Discovery and characterisation of broad-host-range transcriptional terminators for use in solventogenic Clostridia

Ivan Slavkov Gyulev

Doