C (RiboCas vector lacking both homology arms and sgRNA) on selective media in the presence and absence of the inducer theophylline. **(c)** CFU obtained from conjugation of pRECas1G-IIR (RiboCas vector lacking the homology arms) on selective media in the presence and absence of theophylline. **(d)** Confirmation of the gene deletion using colony PCR. L is 2-log ladder; lanes 1-8, colony PCR on colonies from selective plates containing theophylline; lanes 9-16, colony PCR on colonies from selective plates lacking theophylline; WT (wildtype).

Individual colonies harbouring the different CRISPR/Cas9 vectors and the DNA editing template were picked and transferred to selective plates in the presence of theophylline. In all cases, amplicons of ~2.2 kbp instead of ~2.9 kbp size were detected after screening, implying the expected ~0.7 kbp deletion of *spoIIR* and an editing efficiency of 100% (Figure 7.2d). Desired deletions were confirmed by sequencing the PCR products. Screening of uninduced colonies generated the amplicon of the WT strain (2.9 kbp); sequencing of these PCR products did not show insertions/deletions. These results suggest that the dynamic range of riboswitch-E is appropriate for a Cas9-base editing system,

providing both undetectable off-target activity and high selection capacity. The consistency of these results was confirmed by three independent experiments, verifying the robustness of the RiboCas system and its suitability for more ambitious genome editing strategies.

Strain engineering usually requires sequential rounds of genome editing. For this reason, the plasmid must be cured from the engineered host, which might also prevent undesired off-target events associated with the presence of Cas9 in the cell. Accordingly, successfully edited *C. sporogenes* cells (*C. sporogenes* Δ*spoIIR*) harbouring the RiboCas plasmid were cultured in non-selective liquid TYG medium. After 24 hours of growth, individual colonies were isolated from non-selective plates and screened for their ability to grow on thiamphenicol by patch plating them onto plates in the presence and absence of antibiotic supplementation. Cells susceptible to the antibiotic were PCR screened using primers IC273-r and IC274-f. A plasmid curing efficiency of 60% was obtained (18/30 colonies screened), indicating that the replicon pCB102, proven to be stable in *C. sporogenes* NCIMB 10696[34], is suitable for genome editing purposes in such strain.

7.3.4 RiboCas for more ambitious genome editing strategies

To further illustrate the versatility of the RiboCas system it was attempted to both delete and integrate larger fragments into the genome of *C. sporogenes*. The 2.4 kbp *spoIIE* gene (*Clspo_c37040*) was selected as the target locus, on the basis of its large size and potential usefulness in the work related to the conditionally sporulating strain. Two sets of homology arms were designed to either delete the target sequence or to integrate a 3 kbp cassette

(Figure 7.3a). The editing templates and the sgRNA sequences targeting *spoIIE* were cloned into pRECas1 (the most efficient RiboCas system in *C. sporogenes* in terms of conjugation efficiency) generating pRECas1-IIE (deletion vector) and pRECas1_IIEin (integration vector). Both vectors were successfully conjugated into *C. sporogenes.* After conjugation, induction of *cas9* expression was performed by re-streaking individual colonies harbouring pRECas1-IIE and pRECas1-IIEin onto plates supplemented with thiamphenicol and 5mM theophylline. PCR was performed on uninduced cultures and on colonies that had grown for 24 hr in the presence of the inducer. All 100% of the screened colonies were confirmed as positive mutants for *spoIIE* deletion and integration of the cassette, respectively (Figure 7.3b). Generally, PCR products obtained from colonies grown in the absence of the inducer were found to be from unedited cells, demonstrating the tightness of RiboCas system in the absence of the inducer. Only a few colonies (<10%) were composed of mixtures of mutant and WT cells. A plasmid curing efficiency of approximately 70% was obtained on this occasion.

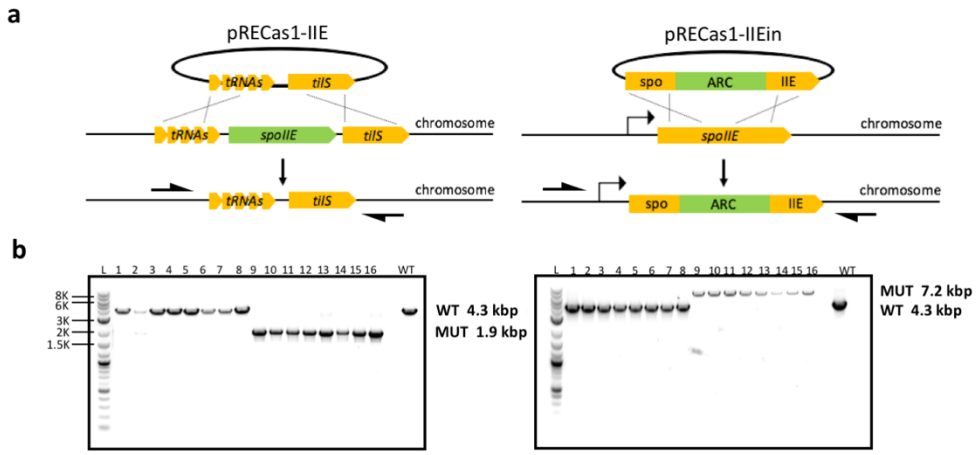


Figure 7.3 Deletion of *spoIIE* and integration of ARC. **(a)** Schematic representation of the deletion of *spoIIE* (via pRECas1_IIE) and the integration of an ARC in the *spoIIE* locus (via pRECas1_IIEin) in the chromosome of *C. sporogenes*. The ARC comprises the *ermB* gene that confers resistance to the antibiotic erythromycin and the sequence encoding the reporter enzyme ß-glucuronidase from *E. coli*. (b) PCR screening of the *spoIIE* deletion and integration of the ARC. The 1.9, 7.2 and 4.3 kbp bands represent the *spoIIE* deletion, ARC integration and WT genotypes, respectively. Lane L, 2-log DNA marker with sizes (kbp) on the left; lanes 1-8 uninduced colonies; lanes 9-16 induced colonies.

7.3.5 Demonstration of the universality of RiboCas

To determine the applicability of the theophylline responsive riboswitches and the RiboCas editing tool in other clostridial species it was firstly sought to delete the same gene, *spoIIE*, in the solventogenic species *C. pasteurianum* and in the pathogens *C. difficile* and *C. botulinum*. Previous work had enabled the use of Cas9 for editing purposes in *C. pasteurianum*[231,232]; however, the strategies employed suffered from very low transformation efficiencies that might compromise more ambitious applications. This is also the case in *C. difficile*[237,238], where the performance of CRISPR was shown to be suboptimal. In particular, conjugation efficiencies of the CRISPR plasmid was invariably very low. This is likely due to the high toxicity associated with the

Cas9 endonuclease activity. The recent application of a CRISPR-Cpf1 editing tool enabled the deletion of various genes with high efficiency[241]. However, because Cpf1 nuclease recognizes a T-rich PAM sequence (5′-TTTN-3′), it is hypothesised that off-target events across T-rich genomes (such as those from species belonging to the genus *Clostridium*) are more likely to occur if compared with Cas9, which recognises the PAM sequence 5′-GGN-3′[242].

Vectors pRECas1, containing the homology arms and the sgRNA to target the *spoIIE* gene in each species were introduced into the host organisms *C. pasteurianum* DSM 525-H1[53], *C. difficile* 630 and the group I *C. botulinum* strain ATCC 3502. *C. pasteurianum* was transformed via electroporation whereas plasmids were introduced into *C. difficile* and *C. botulinum* by conjugative plasmid transfer from an *E. coli* donor. Simultaneously, the control plasmid pRECas1C (RiboCas vector lacking both homology arms and sgRNA), was also introduced into the different hosts to determine the leakiness of the system and its impact on the transformation/conjugation process. All transformations were plated on appropriate media supplemented with thiamphenicol, both in the presence and absence of theophylline. The transformation frequency of the targeting RiboCas vector in *C. pasteurianum* was only 20% lower than that of the control vector (5000 CFU/μg DNA in pRECas1 *vs* 6500 CFU/μg DNA in pRECas1C) (Figure 7.4). On average a transformation efficiency of 75 CFU/μg DNA was obtained on plates containing theophylline, demonstrating that the theophylline responsive riboregulator is functional in *C. pasteurianum*. PCR screening of these colonies confirmed the mutant genotype of the edited cells without the need for further re-streaking; the unedited phenotype was maintained on plates lacking the inducer. Colonies from

non-induced plates were individually streaked onto plates containing both thiamphenicol and theophylline and colony PCR used to confirm their identity as deletion mutants in 24 out of the 24 colonies screened (100% efficiency) (Table 7.1). A similar transformation pattern was found when conjugating the RiboCas vectors into *C. difficile* and *C. botulinum* (data not shown), validating the induction ability of riboswitches and the efficiency of the RiboCas tool in both organisms. Screening of *C. difficile* colonies after induction confirmed a mutation efficiency of 100% while maintaining the WT phenotype on colonies not exposed to theophylline (Table 7.1). Interestingly, and in contrast to the results observed in other clostridial species, *C. botulinum* colonies on plates lacking theophylline were composed of either mutants or a mixture of mutant and WT populations. This may suggest that although inducible, riboswitches exhibit a higher background expression in *C. botulinum* but that the selective capacity of Cas9 still allows the selection of only edited cells. The efficiency of plasmid curing in *C. pasterurianum*, *C. difficile* and *C. botulinum* was approximately of 80, 5 and 60%, respectively. Whilst the efficiencies observed in *C. pasteurianum* and *C. botulinum* are sufficient for genome editing purposes, the low efficiency obtained in *C. difficile* indicates that a more appropriate replicon is needed to favour plasmid loss in this strain.

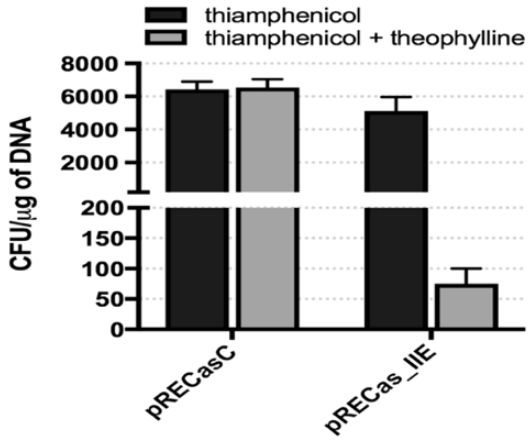


Figure 7.4 Transformation efficiencies of RiboCas plasmids in *C. pasteurianum*, without (pRECasC) or with (pRECas_IIE) the *spoIIE*-targeting sgRNA and the DNA editing template, when plated on selective media in the absence and presence of the inducer theophylline.

A summary of all gene editing attempts carried out for the validation of the RiboCas system in *Clostridium*, including insertions, deletions and nucleotide substitutions are summarized in Table 7.1. This work was done in collaboration with two other PhD students of the group, Maria Zygouropoulou, who worked on *C. sporogenes* and Daphne Groothuis, who carried out all *C. botulinum* work. It is worth noting that, despite pRECas1 being the most efficient in this study (in terms of conjugation efficiency), vectors incorporating the promoter J23119 might be suitable for organisms where P_{araE} is expected to be natively repressed (i.e., in *C. acetobutylicum*).

Finally, to validate the efficiency and detect off-target events that could have been caused by this tool, the genome of six independent mutants of *C. sporogenes* (*C. sporogenes ΔclosA-I* and *C. sporogenes ΔclosA-I-NH* (Table 7.1); three replicas of each) were sequenced. In all of them, no modifications (apart from the intended editions) were detected when compared to the re-sequenced *C. sporogenes* parent strain. This work was done in collaboration

with another PhD student of the group, Maria Zygourupoulou, who had carried out the genome editing process in these strains.

In summary, a novel CRISPR Cas9-based system – RiboCas – that successfully works in several clostridial species, including the non-pathogen *C. sporogenes*, the solventogenic *C. pasteurianum* and the pathogens *C. difficile* and *C. botulinum* has been generated in this study. This novel system overcomes the main obstacles associated with Cas9 editing tools, including the very low transformation efficiencies and the inability to select only edited cells as a result of excessive Cas9 toxicity[243,244]. For this reason, it is suggested that temporal and minimum exposure of the cells to Cas9 has three main benefits: (i) it provides time for the homologous recombination event to occur before Cas9 is expressed; (ii) it reduces the likelihood of potential off target effects, and; (iii) it diminishes the possibility of mutations that inactivate Cas9 or the sgRNA, which result in "escaper" colonies that cannot be distinguished from the edited cells. Because riboswitches can be adjusted to the desired regulatory window, a riboswitch-based CRISPR system allows very low basal expression of Cas9 in the absence of inducer and an adequate level of gene expression after induction, fulfilling the benefits of an inducible Cas9-based system.

Table 7.1 Summary of genome editing results in different *Clostridium* species

Strain	Target	Modification		Protospacer sequence	[a]efficiency (%)
		Deletion (kpb)	Insertion (kbp)		
C.pasteurianum DSM 525	*spoIIE*	2.4	---	ggaaactcctaaatatcacg	100
C.difficile 630	*spoIIE*	2.4	---	ggaattttagatttaaaacg	100
C.botulinum ATCC 3502	*spoIIE*	2.4	---	ttttataaagcgaggaactg	100
C.sporogenes NCBI 10696	*spoIIE*	2.4	---	ttaattataggatatgtaaa	100
	spoIIE	---	2.9	ttaattataggatatgtaaa	100
	spoIIR	0.7	---	ggtaatataacattgccaca	100
	c16380	0.63	---	acagtttttttcaattttcgg	100
	c03700	silent mutation	silent mutation	ttacaactactatccaaaga	94
	c01750	1.4	---	aagtactccgattaaaacag	84
	c04580	1	1	ggtttagaagaaatagcggg	83
	c14540	---	0.8	gacaaaacaagtcaaataga	67
	*c05250**	8.5	---	ggactcttaaccccaactaa	57
	c14780	0.6	---	ccaaatatagtccaacctaa	56
	*closC749::CT***	10	---	tggatattcaccgaacacta	43
	pyrE	0.4	0.8	atagtaggaccagctatggg	42
	c31080	0.24	---	gttgttttaattgatcacaa	38
	c29800	1.3	---	acttgtagcggcagaagcac	36

[a](Number of correctly edited transformants/total number of transformants screened) x 100
Resultant strains * *C. sporogenes ΔclosA-I* ** *C. sporogenes ΔclosA-I-NH*

7.4 Key outcomes

1. Generation of RiboCas, a highly efficient, riboswitch-Cas9-based editing tool that regulates the expression of Cas9 at the level of translation. The tool allows high transformation efficiency and selection capacity in the absence and presence of induction, respectively.
2. Demonstration of the functionality of RiboCas for large gene deletion and insertion in *C. sporogenes*.
3. Demonstration of the universality of RiboCas by obtaining chromosomal modifications in four different clostridial species: *C. sporogenes*, *C. pasteurianum*, *C. difficile* and *C. botulinum*.

Chapter 8
/ General Discussion

As a scientist, I clearly see the potential for harnessing the power of nature
— Craig Venter

8.1 Context

Through the application of engineering principles, synthetic biology provides the means to harness the power of *living things* for the good of humankind. Central to synthetic biology is the ability to assemble standardized biological modules to create systems with useful applications[245]. For this reason, one of the main efforts of synthetic biologists in the last few years has been to characterize and standardize biological parts[75–79]. However, these parts have been largely created for model organisms (i.e., *E. coli, B. subtilis, S. cerevisiae*), with only a very limited selection available for non-model hosts, such as *Clostridium* spp.

Why *Clostridium*? Clostridial strains have been important to biotechnology for years. One such example is the solvent-producing clostridial species *C. acetobutylicum*, which since the beginning of the 20th century has been exploited for its ability to convert sugars into acids and solvents via the well-known acetone-butanol-ethanol (ABE) fermentation[246]. Today, genetic engineering techniques not only allow manipulation of the genome of *C. acetobutylicum* to optimize solvent production, but also to introduce genes and pathways from diverse organisms to produce all manner of things, including drugs, foods or fuels. This also applies to the large variety of clostridial organisms that can be used as "production chassis". The exploitation of their natural abilities (e.g., carbon fixation, tumour colonization, cellulosic and hemicellulosic biomass degradation[84]), can serve, for instance, to produce valuable chemicals whilst reducing the amount of CO_2 present in the atmosphere or allow them to be utilized as anticancer agents. Synthetic biology tools for

manipulating clostridial hosts can facilitate these uses. Consequently, biological parts and systems that are functional in *Clostridium* are needed.

Besides the aforementioned attributes, clostridia cells also have the ability to sporulate. Interestingly, the resilient properties of spores make them attractive for use in a wide range of biotechnological applications. Nevertheless, from a safety perspective, spores enable the organism to survive for prolonged periods of times if release into the environment. As such, there is need for biological containment.

A number of bio-strategies exist to minimize the "escape" of GMOs into the environment[41]. Thus far, none of the individual strategies have been proven to be completely successful as containment measurements and it seems that the combination of more than one "containment layer" is necessary to completely contain GMOs. Nevertheless, the potential risk of recombinant spores shed into the environment has not been, as yet, addressed. In this work, it is proposed that a conditionally sporulating strain, meaning a strain capable of sporulating only under certain growth conditions, could prevent the bacterial spores of a GMO from being formed and subsequently released into the environment.

Controlling the flow of information in a genetic network is probably one of the main obstacles in genetic engineering projects. For example, in order to construct a molecular machine or a genetic circuit, the expression of proteins or RNA molecules that control each other's levels within the circuit needs to be balanced so that it performs the predicted logical function. Balancing expression can be achieved through a variety of genetic parts, of which the most precious are those that can be regulated through the agency of an external stimulus. These parts (i.e., inducible systems) allow, for instance, re-wiring of a specific genetic

network so that a certain process is temporarily regulated (i.e., sporulation). Additionally, inducible systems can be useful to temporarily regulate the expression of detrimental or toxic proteins (i.e., Cas9 in CRISPR editing approaches) or to fine-tune the amount of each component in a genetic circuit or pathway so that it behaves optimally.

Therefore, the projects presented in this thesis have attempted to: (i) provide novel genetic tools (and biological parts) to make *Clostridium* a more amenable microbial chassis for synthetic biology applications, and; (ii) provide biological containment of clostridial spores to facilitate their future implementation in biotechnology applications.

8.2 Original contributions

The design of novel and synthetic "devices" capable of tightly regulating gene expression requires, at the very least, the pre-characterization of the individual composite parts. For this reason, here it was decided to characterize one of the most effective and widely used clostridial inducible sytem, the LAC system from *C. perfringens strain* 13 (Chapter 3). The host, which would be used to test all the tools generated in this study, was *C. sporogenes* NCIMB 10696. This strain represents the most promising candidate for clostridial-based cancer therapies and a focus of this research group (SBRC). Results demonstrated that the LAC system from *C. perfringens* strain 13 is functional in *C. sporogenes*. Substantial background expression, however, was observed in the absence of inducer. This, in turn, compromised the dynamic range of the system (this was approximately 15-fold). Characterization of the LAC system provided information on its functionality that had not previously been reported.

This included the identification of the transcriptional start site and the putative -10 and -35 promoter elements of both, the transcriptional regulator (BgaR) and the regulated gene, as well as the region to which BgaR binds.

In this thesis, it has also been demonstrated, for the first time, that theophylline-dependent riboswitches function in a clostridial host, with a response that is dependent on inducer concentration (Chapter 4). In particular, the novel riboswitch-G, which was rationally designed to function in *Clostridium*, outperformed previously published riboswitches, exhibiting a dynamic range of approximately 15-fold and very tight repression. It was also shown that by replacing the promoter and modifying the 5′ UTR sequence located upstream of the riboswitch, libraries of inducible switches with a broad spectrum of dynamic ranges could be created. These can be suitable for applications where high levels of gene expression are needed as well as for the expression of detrimental proteins. In all cases, the riboswitch was able to respond to the inducer theophylline, ensuring its performance independent of the genetic context and consolidating the theophylline responsive riboswitch as a novel device to regulate gene expression in clostridia.

The systems characterized in chapters 3 and 4, one regulating transcription (LAC system) and the other translation (theophylline responsive riboswitches) resulted in very similar dynamic ranges (approximately 15-fold). However, whereas the LAC system exhibited high levels of gene expression and substantial leakiness in the presence and absence of the inducer, respectively, the theophylline responsive riboswitches were tightly repressed only if low levels of expression were obtained after induction. Importantly, the synthetic biology toolbox of a certain organism (or genus) requires inducible systems with

different dynamic ranges so that they can be used in a wide range of applications, but *Clostridium* was until now, lacking a system exhibiting both reduced background expression and high levels of expression in the OFF and ON states, respectively.

The clue as to how to achieve such a system was obtained when characterizing the LAC system from *C. perfringens* strain 13. The deletion of *bgaR* from the plasmid containing the LAC system resulted in reduced background expression in the absence of inducer, as well as in the inability to activate the expression of the reporter gene. If expression of the transcriptional activator BgaR was tightly repressed, leakiness could be potentially reduced. Accordingly, the theophylline responsive riboswitch was combined with the transcriptionally regulated LAC system to generate a dual translational-transcriptional ON device that controls the expression of both, *bgaR* and the gene of interest (Chapter 5). In this system, termed RiboLac, optimal expression of the target gene requires induction with both theophylline and lactose. Its reduced basal expression resulted in the improvement of the dynamic range from approximately 15-fold to 40-fold. Additionally, the enhanced performance of the RiboLac was further demonstrated by direct comparison with an engineered system where the dual control is only exerted over the target gene.

Additionally, and as one of the main aims of this thesis, a strain with conditional control of spore formation was generated (Chapter 6). The hypothesis was that the ability to sporulate could perhaps be regulated by placing a gene that is essential for this process under the control of an inducible system. As (by definition) a typical synthetic biology approach, the creation of an adequate conditionally sporulating strain required several cycles of design,

building and testing. This was mainly related to the fact that the appropriate behaviour of the strain (no spores in the absence of the inducer and ability to sporulate like WT after induction) needed a system that maintained the level of expression of the selected key sporulation gene within a certain range. The gene encoding the master regulator Spo0A was the first candidate considered, and as such, a strain deficient in this gene – and host to test the different inducible systems – was generated. Plasmid-based analysis involving the expression of *spo0A* under the control of the systems developed in this study (i.e., LAC system, theophylline responsive riboswitches and RiboLac) were initially performed. Results showed that the systems that exhibited the tightest repression when analysing reporter systems in Chapters 3, 4 and 5 (P_{fdx4}-E and P_{fdx}*-E) were the only ones capable of preventing spore formation in the absence of inducer. However, these strains were incapable of sporulating in a WT manner after induction, indicating that a wider dynamic range was necessary to regulate sporulation in *C. sporogenes*. Integration of the different systems in the chromosome of *C. sporogenes* (specifically in the *pyrE* locus) resulted in one strain, *C. sporogenes*::RiboLac-*spo0A*, exhibiting the desired sporulation profile: no spores in the absence of inducers and a sporulation capacity comparable to WT after induction. This further demonstrated the superior ability of the dual system RiboLac to tightly regulate gene expression (the strain containing *spo0A* under the control of the unmodified LAC system did sporulate in the absence of inducer). Additionally, the inducible regulation of Spo0A with systems that had exhibited different levels of CAT expression indicated that a certain threshold of Spo0A might be necessary to activate the sporulation program in *C. sporogenes,* in contrast to what had been recently observed in *C.*

difficile[220]. The strain generated in this study, in addition to offering an ideal means of biocontainment in spore-based applications, could be used to study the *spo0A* regulon in *C. sporogenes*, since this species is also used by the food industry as surrogate of the neurotoxin-producing *C. botulinum* group I strains[247]. Spores play a crucial role in the food contamination and poisoning processes associated with *C. botulinum*[248] and, for this reason, sporulation represents a critical intervention point to improve food safety. Additionally, the transferability of this system to other *Clostridium* species could also serve to analyse certain aspects linked to the *spo0A* regulon, including the regulation of solvent production[249].

As an alternative approach, generating a conditionally sporulating strain of *C. sporogenes* was attempted via regulation of the expression of the *spoIIR* gene, encoding a stage II sporulation protein. In this case, and because *spoIIR* is transcribed in association with the sporulation specific sigma factor σ^F, the regulatory elements of the promoter to which this sigma factor binds had to be maintained. For this reason, only the theophylline responsive riboswitch (in particular, riboswitch-G) was considered, as riboswitches can be located downstream of any promoter. The synthetic regulation of *spoIIR* achieved by placing the riboswitch downstream of its own promoter (P_{spoIIR}*) could not prevent spore formation in the absence of the inducer, indicating that an even tighter repression of the gene was required in this case. Interestingly, generation of the *C. sporogenes* Δ*spoIIR* strain and its subsequent complementation revealed that *spoIIR* might have a "positional regulation" in *C. sporogenes*.

Finally, and to further contribute to the clostridial synthetic biology "toolkit", the theophylline responsive riboswitches were used to generate a

universal and highly efficient inducible Cas9-based system, termed RiboCas (Chapter 7). This novel system overcome the main obstacles associated with Cas9-based editing tools, including the very low transformation efficiencies and the inability to select only edited cells as a result of excessive Cas9 toxicity. Its universality in clostridia was demonstrated by generating mutants in three other different *Clostridium* species: *C. botulinum, C. difficile and C. pasteurianum*. Therefore, RiboCas will be possibly the method of choice for future genome editing purposes in *Clostridium*.

It is important to note that since the riboswitches were functional in all clostridial hosts tested thus far (exemplified through the RiboCas tool), we anticipate that the RiboLac should be functional, at least, in those species where the unmodified LAC system has been proven to work, including *C. difficile*[241], *C. perfringens*[102], *C. ljungdahlii*[107], *C. tyrobutyricum*[146], *Clostridium saccharoperbutylacetonicum*[250], *C. beijerickii*[236] and *C. acetobutylicum*[233].

8.3 Perspectives

At the onset of this project, there were only a limited number of genetic tools available to engineer *Clostridium*. These included procedures for creating random[224,251] and directed mutagenesis[222], methods for genome editing (recently reviewed in[84]), and the formulation of standardized plasmid systems[109]. Nevertheless, a wide range of inducible systems, indispensable for the generation of novel genetic networks, was missing. Now, having been added to the toolbox, there are some other basic parts that are still needed, including appropriate insulators or a large library of RBSs, which are expected to be developed in the near future. Furthermore, adoption of a single reporter system,

which could be used to measure the activity of all different parts tested, has not yet been implemented. This is necessary to facilitate the comparison of data between different laboratories.

To my knowledge, this is the first time that a riboregulator has been used to control the expression of Cas9 in CRISPR applications. Besides its genetic engineering prospects in *Clostridium*, where it has proven to be extremely efficient, the tandem riboswitch-*cas9* could be readily transferred to other genera or kingdoms. Riboswitches are not species-specific, so they could potentially work on any organism.

The reason behind utilising *C. sporogenes* and not any other industrially relevant clostridial strain to generate a conditionally sporulating strain was the direct involvement of this research group in clostridial-based cancer therapies. Substantial efforts have been undertaken so far to bring such therapies to the clinic, and biocontainment could now be a potentially surmountable obstacle. In this context, it would be necessary to evaluate if the conditionally sporulating strain behaves as such in animal models and remains asporogeneous in the absence of inducers as well as the robustness and stability of conditionally sporulating after long periods of time. Additionally, albeit beyond the scope of this work, it is important to evaluate the potential effect that the engineered strain could have in expressing the anticancer "cargo". For instance, it could be possible that the colonization capacity of the strain is affected by the synthetic regulation of *spo0A*, in which case, new target genes would need to be evaluated, built and tested.

Despite the successful development of a conditionally sporulating strain, as a biocontainment measure, *C. sporogenes*::RiboLac-*spo0A* could only be

used in applications where the germinated cell would not be in contact with lactose or theophylline. For example, biocontainment obtained via conditionally regulating sporulation in strains intended to be used as probiotics would (ideally) require inducible systems that are not induced by molecules found in edible products (i.e., lactose or theophylline). As a proof-of-concept, this work has shown that a given inducible system can be engineered to improve its dynamic range; however, other natural or synthetic inducible systems can also be envisioned as potential regulators of the sporulation cascade.

Bibliography

1. Deichmann, U. Chemistry and the Engineering of Life Around 1900: Research and Reflections by Jacques Loeb. *Biol. Theory* **4,** 323–332 (2009).
2. Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science* **355,** 1040–1044 (2017).
3. Agapakis, C. M. Designing synthetic biology. *ACS Synh. Biol.* **3,** 121–128 (2014).
4. Royal Academy of Engineering. *Synthetic Biology: scope, applications and implications*. (2009). Available at: http://www.raeng.org.uk/societygov/policy/current_issues/synthetic_biology/pdf/Synthetic_biology.pdf.
5. Keller, E. F. Knowing As Making, Making As Knowing: The Many Lives of Synthetic Biology. *Biological Theory* **4**(4), 333–339 (2009)
6. Stephanopoulos, G. Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* **1,** 514–525 (2012).
7. Bradley, R. W., Buck, M. & Wang, B. Tools and Principles for Microbial Gene Circuit Engineering. *J. Mol. Biol.* **428,** 862–888 (2016).
8. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–6 (2010).
9. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351,** aad6253 (2016).
10. Shao, Y. *et al.* Creating a functional single-chromosome yeast. *Nature* **560,** 331–335 (2018).
11. Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.* **13,** 406–420 (2012).
12. Quianzon, C. C. & Cheikh, I. History of insulin. *J. community Hosp. Intern. Med. Perspect.* **2,** (2012).
13. Jackson, D. A., Symonst, R. H., Berg, P., Lobban, P. & Kaiser, A. D. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. *Proc. Nat. Acad. Sci.* **69,** (1972).
14. Cohen, S. N., Chang, A. C. Y., Boyert, H. W. & Hellingt, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc. Nat. Acad. Sci.* **70,** (1973).
15. Khan, S. *et al.* Role of Recombinant DNA Technology to Improve Life. *Int. J. Genomics* **2016,** 1–14 (2016).
16. Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J. & Setlow, P. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol. Mol. Biol. Rev.* **64,** 548–72 (2000).
17. Setlow, P. Spore Resistance Properties. *Microbiol. Spectr.* 1–14 (2014).
18. Duc, L. H., Hong, H. A., Barbosa, T. M., Henriques, A. O. & Cutting, S. M. Characterization of *Bacillus* probiotics available for human use. *Appl. Environ. Microbiol.* **70,** 2161–71 (2004).
19. Zhang, L. *et al.* Effects of dietary supplementation of probiotic, *Clostridium butyricum*, on growth performance, immune response, intestinal barrier function, and digestive enzyme activity in broiler chickens challenged with *Escherichia coli* K88. *J. Anim. Sci. Biotechnol.* **7,** 3 (2016).
20. Cassir, N., Benamar, S. & La Scola, B. *Clostridium butyricum*: from beneficial to a new emerging pathogen. *Clin. Microbiol. Infect.* **22,** 37–45 (2016).
21. Potot, S., Serra, C. R., Henriques, A. O. & Schyns, G. Display of recombinant proteins on *Bacillus subtilis* spores, using a coat-associated enzyme as the

carrier. *Appl. Environ. Microbiol.* **76,** 5926–33 (2010)..
22. Chen, H. *et al.* Orlistat response to missense mutations in lipoprotein lipase. *Biotechnol. Appl. Biochem.* **64,** 464–470 (2017).
23. Mauriello, E. M. F. *et al.* Display of heterologous antigens on the *Bacillus subtilis* spore coat using CotC as a fusion partner. *Vaccine* **22,** 1177–1187 (2004).
24. Valdez, A., Yepiz-Plascencia, G., Ricca, E. & Olmos, J. First *Litopenaeus vannamei* WSSV 100% oral vaccination protection using CotC::Vp26 fusion protein displayed on *Bacillus subtilis* spores surface. *J. Appl. Microbiol.* **117,** 347–357 (2014).
25. Nguyen, A. T. V. *et al. Bacillus subtilis* spores expressing the VP28 antigen: a potential oral treatment to protect *Litopenaeus vannamei* against white spot syndrome. *FEMS Microbiol. Lett.* **358,** 202–208 (2014).
26. Zhou, Z. *et al.* Expression of *Helicobacter pylori* urease B on the surface of *Bacillus subtilis* spores. *J. Med. Microbiol.* **64,** 104–110 (2015).
27. Sibley, L. *et al.* Recombinant *Bacillus subtilis* spores expressing MPT64 evaluated as a vaccine against tuberculosis in the murine model. *FEMS Microbiol. Lett.* **358,** 170–179 (2014).
28. Chang, J. & Erler, J. Hypoxia-mediated metastasis. *Adv. Exp. Med. Biol.* **772,** 55–81 (2014).
29. Liu, S.-C., Minton, N., Giaccia, A. & Brown, J. Anticancer efficacy of systemically delivered anaerobic bacteria as gene therapy vectors targeting tumor hypoxia/necrosis. *Gene Ther.* **9,** 291–296 (2002).
30. Minton, N. P. Clostridia in cancer therapy. **1,** 1–7 (2003).
31. Heap, J., Theys, J., Ehsaan, M. & Kubiak, A. Spores of *Clostridium* engineered for clinical efficacy and safety cause regression and cure of tumors *in vivo*. *Oncotarget* **5,** 1761–1769 (2014).
32. Kubiak, A. M. & Minton, N. P. The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy. *Res. Microbiol.* ***4,*** 244–254 (2015).
33. Brüggemann, H & Gottschalk, G. *Clostridia: Molecular Biology in the Post-genomic Era* (Caister Academic Press, 2009).
34. Pennington, O. J. The Development of Molecular Tools for the Expression of Prodrug Converting Enzymes in *Clostridium Sporogenes*. *Dr. Phylosophy* (2006).
35. Dang, L. H., Bettegowda, C., Huso, D. L., Kinzler, K. W. & Vogelstein, B. Combination bacteriolytic therapy for the treatment of experimental tumors. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 15155–60 (2001).
36. Roberts, N. J. *et al.* Intratumoral injection of *Clostridium novyi*-NT spores induces antitumor responses. *Sci. Transl. Med.* **6,** 1–26 (2014).
37. Mowday, A. *et al.* Advancing Clostridia to Clinical Trial: Past Lessons and Recent Progress. *Cancers* **8,** 63 (2016).
38. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12,** 415–418 (2015).
39. Haynes, K. A. Synthetic biology: Building genetic containment. *Nat. Chem. Biol.* **12,** 55–56 (2016).
40. Miller, S. R. & Bergmann, D. Biocontainment design considerations facilities for biopharmaceutical. *Ind. Biotechnol.* **11,** 223–234 (1993).
41. Torres, L., Krüger, A., Csibra, E., Gianni, E. & Pinheiro, V. B. Synthetic biology approaches to biological containment: pre-emptively tackling potential risks. *Essays Biochem.* **60,** 393–410 (2016).
42. Chan, C. T. Y., Lee, J. W., Cameron, D. E., Bashor, C. J. & Collins, J. J. 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat. Chem. Biol.* **12,** 82–86 (2016).
43. Wright, O., Delmans, M., Stan, G. B. & Ellis, T. GeneGuard: A modular plasmid

system designed for biosafety. *ACS Synth. Biol.* **4,** 307–316 (2015).
44. Hosseini, S., Curilovs, A. & Cutting, S. M. Biological Containment of Genetically Modified *Bacillus subtilis*. *Appl. Environ. Microbiol.* **84,** 1–15 (2018).
45. Rovner, A. J. *et al.* Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518,** 89–93 (2015).
46. Moe-Behrens, G. H. G., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4,** 5 (2013).
47. Blaschek, H. P. *Encyclopedia of Food Microbiology* (Elsevier, 2014).
48. Minton, N. P. *et al.* A roadmap for gene system development in *Clostridium*. *Anaerobe* **41,** 104–112 (2016).
49. Uzal, F. A. *et al.* Towards an understanding of the role of *Clostridium perfringens* toxins in human and animal disease. *Future Microbiol.* **9,** 361–377 (2014).
50. Ofosu, A. *Clostridium difficile* infection: a review of current and emerging therapies. *Ann. Gastroenterol.* **29,** 147–154 (2016).
51. Carter, A. T. & Peck, M. W. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. *Res. Microbiol.* **166,** 303–317 (2015).
52. Sabra, W., Wang, W., Surandram, S., Groeger, C. & Zeng, A. Fermentation of mixed substrates by *Clostridium pasteurianum* and its physiological, metabolic and proteomic characterizations. *Microb. Cell Fact.* **15,** (2016).
53. Schwarz, K. M. *et al.* Towards improved butanol production through targeted genetic modification of *Clostridium pasteurianum*. *Metab. Eng.* **40,** 124–137 (2017).
54. Xue, C., Zhao, J., Chen, L., Yang, S.-T. & Bai, F. Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by *Clostridium acetobutylicum*. *Biotechnol. Adv.* **35,** 310–322 (2017).
55. Heinze, S. *et al.* Identification of endoxylanase XynE from *Clostridium thermocellum* as the first xylanase of glycoside hydrolase family GH141. *Sci. Rep.* **7,** 11178 (2017).
56. Maru, B. T., Munasinghe, P. C., Gilary, H., Jones, S. W. & Tracy, B. P. Fixation of CO2 and CO on a diverse range of carbohydrates using anaerobic, non-photosynthetic mixotrophy. *FEMS Microbiol. Lett.* **365,** (2018).
57. Xu, H. *et al.* A study of CO/syngas bioconversion by *Clostridium autoethanogenum* with a flexible gas-cultivation system. *Enzyme Microb. Technol.* **101,** 24–29 (2017).
58. Mukherjee, S. *The gene: an intimate history* (Scribner, 2016).
59. Dürre, P. (Peter). *Handbook on Clostridia*. (Taylor & Francis, 2005).
60. Setlow, P. Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. *J. Appl. Microbiol.* **101,** 514–525 (2006).
61. Saujet, L., Pereira, F. C., Henriques, A. O. & Martin-Verstraete, I. The regulatory network controlling spore formation in *Clostridium difficile. FEMS Microbiol. Lett.* **358,** 1–10 (2014).
62. Khvorova, A., Chary, V. K., Hilbert, D. W. & Piggot, P. J. The Chromosomal Location of the *Bacillus subtilis* Sporulation Gene *spoIIR* Is Important for Its Function. *J. Bacteriol.* **182,** (2000).
63. Talukdar, P. K., Olguín-Araneda, V., Alnoman, M., Paredes-Sabja, D. & Sarker, M. R. Updates on the sporulation process in *Clostridium* species. *Res. Microbiol.* **166,** 225–235 (2015).
64. Gil, F., Lagos-Moraga, S., Calderón-Romero, P., Pizarro-Guajardo, M. & Paredes-Sabja, D. Updates on *Clostridium difficile* spore biology. *Anaerobe* **45,** 3–9 (2017).
65. Tan, I. & Ramamurthi, K. Spore formation in *Bacillus subtilis*. *Environ. Microbiol. Rep.* **6,** 212–225 (2014).
66. Al-Hinai, M. A., Jones, S. W. & Papoutsakis, E. T. The *Clostridium* Sporulation

Programs: Diversity and Preservation of Endospore Differentiation. *Microbiol. Mol. Biol. Rev.* **79,** 19–37 (2015).
67. Deakin, L. J. *et al.* The *Clostridium difficile spo0A* gene is a persistence and transmission factor. *Infect. Immun.* **80,** 2704–11 (2012).
68. Fujita, M. & Losick, R. Evidence that entry into sporulation in *Bacillus subtilis* is mediated by gradual activation of a master regulator. *Genes Dev.* **19,** 2236–2244 (2005).
69. Hilbert, D. W. & Piggot, P. J. Compartmentalization of Gene Expression during *Bacillus subtilis* Spore Formation. *Microbiol. Mol. Biol. Rev.* **68,** 234–262 (2004).
70. Parker, G. F., Daniel, R. A. & Errington, J. Timing and genetic regulation of commitment to sporulation in *Bacillus subtilis*. *Microbiology* **142,** 3445–3452 (1996).
71. Fimlaid, K. A. *et al.* Global analysis of the sporulation pathway of *Clostridium difficile*. *PLoS Genet.* **9,** 1–20 (2013).
72. Wolf, L., Silander, O. K. & van Nimwegen, E. Expression noise facilitates the evolution of gene regulation. *Elife* **4,** 1–48 (2015).
73. Endy, D. Foundations for engineering biology. *Nature* **438,** 449–453 (2005).
74. Lu, T. K., Khalil, A. S. & Collins, J. J. Next-generation synthetic gene networks. *Nat. Biotechnol.* **27,** 1139–1150 (2009).
75. Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M. & Gennari, J. H. Standard Biological Parts Knowledgebase. *PLoS One* **6,** 1–8 (2011).
76. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).
77. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30,** 1137–42 (2012).
78. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated Design of Synthetic Ribosome Binding Sites to Precisely Control Protein Expression. *Nat. Biotechnol.* **27,** 946–950 (2009).
79. Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. & Arkin, A. P. RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30,** 1002–1006 (2012).
80. Lodish, H. *et al. DNA Cloning with Plasmid Vectors*. 4th edn (Freeman and Company, 2000).
81. Sasson, V., Shachrai, I., Bren, A., Dekel, E. & Alon, U. Mode of Regulation and the Insulation of Bacterial Gene Expression. *Mol. Cell* **46,** 399–407 (2012).
82. Chen, Y. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10,** 659–664 (2013).
83. Chen, Y. *et al.* Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **9,** 64 (2018).
84. Gyulev, I. S. *et al.* Part by Part: Synthetic Biology Parts Used in Solventogenic Clostridia. *ACS Synth. Biol.* **7,** 311–327 (2018).
85. Müller-Hill, B. *The lac Operon : a short history of a genetic paradigm*. (Walter de Gruyter, 1996).
86. Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem. Biol.* **14,** 173–84 (2007).
87. Brewster, R. C. *et al.* The transcription factor titration effect dictates level of gene expression. *Cell* **156,** 1312–1323 (2014).
88. Collins, C. H., Leadbetter, J. R. & Arnold, F. H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24,** 708–712 (2006).
89. Saenger, W., Hinrichs, W., Orth, P., Schnappinger, D. & Hillen, W. Structural

basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nat. Struct. Biol.* **7,** 215–219 (2000).
90. Arfman, N., Worrell, V. & Ingram, L. O. Use of the *tac* promoter and *lacIq* for the controlled expression of *Zymomonas mobilis* fermentative genes in *Escherichia coli* and *Zymomonas mobilis*. *J. Bacteriol.* **174,** 7370–8 (1992).
91. Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6785–9 (1983).
92. Hoffmann, J. & Altenbuchner, J. Functional characterization of the mannitol promoter of *Pseudomonas fluorescens* DSM 50106 and its application for a mannitol-inducible expression system for *Pseudomonas putida* KT2440. *PLoS One* **10,** (2015).
93. Schindler, B. D., Seo, S. M., Birukou, I., Brennan, R. G. & Kaatz, G. W. Mutations within the *mepA* operator affect binding of the MepR regulatory protein and its induction by MepA substrates in *Staphylococcus aureus*. *J. Bacteriol.* **197,** 1104–14 (2015).
94. Bertram, R. & Hillen, W. The application of Tet repressor in prokaryotic gene regulation and expression. *Microb. Biotechnol.* **1,** 2–16 (2008).
95. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491,** 249–53 (2012).
96. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wiresg'. *Nature* **469,** 212–215 (2011).
97. Liu, C. C. *et al.* An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods* **9,** 1088–1094 (2012).
98. Findeiß, S., Etzel, M., Will, S., Mörl, M. & Stadler, P. F. Design of Artificial Riboswitches as Biosensors. *Sensors* **17,** (2017).
99. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* **41,** 2541–51 (2013).
100. Wachsmuth, M. *et al.* Design criteria for synthetic riboswitches acting on transcription. *RNA Biol.* **12,** 221–231 (2015).
101. Topp, S. *et al.* Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol.* **76,** 7881–4 (2010).
102. Hartman, A. H., Liu, H. & Melville, S. B. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. *Appl. Environ. Microbiol.* **77,** 471–8 (2011).
103. Zhang, J., Liu, Y.-J., Cui, G.-Z. & Cui, Q. A novel arabinose-inducible genetic operation system developed for *Clostridium cellulolyticum*. (2011).
104. Dong, H., Tao, W., Zhang, Y. & Li, Y. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab. Eng.* **14,** 59–67 (2012).
105. Girbal, L. *et al.* Development of a sensitive gene expression reporter system and an inducible promoter-repressor system for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **69,** 4985–8 (2003).
106. Nariya, H., Miyata, S., Kuwahara, T. & Okabe, A. Development and Characterization of a Xylose-Inducible Gene Expression System for *Clostridium perfringens*. *Appl. Environ. Microbiol.* **77,** 8439–8441 (2011).
107. Banerjee, A., Leang, C., Ueki, T., Nevin, K. P. & Lovley, D. R. Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Appl. Environ. Microbiol.* **80,** 2410–6 (2014).
108. Girbal, L. *et al.* Development of a sensitive gene expression reporter system and an inducible promoter-repressor system for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **69,** 4985–8 (2003).

109. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78,** 79–85 (2009).
110. Mojica, F. J. M., Juez, G. & Rodriguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9,** 613–621 (1993).
111. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).
112. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
113. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).
114. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315,** 1709–1712 (2007).
115. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. & van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol.* **34,** 575–587 (2016).
116. Selle, K. & Barrangou, R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends Microbiol.* **23,** 225–232 (2015).
117. Su, T. *et al.* A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Sci. Rep.* **6,** 37895 (2016).
118. Carroll, D. Genome Engineering With Zinc-Finger Nucleases. *Genetics* **188,** 773–782 (2011).
119. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326,** 1509–1512 (2009).
120. Cho, S., Shin, J. & Cho, B.-K. Applications of CRISPR/Cas System to Bacterial Metabolic Engineering. *Int. J. Mol. Sci.* **19,** (2018).
121. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36,** 765 (2018).
122. Lovitt, R. W., Kell, D. B. & Morris, J. G. The physiology of *Clostridium sporogenes* NCIB 8053 growing in defined media. *J. Appl. Bacteriol.* **62,** 81–92 (1987).
123. Horton, R. M. PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. *Mol. Biotechnol.* **3,** 93–9 (1995).
124 Tellmann, G. The E-Method: a highly accurate technique for gene-expression analysis. *Nat. Methods* **3,** i–ii (2006).
125. Heap, J. T. *et al.* Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res.* **40,** 1–10 (2012).
126. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78,** 79–85 (2009).
127. Huang, H. *et al.* CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synth. Biol.* **5,** 1355–1361 (2016).
128. Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. The pMTL nic− cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68,** 139–149 (1988).
129. Mordaka, P. M. & Heap, J. T. Stringency of synthetic promoter sequences in *Clostridium* revealed and circumvented by tuning promoter library mutation rates. *ACS Synth. Biol.* **7,** 672–681 (2017).
130. Silverstein, R. M. The determination of the molar extinction coefficient of reduced DTNB. *Anal. Biochem.* **63,** 281–2 (1975).
131. Shaw, W. V. Chloramphenicol acetyltransferase from chloramphenicol-

resistant bacteria. *Methods Enzymol.* **43,** 737–55 (1975).
132. Martin, R. G. & Rosner, J. L. The AraC transcriptional activators. *Curr. Opin. Microbiol.* **4,** 132–137 (2001).
133. Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J. L. Arac/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **61,** 393–410 (1997).
134. Sheppard, D. & Englesberg, E. Positive Control in the L-arabinose Gene-Enzyme Complex of *Escherichia coli* B/r as Exhibited with Stable Merodiploids. *Cold Spring Harb. Symp. Quant. Biol.* **31,** 345–347 (1966).
135. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of *mar/sox/rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* **34,** 431–441 (1999).
136. Martin, R. G. & Rosner, J. L. Promoter discrimination at class I MarA regulon promoters mediated by glutamic acid 89 of the MarA transcriptional activator of *Escherichia coli*. *J. Bacteriol.* **193,** 506–15 (2011).
137. Sharma, P. *et al.* The multiple antibiotic resistance operon of enteric bacteria controls DNA repair and outer membrane integrity. *Nat. Commun.* **8,** 1444 (2017).
138. Ellenberger, T., Kwon, H. J., Bennik, M. H. J. & Demple, B. Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat. Struct. Biol.* **7,** 424–430 (2000).
139. Griffith, K. L. & Wolf, J. Systematic mutagenesis of the DNA binding sites for SoxS in the *Escherichia coli zwf* and *fpr* promoters: Identifying nucleotides required for DNA binding and transcription activation. *Mol. Microbiol.* **40,** 1141–1154 (2001).
140. Niland, P., Hühne, R. & Müller-Hill, B. How AraC Interacts Specifically with its Target DNAs. *J. Mol. Biol.* **264,** 667–674 (1996).
141. González-Pérez, M. M., Ramos, J. L., Gallegos, M. T. & Marqués, S. Critical nucleotides in the upstream region of the XylS-dependent TOL meta-cleavage pathway operon promoter as deduced from analysis of mutants. *J. Biol. Chem.* **274,** 2286–90 (1999).
142. Ni, L., Tonthat, N. K., Chinnam, N. & Schumacher, M. A. Structures of the *Escherichia coli* transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain. *Nucleic Acids Res.* **41,** 1998–2008 (2013).
143. Schleif, R. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34,** 779–96 (2010).
144. Gillette, W. K., Martin, R. G. & Rosner, J. L. Probing the *Escherichia coli* transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. *J. Mol. Biol.* **299,** 1245–1255 (2000).
145. Al-Hinai, M. A., Fast, A. G. & Papoutsakis, E. T. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Appl. Environ. Microbiol.* **78,** 8112–21 (2012).
146. Zhang, J., Zong, W., Hong, W., Zhang, Z. T. & Wang, Y. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab. Eng.* **47,** 49–59 (2018).
147. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139,** 1366–75 (2009).
148. Dennis, P. P. & Bremer, H. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3,** (2008).

149. Acar, M., Mettetal, J. T. & Van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40,** 471–475 (2008).

150. Blake, W. J. *et al.* Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Mol. Cell* **24,** 853–865 (2006).
151. Huang, L., Yuan, Z., Liu, P. & Zhou, T. Effects of promoter leakage on dynamics of gene expression. *BMC Syst. Biol.* **9,** 16 (2015).
152. Fraser, D. & Kaern, M. A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.* **71,** 1333–1340 (2009).
153. Perdew, G. H., Vanden Heuvel, J. P. & Peters, J. M. *Regulation of gene expression : molecular mechanisms*. (Humana Press, 2006).
154. Yu, L. *et al.* The mechanism of variability in transcription start site selection. *Elife* **6,** e32038 (2017).
155. Matteau, D. & Rodrigue, S. Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5'-Rapid Amplification of cDNA Ends (5'-RACE). *Methods Mol. Biol*. **1334,** 143–159 (2015).
156. Kumar, A., Garg, S. & Garg, N. in *Encyclopedia of Molecular Cell Biology and Molecular Medicine* 1–59 (Wiley-VCH Verlag GmbH & Co. KGaA, 2014).
157. Hughes, J. R. *et al.* Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9830–5 (2005).
158. Francke, C., Kerkhoven, R., Wels, M. & Siezen, R. J. A generic approach to identify Transcription Factor-specific operator motifs; Inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1. *BMC Genomics* **9,** 145 (2008).
159. Huynen, M. A., Gabaldón, T. & Snel, B. Variation and evolution of biomolecular systems: searching for functional relevance. *FEBS Lett.* **579,** 1839–45 (2005).
160. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–9 (2013).
161. Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34,** 369-73 (2006).
162. Carey, M. F., Peterson, C. L. & Smale, S. T. Electrophoretic Mobility-Shift Assays. *Cold Spring Harb. Protoc.* (2013).
163. Carey, M. F., Peterson, C. L. & Smale, S. T. DNase I Footprinting. *Cold Spring Harb. Protoc.* (2013).
164. Hermser, R., Ursen, B. & Wolde, P.R. Combinatorial gene regulation using auto-regulation. *PLoS Comput. Biol*. **6**, e1000813 (2010).
165. Keseler, I.M. et al. EcoCyc: a comprehensive database resource for *Escherichia coli. Nucleic Acid Res*. **33**, 334–337 (2004).
166. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15,** 423–437 (2014).
167. Villa, J. K., Su, Y., Contreras, L. M. & Hammond, M. C. Synthetic Biology of Small RNAs and Riboswitches. *Microbiology spectrum* **6,** 1–18 (2018).
168. Berens, C., Groher, F. & Suess, B. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* **10,** 246–57 (2015).
169. Barrick, J. E. *et al.* New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6421–6 (2004).
170. Mellin, J.R. & Crossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet*. **3**, 150–156 (2015).

171. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822 (1990).
172. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
173. Mironov, A. S. *et al.* Sensing Small Molecules by Nascent RNA. *Cell* **111,** 747–756 (2002).
174. Gao, X. *et al.* Engineering of *Bacillus subtilis* strains to allow rapid characterization of heterologous diguanylate cyclases and phosphodiesterases. *Appl. Environ. Microbiol.* **80,** 6167–74 (2014).
175. Fowler, C. C., Brown, E. D. & Li, Y. Using a Riboswitch Sensor to Examine Coenzyme B12 Metabolism and Transport in *E. coli*. *Chem. Biol.* **17,** 756–765 (2010).
176. Zhou, L.-B. & Zeng, A.-P. Exploring Lysine Riboswitch for Metabolic Flux Control and Improvement of l -Lysine Synthesis in *Corynebacterium glutamicum*. *ACS Synth. Biol.* **4,** 729–734 (2015).
177. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–10 (1990).
178. Weigand, J. E. *et al.* Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity. *Nucleic Acids Res.* **39,** 3363–72 (2011).
179. Lee, S. W., Zhao, L., Pardi, A. & Xia, T. Ultrafast dynamics show that the theophylline and 3-methylxanthine aptamers employ a conformational capture mechanism for binding their ligands. *Biochemistry* **49,** 2943–51 (2010).
180. Wunnicke, D. *et al.* Ligand-induced conformational capture of a synthetic tetracycline riboswitch revealed by pulse EPR. *RNA* **17,** 182–188 (2011).
181. Espah Borujeni, A., Mishler, D. M., Wang, J., Huso, W. & Salis, H. M. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res.* **44,** 1–13 (2016).
182. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263,** 1425–9 (1994).
183. Desai, S. K. & Gallivan, J. P. Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation. *J. Am. Chem. Soc.* **126,** 247–254 (2004).
184. Topp, S. & Gallivan, J. P. Random Walks to Synthetic Riboswitches—A High-Throughput Selection Based on Cell Motility. *ChemBioChem* **9,** 210–213 (2008).
185. Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **37,** 184–92 (2009).
186. Cui, W. *et al.* Engineering an inducible gene expression system for *Bacillus subtilis* from a strong constitutive promoter and a theophylline-activated synthetic riboswitch. *Microb Cell Fact* **15,** 2–13(2016).
187. Dwidar, M. & Yokobayashi, Y. Controlling *Bdellovibrio* bacteriovorus Gene Expression and Predation Using Synthetic Riboswitches. *ACS Synth. Biol.* **6,** 2035–2041 (2017).
188. Rudolph, M. M., Vockenhuber, M. P. & Suess, B. Conditional control of gene expression by synthetic riboswitches in streptomyces coelicolor. *Methods Enzymol.* **550,** 283–299 (2015).
189. Nakahira, Y., Ogawa, A., Asano, H., Oyama, T. & Tozawa, Y. Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in *Cyanobacterium Synechococcus* elongatus PCC 7942. *Plant Cell Physiol.* **54,** 1724–35 (2013).
190. Ohbayashi, R., Akai, H., Yoshikawa, H., Hess, W. R. & Watanabe, S. A tightly inducible riboswitch system in *Synechocystis* sp. PCC 6803. *J. Gen. Appl.*

Microbiol. **62,** 154–159 (2016).

191. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39,** 1131–41 (2011).
192. Codon, D. N. *et al.* Accessibility of the Shine-Dalgarno Sequence Article Accessibility of the Shine-Dalgarno Sequence. *Mol. Cell* **70,** 894–905.e5 (2018).
193. Mustoe, A. M. *et al.* Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **173,** 181–195.e18 (2018).
194. Dorak, M. T. & Dorak, M. T. *Real-Time PCR.* (Taylor and Francis, 2006).
195. Rocha, D. J. P., Santos, C. S. & Pacheco, L. G. C. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek* **108,** 685–693 (2015).
196. Kirk, D. G., Palonen, E., Korkeala, H. & Lindström, M. Evaluation of normalization reference genes for RT-qPCR analysis of *spo0A* and four sporulation sigma factor genes in *Clostridium botulinum* Group I strain ATCC 3502. *Anaerobe* **26,** 14–19 (2014).
197. Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* **29,** 23–39 (2002).
198. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55,** 611–22 (2009).
199. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J. Am. Chem. Soc.* **126,** 13247–54 (2004).
200. Morra, R. *et al.* Dual transcriptional- Translational cascade permits cellular level tuneable expression control. *Nucleic Acids Res.* **44,** e21 (2016).
201. Horbal, L. & Luzhetskyy, A. Dual control system - A novel scaffolding architecture of an inducible regulatory device for the precise regulation of gene expression. *Metab. Eng.* **37,** 11–23 (2016).
202. Westbrook, A. M. & Lucks, J. B. Achieving large dynamic range control of gene expression with a compact RNA transcription-translation regulator. *Nucleic Acids Res.* **45,** 5614–24 (2017).
203. Ng, Y. K. *et al.* Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* **8,** e56051 (2013).
204. Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. *Appl. Environ. Microbiol.* **78,** 4683–4690 (2012).
205. Ghim, S.-Y., Nielsen, P. & Neuhard, J. Molecular characterization of pyrimidine biosynthesis genes from the thermophile *Bacillus caldolyticus*. *Microbiology* **140,** 479–491 (1994).
206. *Registry of standard biological parts. http://www.parts.igem.org/.*
207. Willson, B. J. *et al.* Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* **9,** 109 (2016).
208. Burbulys, D., Trach, K. A. & Hoch, J. A. Initiation of Sporulation in *B. subtilis* Is Controlled by a Multicomponent Phosphorelay. *Cell* **64,** (1991).
209. Paredes, C. J., Alsaker, K. V & Papoutsakis, E. T. A comparative genomic view of clostridial sporulation and physiology. *Nat. Rev. Microbiol.* **3,** 969–78 (2005).
210. Fujita, M., Gonzalez-Pastor, J. E. & Losick, R. High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. *J. Bacteriol.* **187,** 1357–1368 (2005).
211. Vishnoi, M. *et al.* Triggering sporulation in *Bacillus subtilis* with artificial two-component systems reveals the importance of proper Spo0A activation

dynamics. *Mol. Microbiol.* **90,** 181–194 (2013).

212. Molle, V. *et al.* The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* **50,** 1683–1701 (2003).
213. Saujet, L. *et al.* Genome-Wide Analysis of Cell Type-Specific Gene Transcription during Spore Formation in *Clostridium difficile*. *PLoS Genet.* **9,** 1–10 (2013).
214. Steil, L., Serrano, M., Henriques, A. O. & Völker, U. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology* **151,** 399–420 (2005).
215. Dürre, P. & Hollergschwandner, C. Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* **10,** 69–74 (2004).
216. Yasugi, M. *et al.* Transcriptional Profile during Deoxycholate-Induced Sporulation in a *Clostridium perfringens* Isolate Causing Foodborne Illness. *Appl. Environ. Microbiol.* **82,** 2929–2942 (2016).
217. Harris, L. M., Welker, N. E. & Papoutsakis, E. T. Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* **184,** 3586–3597 (2002).
218. Vohra, P. & Poxton, I. R. Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. *Microbiology* **157,** 1343–1353 (2011).
219. Stephens, C. Bacterial sporulation: A question of commitment? *Curr. Biol.* **8,** R45–R48 (1998).
220. Dembek, M. *et al.* Inducible expression of *spo0A* as a universal tool for studying sporulation in *Clostridium difficile*. *Front. Microbiol.* **8,** 1–12 (2017).
221. Galperin, M. Y. *et al.* Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ. Microbiol.* **14,** 2870–90 (2012).
222. Heap, J. T. *et al.* The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* **80,** 49–55 (2010).
223. Nagaraju, S., Davies, N. K., Walker, D. J. F., Köpke, M. & Simpson, S. D. Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Biotechnol. Biofuels* **9,** 219 (2016).
224. Dembek, M. *et al.* High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *MBio* **6,** e02383–14 (2015).
225. Karow, M. L., Glaser, P. & Piggot, P. J. Identification of a gene, *spoIIR*, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2012–6 (1995).
226. Monod, M., Denoya, C. & Dubnau, D. Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. *J. Bacteriol.* **167,** 138–47 (1986).
227. Serra, C. R., Earl, A. M., Barbosa, T. M., Kolter, R. & Henriques, A. O. Sporulation during growth in a gut isolate of Bacillus subtilis. *J. Bacteriol.* **196,** 4184–96 (2014).
228. Garmendia, E., Brandis, G. & Hughes, D. Transcriptional Regulation Buffers Gene Dosage Effects on a Highly Expressed Operon in *Salmonella*. *MBio* **9,** e01446-18 (2018).
229. Yang, W.W., Crow-Willard, E. N. & Ponce, A. Production and characterization of pure *Clostridium* spore suspensions. *J. Appl. Microbiol.* **106,** 27–33 (2009).
230. Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70,** 452–64 (2007).
231. Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional

Engineering in the Genus *Clostridium*. *Appl. Environ. Microbiol.* **82,** 6109–6119 (2016).
232. Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A. & Chou, C. P. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.* **6,** 25666 (2016).
233. Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M. & Lopes Ferreira, N. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J. Microbiol. Methods* **140,** 5–11 (2017).
234. Li, Q. *et al.* CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol. J.* **11,** 961–972 (2016).
235. Wang, Y. *et al.* Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J. Biotechnol.* **200,** 1–5 (2015).
236. Wang, Y. *et al.* Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable 'clean' Mutant Selection in *Clostridium beijerinckii* as an Example. *ACS Synth. Biol.* **5,** 721–732 (2016).
237. Wang, S. *et al.* Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* **24,** 1095–1099 (2018).
238. McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T. & Sorg, J. A. Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* **7,** 14672 (2017).
239. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11,** 399–402 (2014).
240. Xu, T. *et al.* Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Appl. Environ. Microbiol.* **81,** 4423–4431 (2015).
241. Hong, W., Zhang, J., Cui, G., Wang, L. & Wang, Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of *C. difficile* Infection. *ASC Synth. Biol.* **7,** 1588–1600 (2018).
242. Mougiakos, I., Bosma, E. F., Ganguly, J., van der Oost, J. & van Kranenburg, R. Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. *Curr. Opin. Biotechnol.* **50,** 146–157 (2018).
243. Peters, J. M. *et al.* Bacterial CRISPR: accomplishments and prospects. *Curr. Opin. Microbiol.* **27,** 121–6 (2015).
244. Selle, K. & Barrangou, R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Cell Press* **24,** 225–232 (2015).
245. Engstrom, M. D. & Pfleger, B. F. Transcription control engineering and applications in synthetic biology. *Synth. Syst. Biotechnol.* **2,** 176–191 (2017).
246. Sauer, M. Industrial production of acetone and butanol by fermentation-100 years later. *FEMS Microbiol. Lett.* **363,** (2016).
247. Bradbury, M. *et al.* Draft genome sequence of *Clostridium sporogenes* PA 3679, the common nontoxigenic surrogate for proteolytic *Clostridium botulinum*. *J. Bacteriol.* **194,** 1631–2 (2012).
248. World Health Organisation. Botulism. at <http://www.who.int/news-room/fact-sheets/detail/botulism> (2018).
249. Seo, S. O., Wang, Y., Lu, T., Jin, Y. S. & Blaschek, H. P. Characterization of a *Clostridium beijerinckii spo0A* mutant and its application for butyl butyrate production. *Biotechnol. Bioeng.* **114,** 106–112 (2017).
250. Wang, S., Dong, S., Wang, P., Tao, Y. & Wang, Y. Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **83,** e00233-17 (2017).
251. Zhang, Y. *et al.* A Universal Mariner Transposon System for Forward Genetic Studies in the Genus *Clostridium*. *PLoS One* **10,** e0122411 (2015).

Appendix
/ Supplementary Tables and Figures

List of supplementary tables

Table S1 Chemicals, kits and suppliers used in this study 207
Table S2 Growth media composition 209
Table S3 *E. coli* strains used in this study 210
Table S4 *Clostridium* strains used in this study 211
Table S5 List of plasmids used in this study 213
Table S6 List of primers used in this study 221

List of supplementary figures

Figure S1 Expression plasmids: parent vector (pMTL82251) and promoter-less vector (pMTL-IC001) 233
Figure S2 Expression plasmids containing the LAC system or the dual mechanisms RiboLac and Lactoswitch 234
Figure S3 Expression plasmids containing theophylline responsive riboswitches upstream of *catP* or *spo0A* 235
Figure S4 Plasmids used for in-frame deletions (or integrations) via Allelic-Exchange 236
Figure S5 Plasmids used for chromosomal integrations via ACE 237
Figure S6 RiboCas plasmids 238
Figure S7 Supplementary DNA sequences 239

Table S1 Chemicals, kits and suppliers used in this study

Material	Supplier
Acetyl-Coa, sodium salt	Sigma-Aldrich (A2056)
Agar bacteriological (Agar no. 1)	Oxoid (LP0011)
Agar bacteriological (Agar no. 2)	Oxoid (LP0012)
Agarose	Fisher Scientific (BP1356)
Antarctic Phosphatase	New England Biolabs (M0289S)
BHI Agar ready mix	Oxoid (CM1136)
BHI ready mix	Oxoid (CM1135)
BugBuster Master Mix	Novagen (71456)
Chloramphenicol	Alfa Aesar (B20841)
cOmplete™, Protease Inhibitor Cocktail	Roche (11836145001)
DNA Ladder (2-Log, 100 bp)	New England Biolabs
DreamTaq Green PCR Master Mix	Thermo Fisher (K1081)
DTNB (5,5' dithio(nitrobenzidene))	Sigma-Aldrich (D1830)
Erythromycin	Sigma-Aldrich (E5389)
Ethanol	Fisher Scientific (G/0650DF/17)
Expand™ High Fidelity PCR System	Roche (11732641001)
Fluoroorotic acid	Source Biosciences
FastRNA PRO™ Blue Kit	MP Biomedicals (116025050)
Gel Loading Dye, Purple (6X)	New England Biolabs (B7024S)
GenElute™ Bacterial Genomic DNA Kit	Sigma-Aldrich (NA2120)
Glucose	Sigma-Aldrich (G8270)
Glycerol	Sigma-Aldrich (G5516)
Glycine	Sigma-Aldrich (G8898)
High Pure PCR Product Purification Kit	Roche (11732668001)
High Pure RNA Isolation Kit	Roche (11828665001)
Kanamycin	*Sigma-Aldrich (K1377)*
KOD Hot Start Master Mix	Novagen (71842)
L-Cysteine Hydrochloride Monohydrate	Fisher Bioreagents (BP376)
Lactose	Sigma-Aldrich (17814)
LB ready mix	Fisher Bioreagents (BP1426)
Lysozyme	Sigma-Aldrich (L6876)
Magnesium chloride	Sigma-Aldrich (M8266)
Monarch DNA Gel Extraction Kit	New England Biolabs (T1020S)

Table S1 Chemicals, kits and suppliers used in this study; continued from previous page

Material	Supplier
Monarch PCR & DNA Cleanup Kit	New England Biolabs (T1030S)
Monarch Plasmid Miniprep Kit	New England Biolabs (T1010S)
MOPS (3-(N-morpholino)propanesulfonic acid)	Sigma-Aldrich (M1254)
Omniscript RT Kit	Qiagen (205111)
Peptone	Novagen (71280)
Phosphate-buffered saline tablets	Oxoid (BR0014G)
Phenol:Chloroform:Isoamyl alcohol	Sigma-Aldrich (P3803)
Pierce™ BCA Protein Assay Kit	Thermo Fisher (23225)
Potassium phosphate dibasic	Sigma-Aldrich (P3786)
Potassium phosphate monobasic	Sigma-Aldrich (P0662)
Protector RNase inhibitor	Roche (3335399001)
Proteinase K	Sigma-Aldrich (P6556)
Random Primer 6	New England Biolabs (S1230S)
RCM agar ready mix	Oxoid (CM0151)
Restriction Enzymes	New England Biolabs
RNAprotect Bacteria Reagent	Qiagen (76506)
RQ1 RNase-Free DNase Kit	Promega (M6101)
SYBR™ Green PCR Master Mix	Thermo Fisher (4309155)
SYBR™ Safe DNA gel stain	Invitrogen (S33102)
Sodium acetate	Sigma-Aldrich (W302406)
Sodium chloride	Sigma-Aldrich (746398)
Sodium dodecyl sulfate	Sigma-Aldrich (436143)
Sodium thioglycolate	Sigma-Aldrich (BP376)
Sucrose	Sigma-Aldrich (S8501)
TAE buffer (10 x)	EMD Millipore (106023)
T4 DNA ligase	Promega (M1801)
T4 Polynucleotide Kinase	New England Biolabs (M0201S)
Theophylline	Sigma-Aldrich (T1633)
Thiamphenicol	MP Biomedicals (QR13918)
Trizma base	Sigma Aldrich (93352)
Tryptone	Bacto (211705)
Yeast extract	Oxoid (LP0021)

Table S2 Growth media composition

Medium	Composition	Amount (g/l)
SOC	Tryptone	20
	Yeast Extract	5
	NaCl	0.584
	KCl	0.186
	$MgCl_2$	0.952
	$MgSO_4$	1.204
	Glucose	3.603
LB (lysogeny broth)	Tryptone	10
	Yeast Extract	5
	NaCl	10
LB - agar	Tryptone	10
	Yeast Extract	5
	NaCl	10
	Bacteriological agar no. 3	15
TYG	Tryptone	30
	Yeast Extract	20
	Sodium Thyoglycolate	1
TYG - agar	Tryptone	30
	Yeast Extract	20
	Sodium Thyoglycolate	1
	Bacteriological agar no. 2	15
Amino Acid Complete and defined Medium (MACC)	[1]	
2xYPG	Peptone	16
	Yeast Extract	10
	NaCl	5
	Glucose	6%
	pH 6.2	
RCM - agar	Yeast extract	3
	Lab-Lemco powder	10
	Peptone	10
	Glucose	5
	Soluble starch	1
	NaCl	5
	Sodium acetate	3
	Cysteine	0.5
	Agar	0.5
BHIS	Brain infusion solids	12.5
	Beef heart infusion solids	5
	Protease peptone	10
	Glucose	2
	NaCl	5
	Na_2HPO_4	2.5
BHIS - agar	Brain infusion solids	12.5
	Beef heart infusion solids	5
	Protease peptone	10
	Glucose	2
	NaCl	5
	Na_2HPO_4	2.5

Table S3 E. coli strains used in this study

Strain	Characteristics	Source
TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *araD*139 Δ(*ara-leu*)7697 *galU galK rpsL end*A1 *nupG*	Life Technologies
CA434	As TOP10 but bearing R702 conjugative plasmid	[2]
TOP10 pCR1	As TOP10 but bearing pCR1 methylation plasmid	[3]

Table S4 Clostridium strains used in this study

Strain	Description/Properties[a]	Source[b]
***C. sporogenes* NCIMB 10696**	Wild type isolate	CRG
***C. sporogenes pyrE*$^{-}$**	*pyrE* deficient mutant	[4]
***C. sporogenes*::LAC**	*C. sporogenes* with supplied LAC-*catP* via ACE	This study
***C. sporogenes*::RiboLac**	*C. sporogenes* with supplied RiboLac-*catP* via ACE	This study
C. sporogenes pyrE*$^{-}$ *Δspo0A	*pyrE* deficient mutant with *spo0A* deletion via AE	This study
C. sporogenes Δspo0A	*spo0A* mutant with *pyrE* repaired via ACE	This study
***C. sporogenes Δspo0A*^{COMP}**	*spo0A* mutant with supplied *spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{fdx4}-*E*-*spo0A	*spo0A* mutant with supplied P_{fdx4}-E-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{fdx4}*-*E*-*spo0A	*spo0A* mutant with supplied P_{fdx4}*-E-*spo0A* via ACE	This study
C. sporogenes Δspo0A*:: P_{fdx}-*E*-*spo0A	*spo0A* mutant with supplied P_{fdx}-E-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{fdx}-*G*-*spo0A	*spo0A* mutant with supplied P_{fdx}-G-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{fdx}*-*E*-*spo0A	*spo0A* mutant with supplied P_{fdx}*-E-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{fdx}*-*G*-*spo0A	*spo0A* mutant with supplied P_{fdx}*-G-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{araE}*-*E*-*spo0A	*spo0A* mutant with supplied P_{araE}*-E-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::P_{araE}*-*G*-*spo0A	*spo0A* mutant with supplied P_{araE}*-G-*spo0A* via ACE	This study

Table S4 *Clostridium* strains used in this study; continued from previous page

Strain	Description/Properties[a]	Source[b]
C. sporogenes Δspo0A*::*LAC-spo0A	*spo0A* mutant with supplied LAC-*spo0A* via ACE	This study
C. sporogenes Δspo0A*::RiboLac-*spo0A	*spo0A* mutant with supplied RiboLac-*spo0A* via ACE	This study
C. sporogenes ΔspoIIR	*spoIIR* mutant via RiboCas	This study
C. sporogenes pyrE⁻ ΔspoIIR	*pyrE* deficient mutant with *spoIIR* deletion via RiboCas	This study
C. sporogenes ΔspoIIR^{COMP}	*spoIIR* mutant with supplied *spoIIR* via ACE	This study
C. sporogenes*::*PspoIIR-G-spoIIR	*spoIIR* mutant with supplied P_{spoIIR}-G-*spoIIR* via AE	This study
C. sporogenes ΔspoIIE	*spoIIE* mutant via RiboCas	This study
C. sporogenes spoIIE⁻*::*ARC	*spoIIE* deficient mutant with insertion of the ARC via RiboCas	This study
C. pasterurianum DSM-525 H1	*C. pasteurianum* strain with increased transformation efficiency	CRG
C. pasterurianum ΔspoIIE	*spoIIE* mutant via RiboCas	This study
C. difficile 630	Wild type isolate	CRG
C. difficile 630 ΔspoIIE	*spoIIE* mutant via RiboCas	This study
***C. botulinum* ATCC 3502**	Wild type isolate	CRG
C. botulinum ΔspoIIE	*spoIIE* mutant via RiboCas	This study

[a]ACE (Allele-Coupled Exchange); AE (Allelic Exchange). [b]CRG (Clostridia Research Group)

Table S5 **List of plasmids used in this study**

Plasmid	Relevant characteristics*	Reference source**
	Modular vectors	
pMTL82251/pMTL82254	Erm^R; modular vector expression assays	[5]
pMTL83151	Cm^R; modular vector used as a chassis for RiboCas	[5]
	Chapter 3 / The LAC system from *C. perfringens* strain 13 (all vectors derived from pMTL82251 with sequences inserted via NotI and XhoI restriction sites)	
pMTL-HZ1	Erm^R; LAC-*catP*	Wang, H. CRG
pMTL-IC001	Erm^R; promoterless *catP* reporter vector	This study
pMTL-IC1201	Erm^R; P_{bgaR}-*catP*	This study
pMTL-ICMO1	Erm^R; LAC-*catP* with O1 deleted	This study
pMTL-ICMO2	Erm^R; LAC-*catP* with O2 deleted	This study
pMTL-ICMO1O2	Erm^R; LAC-*catP* with O1 and O2	This study
pMTL-ICMB	Erm^R; LAC-*catP* lacking *bgaR*	This study
pMTL-ICT1	Erm^R; LAC-*catP* with truncation 1	This study
pMTL-ICT2	Erm^R; LAC-*catP* with truncation 2	This study
pMTL-ICT3	Erm^R; LAC-*catP* with truncation 3	This study
pMTL-ICT4	Erm^R; LAC-*catP* with truncation 4	This study
pMTL-ICT5	Erm^R; LAC-*catP* with truncation 5	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pMTL-ICT6	Erm^R; LAC-*catP* with truncation 6	This study
pMTL-ICT7	Erm^R; LAC-*catP* with truncation 7	This study
pMTL-ICT8	Erm^R; LAC-*catP* with truncation 8	This study
pMTL-ICT9	Erm^R; LAC-*catP* with truncation 9	This study
pMTL-ICT10	Erm^R; LAC-*catP* with truncation 10	This study
pMTL-ICT11	Erm^R; LAC-*catP* with truncation 11	This study
pMTL-ICM1	Erm^R; LAC-*catP* with mutation 1	This study
pMTL-ICM2	Erm^R; LAC-*catP* with mutation 2	This study
pMTL-ICM3	Erm^R; LAC-*catP* with mutation 3	This study
pMTL-ICM4	Erm^R; LAC-*catP* with mutation 4	This study
pMTL-ICM5	Erm^R; LAC-*catP* with mutation 5	This study
pMTL-ICM6	Erm^R; LAC-*catP* with mutation 6	This study
pMTL-ICM7	Erm^R; LAC-*catP* with mutation 7	This study
pMTL-ICM8	Erm^R; LAC-*catP* with mutation 8	This study
pMTL-ICM9	Erm^R; LAC-*catP* with mutation 9	This study
pMTL-ICM10	Erm^R; LAC-*catP* with mutation 10	This study
pMTL-ICM11	Erm^R; LAC-*catP* with mutation 11	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
	Chapter 4 / Theophylline responsive riboswitches in *C. sporogenes* (all vectors derived from pMTL82251 with sequences inserted via NotI and XhoI restriction sites)	
pMTL-IC101	Erm^R; P_{fdx}-*catP*	This study
pMTL-IC111-D	Erm^R; P_{fdx}-D-*catP*	This study
pMTL-IC111-E	Erm^R; P_{fdx}-E-*catP*	This study
pMTL-IC111-E*	Erm^R; P_{fdx}-E*-*catP*	This study
pMTL-IC111-F	Erm^R; P_{fdx}-F-*catP*	This study
pMTL-IC111-G	Erm^R; P_{fdx}-G-*catP*	This study
pMTL-IC111-H	Erm^R; P_{fdx}-H-*catP*	This study
pMTL-IC111-I	Erm^R; P_{fdx}-I-*catP*	This study
pMTL-IC111-J	Erm^R; P_{fdx}-J-*catP*	This study
pMTL-IC121-E	Erm^R; P_{fdx}*-E-*catP*	This study
pMTL-IC121-G	Erm^R; P_{fdx}*-G-*catP*	This study
pMTL-IC121-H	Erm^R; P_{fdx}*-H-*catP*	This study
pMTL-IC201	Erm^R; P_{fdx4} -*catP*	This study
pMTL-IC211-E	Erm^R; P_{fdx4}-E-*catP*	This study
pMTL-IC211-G	Erm^R; P_{fdx4}-G-*catP*	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pMTL-IC211-H	Erm^R; P_{fdx4}-H-*catP*	This study
pMTL-IC221-E	Erm^R; P_{fdx4}*-E-*catP*	This study
pMTL-IC221-G	Erm^R; P_{fdx4}*-G-*catP*	This study
pMTL-IC221-H	Erm^R; P_{fdx4}*-H-*catP*	This study
	Chapter 5 / A dual mechanism to tightly control gene expression - the RiboLac (all expression vectors derived from pMTL82251 with sequences inserted via NotI and XhoI restriction sites)	
pMTL-ICE1	Erm^R; RiboLac with riboswitch E downstream of P_{bgaR}	This study
pMTL-ICG1	Erm^R; RiboLac with riboswitch G downstream of P_{bgaR}	This study
pMTL-ICH1	Erm^R; RiboLac with riboswitch H downstream of P_{bgaR}	This study
pMTL-ICE1*	Erm^R; RiboLac with riboswitch E downstream of P_{bgaR}*	This study
pMTL-ICG1*	Erm^R; RiboLac with riboswitch G downstream of P_{bgaR}*	This study
pMTL-ICH1*	Erm^R; RiboLac with riboswitch H downstream of P_{bgaR}*	This study
pMTL-JH29	Cm^R; integration vector for ACE in *C. sporogenes*	[4]
pMTL-JH29T	Cm^R; integration vector including a terminator downstream of *pyrE*	This study
pMTL-JH29T-L	Cm^R; integration vector for the integration of LAC-*catP* via ACE	This study
pMTL-JH29T-R	Cm^R; integration vector for the integration of RiboLac-*catP* via ACE	This study
pMTL-ICLG1*	Erm^R; Lactoswitch with riboswitch G downstream of P_{bgaL}*	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pMTL-ICLH1*	Erm^R; Lactoswitch with riboswitch H downstream of P_{bgaL}*	This study
	Chapter 6 / Development of a *C. sporogenes* conditionally sporulating strain (all expression vectors derived from pMTL82251 with sequences inserted via NotI and XhoI restriction sites)	
pMTL-CH14	Cm^R; deletion vector harbouring *pyrE* from *C. acetobutylicum* and pCD6 replicon	Humphreys, C. CRG
pMTL-ICD	Cm^R; deletion vector harbouring *pyrE* from *C. acetobutylicum* and pIM13 replicon	This study
pMTl-ICD-*spo0A*	Cm^R; deletion vector to delete *spo0A* via AE	This study
pMTL-JH29-*spo0A*	Cm^R; integration vector to complement *spo0A* via ACE	This study
pMTL-JH29	Cm^R; integration vector to repair *pyrE* via ACE	[4]
pMTL-IC321-E	Erm^R; P_{araE}*-E-*catP*	This study
pMTL-IC321-G	Erm^R; P_{araE}*-G-*catP*	This study
pMTL-IC321-H	Erm^R; P_{araE}*-H-*catP*	This study
pMTL-ICSPO-P_{fdx4}-E	Erm^R; P_{fdx4}-E-*spo0A*	This study
pMTL-ICSPO-P_{fdx4}-G	Erm^R; P_{fdx4}-G-*spo0A*	This study
pMTL-ICSPO-P_{fdx4}-H	Erm^R; P_{fdx4}-H-*spo0A*	This study
pMTL-ICSPO-P_{fdx4}*-E	Erm^R; P_{fdx4}*-E-*spo0A*	This study
pMTL-ICSPO-P_{fdx4}*-G	Erm^R; P_{fdx4}*-G-*spo0A*	This study
pMTL-ICSPO-P_{fdx4}*-H	Erm^R; P_{fdx4}*-H-*spo0A*	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pMTL-ICSPO-P*fdx*-E	Erm^R; P_{fdx}-E-*spo0A*	This study
pMTL-ICSPO-P*fdx*-G	Erm^R; P_{fdx}-G-*spo0A*	This study
pMTL-ICSPO-P*fdx*-H	Erm^R; P_{fdx}-H-*spo0A*	This study
pMTL-ICSPO-P*fdx-E**	Erm^R; P_{fdx}*-E-*spo0A*	This study
pMTL-ICSPO-P*fdx-G**	Erm^R; P_{fdx}*-G-*spo0A*	This study
pMTL-ICSPO-P*fdx-H**	Erm^R; P_{fdx}*-H-*spo0A*	This study
pMTL-ICSPO-L	Erm^R; LAC-*spo0A*	This study
pMTL-ICSPO-R	Erm^R; Ribolac-*spo0A* (RiboLac with riboswitch G downstream of P_{bgaR}*)	This study
pMTL-JH29T-P*fdx4*-E-*spo0A*	Cm^R; integration vector for P_{fdx4}-E-*spo0A* via ACE	This study
pMTL-JH29T-P*fdx4-E-*spo0A***	Cm^R; integration vector for P_{fdx4}*-E-*spo0A* via ACE	This study
pMTL-JH29T-P*fdx*-E-*spo0A*	Cm^R; integration vector for P_{fdx}-E-*spo0A* via ACE	This study
pMTL-JH29T-P*fdx*-G-*spo0A*	Cm^R; integration vector for P_{fdx}-G-*spo0A* via ACE	This study
pMTL-JH29T-P*fdx-E-*spo0A***	Cm^R; integration vector for P_{fdx}*-E-*spo0A* via ACE	This study
pMTL-JH29T-P*fdx-G-*spo0A***	Cm^R; integration vector for P_{fd4}*-G-*spo0A* via ACE	This study
pMTL-JH29T-P*araE-E-*spo0A***	Cm^R; integration vector for P_{araE}*-E-*spo0A* via ACE	This study
pMTL-JH29T-P*araE-G-*spo0A***	Cm^R; integration vector for P_{araE}*-G-*spo0A* via ACE	This study
pMTL-JH29T-L-*spo0A*	Cm^R; integration vector for LAC-*spo0A* via ACE	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pMTL-JH29T-R-*spo0A*	Cm^R; integration vector for RiboLac-*spo0A* via ACE	This study
pMTL-JH29-*spoIIR*	Cm^R; integration vector to complement *spoIIR* via ACE	This study
pMTL-ICD-P_{spoIIR}*-G-*spoIIR*	Cm^R; integration vector to integrate P_{spoIIR}-G-*spoIIR* in the *spoIIR* locus via AE	This study
Chapter 7 / A universal Cas9-based editing tools for the genus *Clostridium*: RiboCas		
pRECasC	Cm^R; P_{fdx}*-E-*cas9*	This study
pRECas1G-IIR	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIR*)	This study
pRECas1-IIR	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA-editing template (*spoIIR*)	This study
pRECas2-IIR	Cm^R; P_{fdx}*-E-*cas9*-J23119-sgRNA-editing template (*spoIIR*)	This study
pRGCasC	Cm^R; P_{fdx}*-G-*cas9*	This study
pRGCas1G-IIR	Cm^R; P_{fdx}*-G-*cas9*-P_{araE}-sgRNA (*spoIIR*)	This study
pRGCas1-IIR	Cm^R; P_{fdx}*-G-*cas9*-P_{araE}-sgRNA-editing template (*spoIIR*)	This study
pRGCas2-IIR	Cm^R; P_{fdx}*-G-*cas9*-J23119-sgRNA-editing template (*spoIIR*)	This study
pRECas1G-IIE	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIE*)	This study
pRECas1-IIE	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA-editing template (*spoIIE)*	This study
pRECas1-IIEin	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIE*)-editing template (ARC)	This study
pRECas1G-IIE-*Cpas*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIE*) – *C. pasteurianum*	This study

Table S5 List of plasmids used in this study; continued from previous page

Plasmid	Relevant characteristics*	Reference source**
pRECas1-IIE-*Cpas*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA-editing template (*spoIIE*) – *C. pasteurianum*	This study
pRECas1G-IIE-*Cdif*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIE*) – *C. difficile*	This study
pRECas1-IIE-*Cdif*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA-editing template (*spoIIE*) – *C. difficile*	This study
pRECas1G-IIE-*Cbot*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA (*spoIIE*) – *C. botulinum*	This study
pRECas1-IIE-*Cbot*	Cm^R; P_{fdx}*-E-*cas9*-P_{araE}-sgRNA-editing template (*spoIIE*) – *C. botulinum*	This study

*ACE (Allele-Coupled Exchange); AE (Allelic Exchange); Erm^R (Erythromycin resistance); Cm^R (Chloramphenicol resistance)
**CRG (Clostridia Research Group)

Table S6 List of primers used in this study

Primer	Sequence	Comment
	Chapter 3 The LAC system from *C. perfringens* strain 13	
IC103-f	TATAGGTCTCAATGGTATTTGAAAAAAT	
IC104-r	TATACTCGAGTTAACTATTTATCAATTCCTG	
IC105-f	CAGCGGCCGCATGGTATTTGAAAAAATTGATAAAAA	
IC301-f	TATAGCGGCCGCAAATATTCTTATATTTAGA	
IC302-r	TATAGGTCTCACCATTTCTACCTCCTA	
IC303-f	AAAGCAGTTTATTTAAAATAAATTCTTCTAAATATAAGAATATTTTAAAGAAATATCTTT	
IC304-r	TATTTTAAATAAACTGCTTTGCAAAGCTCTCATGT	
IC305-f	TTATAAGATTATATAGAACTTTTTAACTATTCTAAAAGATTAATTTACATA	
IC306-r	AAAAGTTCTATATAATCTTATAAATTTTAATAACTAATATATAAAGATA	
IC307-f	CACAGCGGCCGCTTCTACCTCCTAACCTATAAAATTAGCC	
IC308-r	CACACATATGACCCTCCCAATACATTTAAAATAATTATGTATTCATGAAA	
IC309-f	GGCCGGTCTCTTAATAACTAATATATAAAGATATTTCTTTAAAATATTCTTATATTTAGA	
IC310-r	CCGCGGCCGCTAATTTAGATATTAATTCTAAATTAAGTGAAATTAATATAGTAATTAT	
IC311-f	GGCCGGCCGGTCTCTATTATATAAGATTATAAGAAACATTATAACATATTTTAG	
IC312-f	GGCCGGCCGGTCTCTATTAGATTATAAGAAACATTATAACATATTTTAGAACTT	
IC313-f	GGCCGGCCGGTCTCTATTATAAGAAACATTATAACATATTTTAGAACTTTTTAA	
IC314-f	GGCCGGCCGGTCTCTATTAAACATTATAACATATTTTAGAACTTTTTAACTATT	
IC315-f	GGCCGGCCGGTCTCTATTATATAACATATTTTAGAACTTTTTAACTATTCTAAA	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC316-f	GGCCGGCCGGTCTCTATTAACATATTTTAGAACTTTTTAACTATTCTAAAAGATT	
IC317-f	GGCCGGCCGGTCTCTATTATTTTAGAACTTTTTAACTATTCTAAAAGATTAATTT	
IC318-f	GGCCGGCCGGTCTCTATTAGAACTTTTTAACTATTCTAAAAGATTAATTTACATA	
IC319-f	GGCCGGCCGGTCTCTATTATTTTAACTATTCTAAAAGATTAATTTACATATTAAC	
IC320-f	GGCCGGCCGGTCTCTATTAACTATTCTAAAAGATTAATTTACATATTAACATTTA	
IC321-f	GGCCGGCCGGTCTCTATTATCTAAAAGATTAATTTACATATTAACATTTAATTAT	
IC322-f	GGCCGGCCGGTCTCTATTATTTAATATAAGATTATAAGAAACATTATAACATATTTTAGAAC	
IC323-f	GGCCGGCCGGTCTCTATTAAAATTATATTGATTATAAGAAACATTATAACATATTTTAGAACTTTTTA	
IC324-f	GGCCGGCCGGTCTCTATTAAAATTTATAACTAATTAAGAAACATTATAACATATTTTAGAACTTTTTA ACTATT	
IC325-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTAATTCTAACATTATAACATATTTTAGAACTTTTT AACTATT	
IC326-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGATTGTATATAACATATTTTAGAACTTTTT AACTATT	
IC327-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATATATTCATATTTTAGAACTTTTT AACTATT	
IC328-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATTATAAGTATATTTAGAACTTTTT AACTATT	
IC329-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATTATAACATATAAATCAACTTTTT AACTATTCTAAAAG	
IC330-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATTATAACATATTTTAGTTGAATTT AACTATTCTAAAAGATTAATT	
IC331-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATTATAACATATTTTAGAACTTAAA TTCTATTCTAAAAGATTAATTTACA	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC332-f	GGCCGGCCGGTCTCTATTAAAATTTATAAGATTATAAGAAACATTATAACATATTTTAGAACTTTTTA AGATAACTAAAAGATTAATTTACA	
IC263-r	AAATGTGAAATCCGTCACATACT	5´RACE SP1
IC264-r	GGTATCATCTCATCATATATCCCCAATTCACCATCTT	5´RACE SP2
IC265-r	TATAGGCGCGCCTTTACAATCATTGCAATATAATAAAG	5´RACE SP3
M13-f	TGTAAAACGACGGCCAGT	Vector Scr/Seq primers
M13-r	CAGGAAACAGCTATGACC	
IC011-r	TTATATACTTGGTTTATTTACTTGATTATTTCTGTAA	
IC012-f	TATTGTATAGCTTGGTATCATCT	

Chapter 4 / Theophylline responsive riboswitches in *C. sporogenes*

IC101-f	ATATGCGGCCGCGTGTAGTAGCCTGTGAAATAAGTAA	
IC102-r	ATATGGTGTCACCATTAACACACCTCCTTAAAAATTACAC	
IC103-f	TATAGGTCTCAATGGTATTTGAAAAAAT	
IC104-r	TATACTCGAGTTAACTATTTATCAATTCCTG	
IC105-f	CAGCGGCCGCATGGTATTTGAAAAAATTGATAAAAA	
IC106-r	GGCGGTACCGATTTATATTTTACCATGTTTTTATTAATTT	
IC107-f	CCCGGTACCAATACGACTCACTATAGGTTCCGGTGATACCAGCATCGTCTTGATGCCCTTG	
IC108-r	CCGGTCTCACCATGTTGTTGCCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGCT	
IC109-r	CCGGTCTCACCATGTTGTTGCCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGCT	
IC110-r	CCGGTCTCACCATCTTGTTGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC111-r	CCGGTCTCACCATGTTGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATG	
IC112-r	CCGGTCTCACCATTAAGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC113-r	CCGGTCTCACCATTAACACACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC114-r	CCGGTCTCACCATCTTGTTGACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC115-r	CCGGTCTCACCATTGGTCGTACACTCCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC116-r	CCCGGTCTCTATTGAAAAATTACACAACTTTATACGATTTAT	
IC117-f	GGTCTCACAATACGACTCACTATAGGTGATACCAGCATCGTCTTGATGCCCTTG	
IC118-f	AAACTTAACTTCATGTTAAAAACTTGTTAAAATATAAATCGTATAAAGTTGTGTAATTT	
IC119-r	ATTTATATTTTAACAAGTTTTTAACATGAAGTTAAGTTTTTTTTATATAACTGTATAACC	
IC263-r	AAATGTGAAATCCGTCACATACT	5´RACE SP1
IC264-r	GGTATCATCTCATCATATATCCCCAATTCACCATCTT	5´RACE SP2
IC265-r	TATAGGCGCGCCTTTACAATCATTGCAATATAATAAAG	5´RACE SP3
IC257-f	CTACTTTGCAAGTGTACCTTG	RT-qPCR *catP*
IC258-r	ATCCCCAATTCACCATCTTG	
IC259-f	AGCGGTGAAATGCGTAGAGA	RT-qPCR *16Srrn*
IC260-r	GGCACAGGGGGAGTTGATAC	
IC261-f	GAAAGGGTATTCAAGCCACATC	RT-qPCR *gyrA*
IC262-r	CTATCGGTGTGGCTATGGGGGCT	
M13-f	TGTAAAACGACGGCCAGT	Vector Scr/Seq primers
M13-r	CAGGAAACAGCTATGACC	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
	Chapter 5 / A dual mechanism to tightly control gene expression – the RiboLac	
IC333-r	CCGGTCTCAATGCAAATATTGTGGAAAAAGTATGTTAAAGA	
IC334-r	CCGGTCTCTGCATCTTGTTGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC335-r	CCGGTCTCTGCATTAAGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC336-r	CCGGTCTCTGCATTAACACACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC337-f	GGCGGTACCGAGCTTTGCAAAGCAGTTTATCAACATAAA	
IC338-f	GGCGGTACCTATAAAATTAGCCAATTTTATAGTAGTCTT	
IC339-f	GCCCCCTGCAGGAGTAATATAAATGTTATAGATATATTAA	
IC340-r	CCGGTCTCAAACCTTATTTTTGTTCTCTACTACCTGGTTTTACAAA	
IC341-f	CCCGGGTCTCAGGTTTCCCGACTGGAAAGCGGGCAGTGACCAGGCATCAAATAAAACGAAAG GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCT	
IC342-r	CCCGCGGCCGCAGTTAATTATTGATTATAAACGCAGAAAGGCCCACCCGAAGGTGAGCCAGT GTGACTCTAGTAGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCA	
IC343-r	GGCCGGTCTCAATTGATACATTTAAAATAATTATGTATTCATGAAACATGATTGT	
M13-f	TGTAAAACGACGGCCAGT	Vector Scr/Seq primers
M13-r	CAGGAAACAGCTATGACC	
IC011-r	TTATATACTTGGTTTATTTACTTGATTATTTCTGTAA	
IC012-f	TATTGTATAGCTTGGTATCATCT	
IC001-f	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT	Src/Seq mutants via ACE
IC002-r	CAAAAGCACTTGAAAATGCATG	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
	Chapter 6 / Development of a conditionally sporulating *C. sporogenes* strain	
IC401-f	TATAGCTAGCGCAGCATTAGGCCATCCTATCA	
IC402-r	ATTAGCAAACTTCCAATATAAACTACTCTCCT	
IC403-f	TTATATGGAAGTTTGCTAATATGGATATATTAACAG	
IC404-r	TATAGGCGCGCCGTAAGCAAATATAATAGTAATATAAGATATG	
IC405-f	TATAGCGGCCGCTTGAATTTCTTTTGTCGAATAAATAATT	
IC406-r	TATAGCTAGCAAATAGACCCCCTGTATTTGG	
IC407-f	CCCGCGGCCGCTTTATATTTAGTCCCTTGCCTTGCCTA	
IC408-r	GGCCGGTCTCAATTGATACATTTAAAATAATTATGTATTCATGAAACATGATTGT	
IC409-f	CACAGGTCTCAATGGAAGAAACAAAGATCAATG	
IC410-r	TATACTCGAGCATAACAAAATAGACCCCCTG	
IC411-f	GGCCCATATGATGGAAGAAACAAAGATCAATG	
IC412-f	GGCCGCGGCCGCTTTTTAAAAATTGAATATAACTATAAAATTTGGTAATAA	
IC413-r	CCGGGCTAGCTTATTTTTTTAATTTATTAATCTCTT	
IC414-f	GGCCGACGTCTTTTTAAAAATTGAATATAACTATAAAATTTGGTAATAA	
IC415-r	GGCCGGTCTCGATTGGTTGATCTTATTTATTAGTTATTATTACCAAATTTTATAGTTATATTCAATT	
IC416-f	GGTCTCACAATACGACTCACTATAGGTGATACCAGCATCGTCTTGATGCCCTTG	
IC417-r	CCGGTCTCTGCATTAAGTTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGC	
IC418-f	GGCCGGTCTCAATGAAAAGAATATTAGTTTTTATGTTAAGTATT	
IC419-r	GGCCCTCGAGTTATTTTTTTAATTTATTAATCTCTT	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC420-f	GGCCGCTAGCAGGAAGTAAGTTTGACAAAATAAGAGTAACTTATG	
IC421-r	GGCCGACGTCTTATTTTATTGCTCCTATTCTTCTAACTTTTG	
IC422-f	GGCCCTCGAGAAAAAACTCTGTATTGAAAATAATACAGAGTTTTTTTATAATAAAAATTT	
IC423-r	GGCCGGCGCGCCACAAACTTTCTCATTTAATAGATGAATTTGACCCAGATGTTATAATAT	
M13-f	TGTAAAACGACGGCCAGT	Scr/Seq pMTL-ICSPO series vectors
M13-r	CAGGAAACAGCTATGACC	
IC001-f	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT	Scr/Seq mutants via ACE
IC002-r	CAAAAGCACTTGAAAATGCATG	
IC003-f	CCTTTATAAACTGAGGTCGTGCAC	External primers *spo0A* locus. Scr/Seq SC and DC via AE
IC004-r	GGAAAAGAAATAACTAACTCCGAGG	
IC005-f	CAAGTTTATCGCTCTAATGAAC	Scr/Seq pMTL-ICD vectors, SC and plasmid loss
IC006-r	CTGGCCGTCGTTTTACAACGTCGTG	
IC007-f	GATGGAATCATGTGTATACGAAGAGG	Plasmid loss pMTL-JH29(T) in combination with IC008-r
IC008-r	CTTTTTGCTGCTTTTCC	
IC009-f	CAGAAAAATCTGAATTTCCTAACAACTGA	Scr/Seq pMTL-ICD-*spo0A*
IC010-r	AAGGCATGAGTGGAAGTCCTATAATA	
IC014-f	ATGGAAGAAACAAAGATCAATG	
IC011-r	TTATATACTTGGTTTATTTACTTGATTATTTCTGTAA	Scr/Seq vectors with *bgaR*
IC012-f	TATTGTATAGCTTGGTATCATCT	Scr/Seq vector with *catP*
IC013-f	TTCCCGTAGACCTAAAAATATAAACTCTG	Scr/Seq vectors with *spo0A*

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC015-f	ATGAAAAGAATATTAGTTTTTATGTTAAG	Scr/Seq pMTL-ICD-P$_{spoIIR}$-G-*spoIIR*
IC016-f	TCTAATAATATACAACTTATAAATA	
IC247-f	TTTCCGTTCTAGGAGTGGATGCGGGAAAGGA	Scr/Seq *spoIIR* integrants via AE
IC248-r	TGACCATTCAGAGGATGATGAAGGTTTAGCC	

	RiboCas
IC201-f	ATATATATGCGGCCGCTGTATCCATATGACCATGATTACGAATTCTCAGTCACCTCCTAGCTGACT
IC202-r	GCTATGGTCTCCATGGATAAGAAATACTCAATAGGCTTAG
IC203-f	GGCCTCTAGAGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAAAGAAGTAAGTGTTATATATGATG
IC204-f	GGCCGGCGCGCCGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAAAGAAGTAAGTGTT
IC205-f	ATATATTCTAGATTTATATTTAGTCCCTTGCC
IC206-r	GGCCGGCGCGCCATTGACGTCTATGTCGACGAAAACTCCTCCTTAAGA
IC207-f	GTATCATCTAGAGTAAAACGACGGCCAGTTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT
IC208-r	GTATCAGGCGCGCCATTGACGTCTATACTAGTATTATACCTAGGACTGAGCTAGCTGTCAAACTGG
IC209-r	GGCCCCGGGACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGC ACCGACTCGGTGCCACTTTTTCAAGTTG

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
	***C. sporogenes – spoIIR* deletion**	
IC211-f	GGCCGACGTCTCATTAGATGCATATTCAATGCAGGATAGTATAG	LHR
IC212-r	ATTTTCAATACAGAGGTTGATCTTATTTATTAGTTATTATTACCAAATTTTATAGTTATA	
IC213-f	ATAAATAAGATCAACCTCTGTATTGAAAATAATACAGAGTTTTTTTATAATAAAA	RHR
IC214-r	AAGGCGGCGCGCCATTATGAACTACAAACTTTCTCATTTAATAGATGAATTTGACC	
IC215-f	CCGGGTCGACGGTAATATAACATTGCCACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	sgRNA-f-SalI
IC216-f	GGCCACTAGTGGTAATATAACATTGCCACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	sgRNA-f-SpeI
IC247-f	TTTCCGTTCTAGGAGTGGATGCGGGAAAGGA	Screening primers
IC248-r	TGACCATTCAGAGGATGATGAAGGTTTAGCC	
	***C. sporogenes – spoIIE* deletion**	
IC217-f	GGCCGGCGCGCCGCTGGAGTGGCGGAACTGGCAGACGCACAGGACTTAAAATCCTGCGGGCCTAAC AG	LHR
IC218-r	TATATTCTCTTTCTATATTTTAATATAATTGCATAATAATTAATCACCCCCATTAAGTTACTTCTCAAT TATATCA	
IC219-f	ACTTAATGGGGGTGATTAATTATTATGCAATTATATTAAAATATAGAAAGAGAATATAAAGTTAAGT AGTAAATACTTA	RHR
IC220-r	GGCCGACGTCTCATCATCAATATTTTTCACAATAATTTTCTATATCTTCTCTTAATACATTTATTAAAG GTCTTATATATACA	
IC221-f	GGCCGTCGACTTAATTATAGGATATGTAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	sgRNA

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC249-f	ATATAATTATTTATGTCGTCAGCAATGAGAC	Screening primers
IC250-r	CTCTCTTATCTATTATTACCCTTTGTGTT	
	***C. sporogenes* – ARC integration**	
IC222-f	GGCCGACGTCATGCAATATGGCGCAGAACTTTTACCATACCA	LHR
IC223-r	GGCCGCTCTTCTAAGTCATCATCATATCATAGTAATAGATATTAAATTCATTAAATTT	
IC224-f	GGCCGCTCTTCCCTTTAAGGAGGTGTGTTACATATGAACAAAAATATAAAATATTCTCAAAACTTTTTAACGAGTGAAAAAGTACTCAACCAAATA	*ermR*
IC225-r	GTACCGAGCTCGAATTCGTAATTTATTATTATTTCCTCCCGTTAAATAATAGATAACTATTAAAAATAGACAATACTTG	
IC226-f	TCTATTATTTAACGGGAGGAAATAATAATAAATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACGTCACGCGTCC	*lacZα* MCS
IC227-r	AACACACCTCCTTAAATCACTGCCCGCTTTCCAGTCGGGAAACCCTAGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGC	
IC228-f	GGTTTCCCGACTGGAAAGCGGGCAGTGATTTAAGGAGGTGTGTTACATATGTTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAAACTCG	*gusA*
IC229-r	GCTCTTCTCATTGTTTGCCTCCCTGCTGCGGTTTTTCACCGAAGTTCATGC	
IC230-f	GGCCGCTCTTCCATGAGGGAAGTACTAGTGCTTCAGCTATCGGTGTGGCT	RHR
IC231-r	GGCCGGCGCGCCTTTCATTTATTATATAATTATTTACTATATCTTC	
	***C. pasteurianum* – *spoIIE* deletion**	
IC237-f	GGCCGGCGCGCCGTTACGGGAATAGAAGAATCAAATTTGGGCTTTAAGGGTGTTTTAGAAGGAATACAAGGTGATG	LHR
IC238-r	ATACACATAATCTTGTAAACCTTATTTTGTTTGCATAAAAAAACCTCCTCTTTTGTTTATATTAAGAATTGTAAC	

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC234-f	GTTTATAGGTTGGGAGATGTTTGTATGCAATTGTATTAAAATGTTGTTTATTAATTGAATGAGAATTTATAA	RHR
IC235-r	GGCCGACGTCTCATCATCATGGATTAATGTGTTCCTTCTTACAATAGCTTTCTATTTCTTCTCTTCCTATTTTTATT	
IC236-f	GGCCGTCGACGGAAACTCCTAAATATCACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTT	sgRNA
IC253-f	ATGATATATAATAGTTTCTGTTGCAGTAAG	Screening primers
IC254-r	GTTATCTACATTTACTGTATCAGATAATC	
	***C. difficile – spoIIE* deletion**	
IC237-f	GGCCGGCGCGCCGTTACGGGAATAGAAGAATCAAATTTGGGCTTTAAGGGTGTTTTAGAAGGAATACAAGGTGATG	LHR
IC238-r	ATACACATAATCTTGTAAACCTTATTTTGTTTGCATAAAAAAACCTCCTCTTTTGTTTATATTAAGAATTGTAAC	
IC239-f	ATAAACAAAAGAGGAGGTTTTTTTATGCAAACAAAATAAGGTTTACAAGATTATGTGTATAAGATGTGGATAACTTG	RHR
IC240-r	GGCCGACGTCTCACCTCATCAATCAAGTTTTTCTTTAATGTTTAAGGCAAAAGACAAATGTGTTGGCGATTTTGTTA	
IC241-f	GGCCGTCGACGGAATTTTAGATTTAAAACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTT	sgRNA
IC251-f	TTCTGATAAGTGTAATGATGAAAAGTGTG	Screening primers
IC252-r	CTAACTCTAAATCTGAACAAATCCCCATA	
	***C. botulinum – spoIIE* deletion**	
IC242-f	GTTCATGGCGCGCCCTTCTAAATAGTTATTGTCTTTAGAAACTATATTAG	LHR
IC243-r	GCTATGGTCTCCTTATACTAAAATATAGAAAGAGAATATAAAGTTAGGTAG	
IC244-f	GCTACTGGTCTCAATAATTGCATAATAATTAATCACCCCCATTAAG	RHR
IC245-r	GTAGCCGACGTCCAGCGATGAGACATAAACATAAAAC	
IC246-f	TTTTCGTCGACTTTTATAAAGCGAGGAACTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGC	sgRNA

Table S6 List of primers used in this study; continued from previous page

Primer	Sequence	Comment
IC255-f	GAAATAAAGTGTTGACATAGTAAAGTTTG	Screening primers
IC256-r	AAAGCTCTCTTATCTATTATTACCCTTTG	
	Plasmid screening and sequencing	
IC266-r	GAAACTTAATCATATGCGCTAAGG	
IC267-f	ATGGATAAGAAATACTCAATAGGCTTAG	
IC268-f	GCTTTGTCATTGGGTTTGAC	
IC269-f	GTCGATAAAGGTGCTTCAGC	
IC270-f	GAACATATTGCAAATTTAGCTGG	
IC271-f	CTGACTTCCGAAAAGATTTCC	
IC272-f	GAGTTAGAAAACGGTCGTAAACG	
IC273-r	CTGTTATGCCTTTTGACTATC	
IC274-f	ATATATTCTAGATTTATATTTAGTCCCTTGCC	

Scr (screening); Seq (sequencing); ACE (Allele-coupled exchange); AE (Allelic-exchange);
RHR (right homologous region); LHR (left homologous region)

Figure S1 Expression plasmids: parent vector (pMTL82251) and promoter-less vector (pMTL-IC001)

All expression plasmids used in this study derived from the pMTL82251 vector[5]. In all cases, derivative vectors incorporated the "cargo" between the restriction sites NotI and XhoI. The figure includes the screening/sequencing primers M13-r and M13-f.

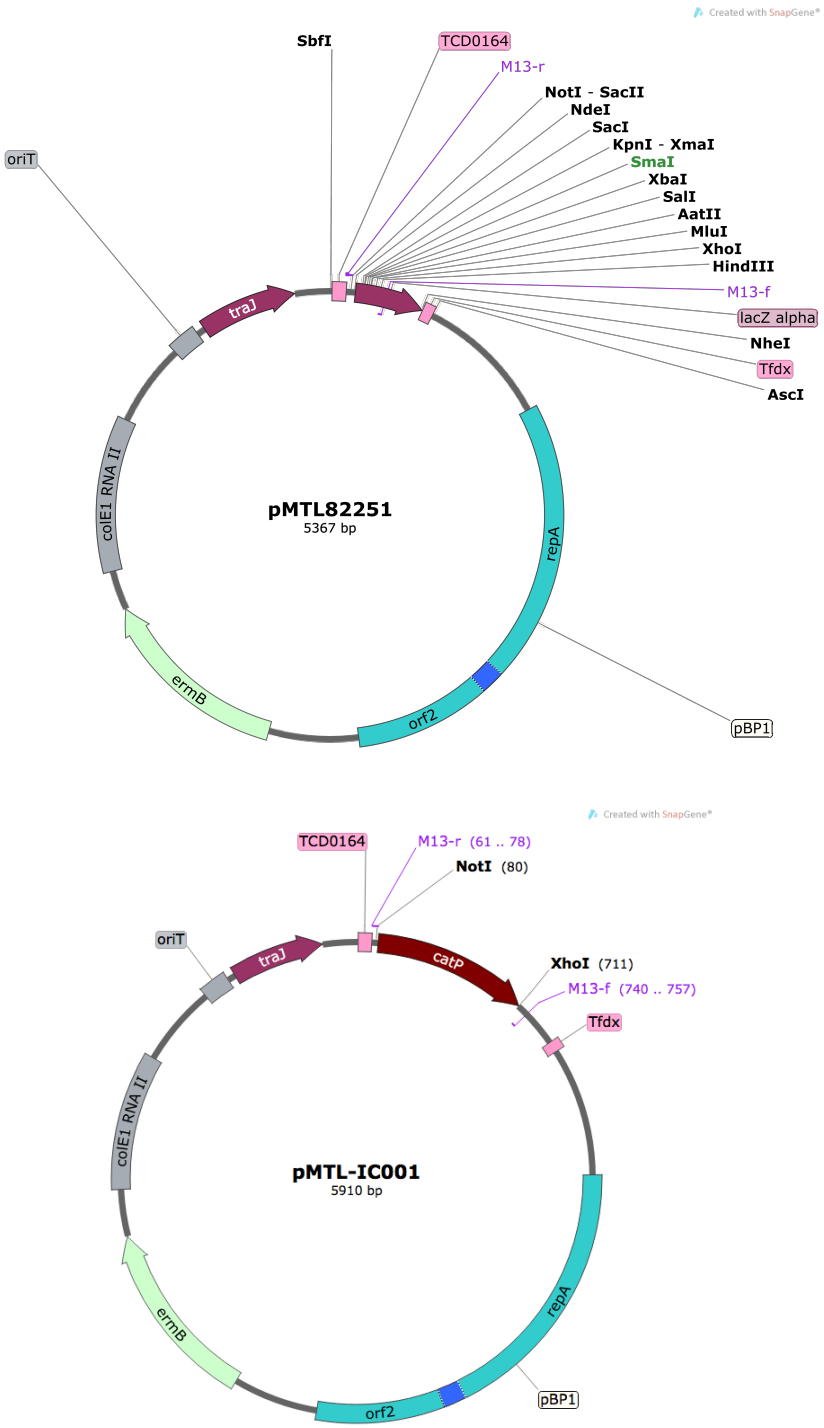
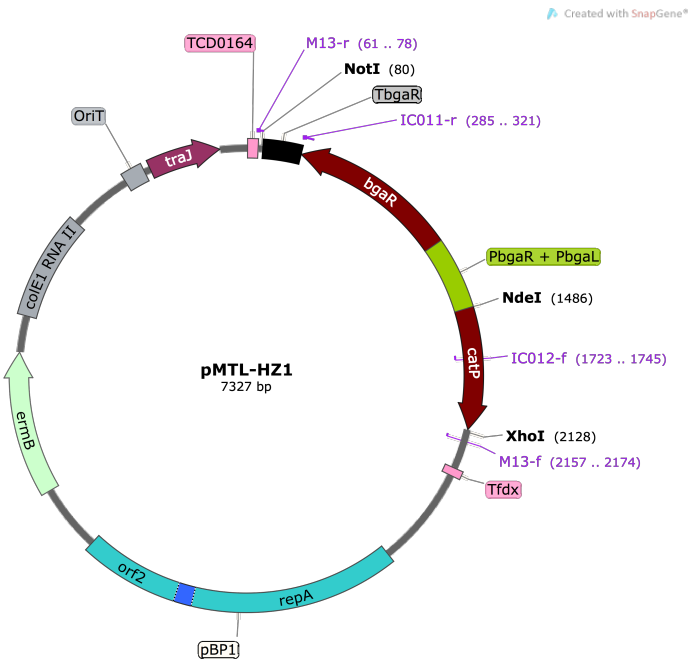


Figure S2 Expression plasmids containing the LAC system or the dual mechanisms RiboLac and Lactoswitch

The figure illustrates pMTL-HZ1 (LAC-*catP*) and pMTL-ICSPO-L (LAC-*spo0A*). Plasmids containing the RiboLac or the Lactoswitch only differ from pMTL-HZ1 in that such dual mechanisms included a riboswitch downstream of P_{bgaR} (or P_{bgaR}*) and P_{bgaL}, respectively. The figure includes the screening/sequencing primers M13-r, IC011-r, M13-f, IC012-f (*catP* specific) and IC014-f (*spo0A* specific).



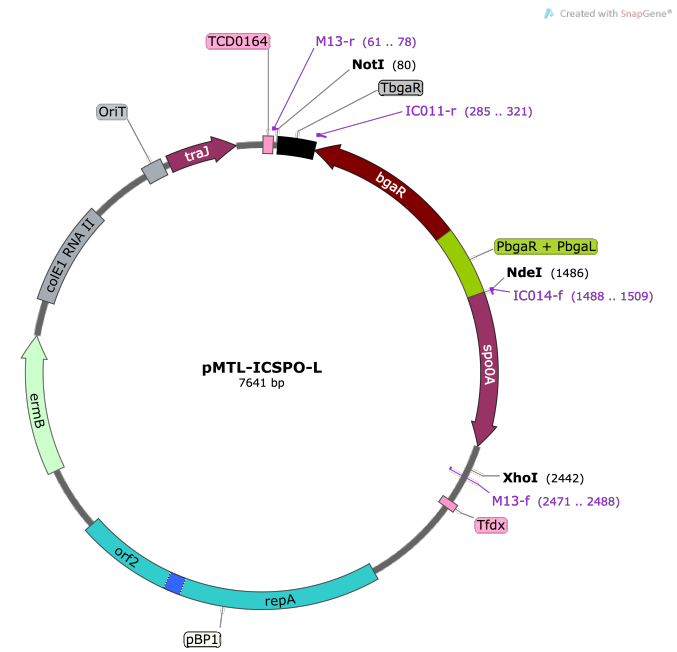


Figure S3 **Expression plasmids containing theophylline responsive riboswitches upstream of catP or spo0A**

As an example, the figure illustrates the expression plasmid pMTL-IC111-G where riboswitch-G, located downstream of P_{fdx}, regulates the expression of *catP*. In the pMTL-SPO-series vectors, *spo0A* is replaced by *catP*. The figure includes the screening/sequencing primers M13-r and M13-f.

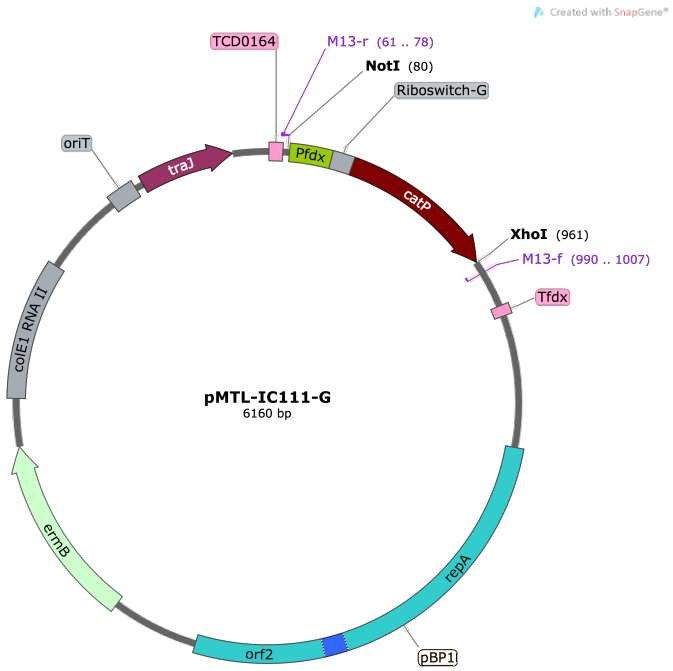
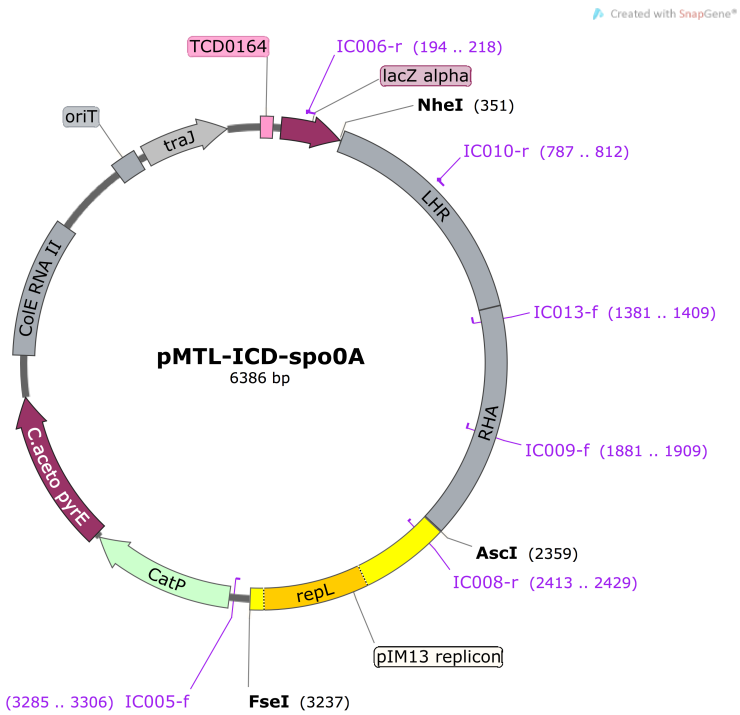


Figure S4 Plasmids used for in-frame deletions (or integrations) via Allelic-Exchange

Vector pMTL-ICD, which contains the pIM13 replicon, was used to create in-frame deletions (or insertions) by Allelic-Exchange. In all cases, the knock-out (in) cluster was inserted into the backbone using the restriction sites NotI and NheI. The figure illustrates the deletion vector pMTL-ICD-*spo0A* and the integration vector pMTL-ICD-P$_{spoIIR}$-G-*spoIIR*. The figure includes the primers used for screening and sequencing. RHR and LHR are right and left homologous regions, respectively.



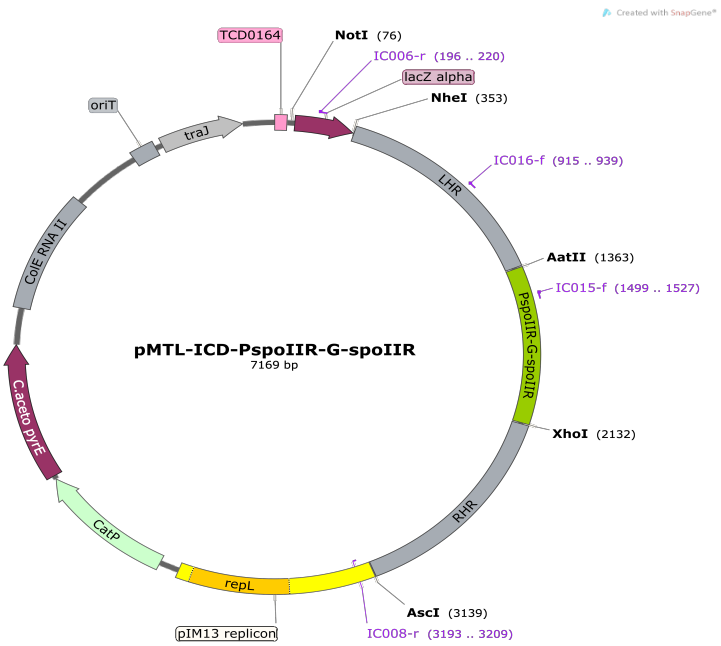
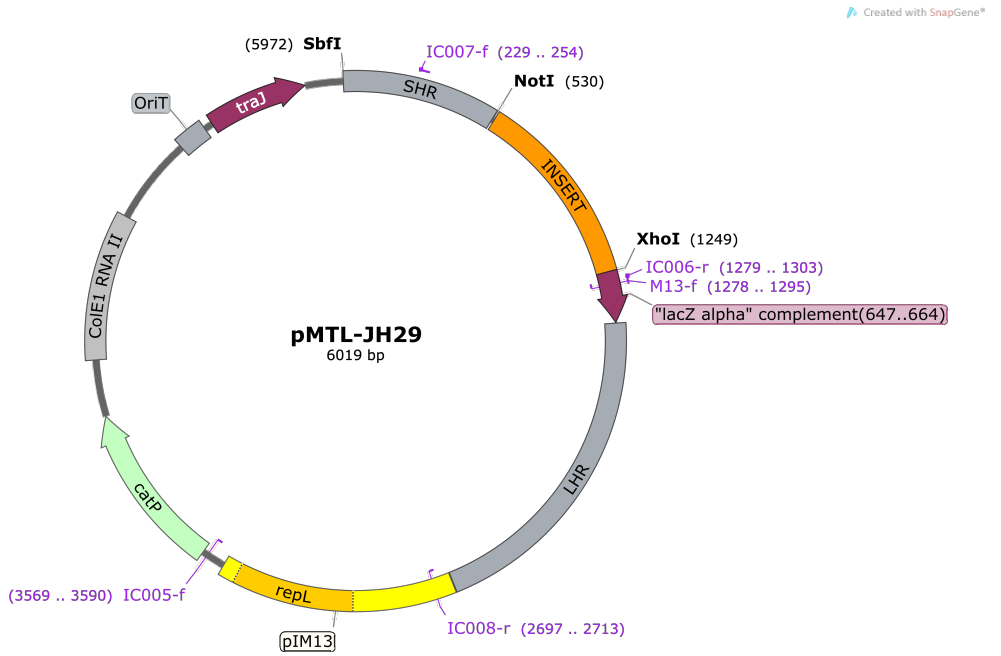


Figure S5 **Plasmids used fr chromosomal integrations via ACE**

Plasmid pMTL-JH29[4] was used to complement *spo0A* and *spoIIR*, as well as for *pyrE* restoration. pMTL-JH29T, which includes a terminator that is incorporated downstream of *pyrE* after chromosomal integration, was used to integrate all *catP/spo0A* expressing systems. The figure illustrates both plasmids with an example "insert", including the primers used for screening and sequencing. In all cases, the insert was cloned into the vector via NotI/XhoI restriction sites. SHR and LHR are short homologous region and long homologous region, respectively.



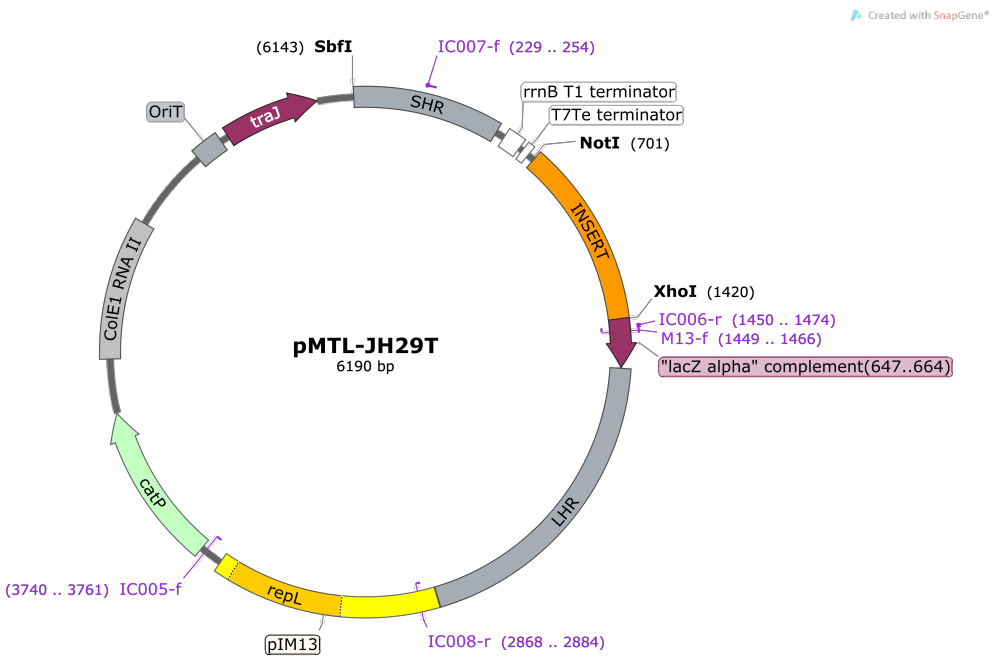


Figure S6 **RiboCas plasmids**

RiboCas plasmids included the specific sgRNA sequence between the restriction sites SalI and AatII, and the editing template between restriction sites AatII and AscI. The figure illustrates pRECas1 and pRECas2, where the sgRNA is driven by the promoters P_{araE} and J23119, respectively. RHR and LHR are right and left homologous regions, respectively. The figure includes the primers used for screening and sequencing.

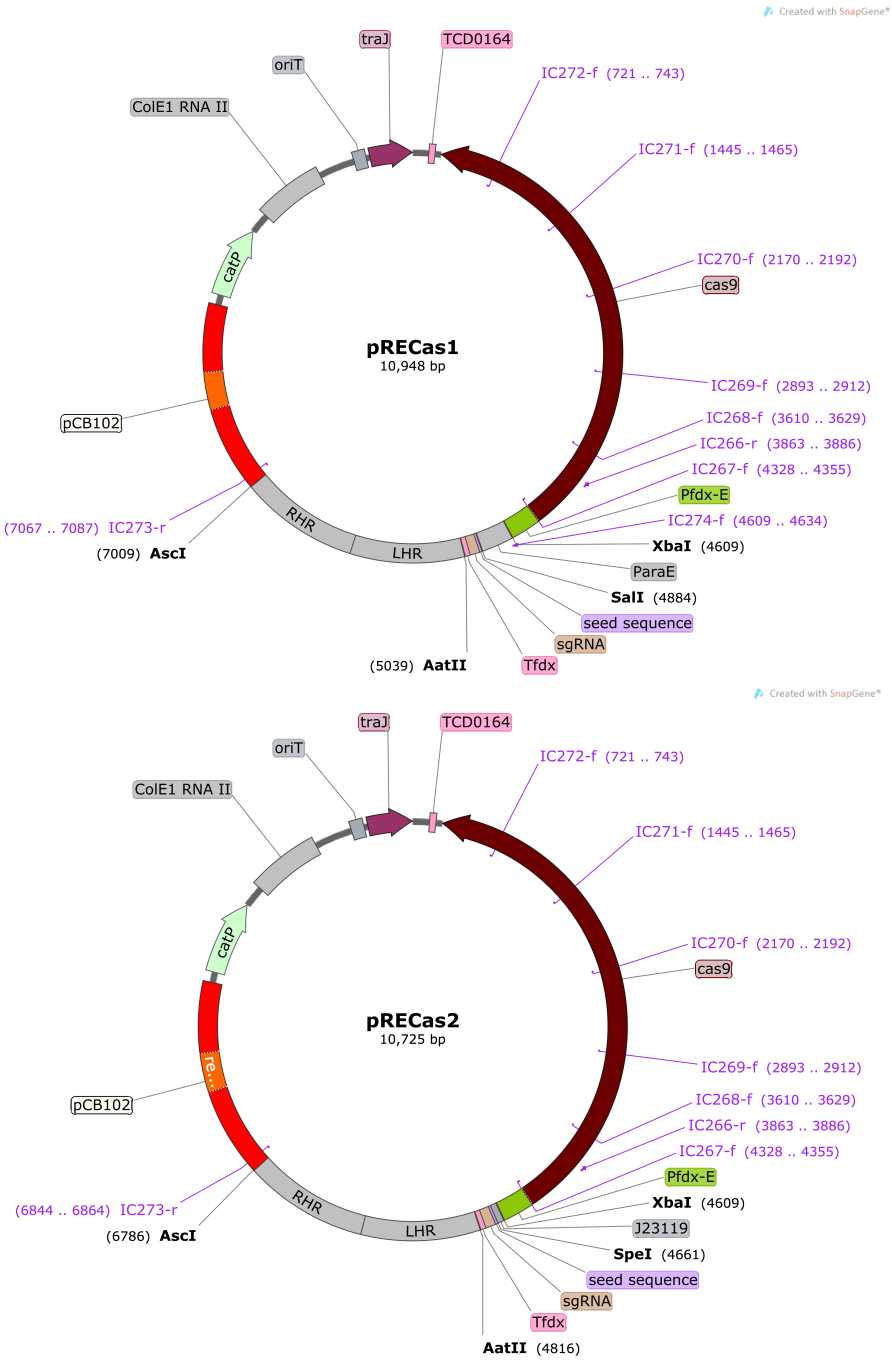


Figure S7 **Supplementary DNA sequences**

catP (translational start site is bold and underlined)

ATGGTATTTGAAAAAATTGATAAAAATAGTTGGAACAGAAAAGAGTATTTTGACC
ACTACTTTGCAAGTGTACCTTGTACCTACAGCATGACCGTTAAAGTGGATATCACA
CAAATAAAGGAAAAGGGAATGAAACTATATCCTGCAATGCTTTATTATATTGCAATG
ATTGTAAACCGCCATTCAGAGTTTAGGACGGCAATCAATCAAGATGGTGAATTGGG
GATATATGATGAGATGATACCAAGCTATACAATATTTCACAATGATACTGAAACATTT
TCCAGCCTTTGGACTGAGTGTAAGTCTGACTTTAAATCATTTTTAGCAGATTATGAA
AGTGATACGCAACGGTATGGAAACAATCATAGAATGGAAGGAAAGCCAAATGCTC
CGGAAAACATTTTTAATGTATCTATGATACCGTGGTCAACCTTCGATGGCTTTAATC
TGAATTTGCAGAAAGGATATGATTATTTGATTCCTATTTTTACTATGGGGAAATATTA
TAAAGAAGATAACAAAATTATACTTCCTTTGGCAATTCAAGTTCATCACGCAGTAT
GTGACGGATTTCACATTTGCCGTTTTGTAAACGAATTGCAGGAATTGATAAATAGTT
AA

P_{araE} (SD is bold and underlined)

TTTATATTTAGTCCCTTGCCTTGCCTACAAGGGATTTCCTATTCCTTTCATTTACAAT
TCATACGTATAAAATCCAAATTTTTCTTGACATTTATACACATAAATATTATGATTTAT
ATAGGTAATCGCTTTCATAAAATATATTACCCTTAGGAAATCAAATGATTATAAGTCA
TATATGAAAACGTTATATATAATTGATATGTTTACATTTGTAACTTAGATTTCTCTTTG
ATTTCCACATATATAAATCTT**AAGGAGG**AGTTTTC

P_{araE}*

TTTATATTTAGTCCCTTGCCTTGCCTACAAGGGATTTCCTATTCCTTTCATTTACAAT
TCATACGTATAAAATCCAAATTTTTCTTGACATTTATACACATAAATATTATGATTTAT
ATAGGTAATCGCTTTCATAAAATATATTACCCTTAGGAAATCAAATGATTATAAGTCA
TATATGAAAACGTTATATATAATTGATATGTTTACATTTGTAACTTAGATTTCTCTTTG
ATTTCCACATATATAAATCTT

Double terminator *rrn*T1 and T7Te, BBa_B0015

CAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTT
GTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCT
GCGTTTATAATCAATAATTAACT

P_{spoIIR} (SD is bold and underlined)

TTTTTAAAAATTGAATATAACTATAAAATTTGGTAATAATAACTAATAAATAAGATCA
AC**GGGGGG**AATATAT

P_{spoIIR}*

TTTTTAAAAATTGAATATAACTATAAAATTTGGTAATAATAACTAATAAATAAGATCA
AC

Appendix bibliography

1. Lovitt, R. W., Kell, D. B. & Morris, J. G. The physiology of *Clostridium sporogenes* NCIB 8053 growing in defined media. *J. Appl. Bacteriol.* **62,** 81–92 (1987).
2. Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from Escherichia coli to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* **46,** (2002).
3. Schwarz, K. M. *et al.* Towards improved butanol production through targeted genetic modification of *Clostridium pasteurianum*. *Metab. Eng.* (2017).
4. Heap, J. T. *et al.* Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res.* **40,** e59 (2012).
5. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78,** 79–85 (2009).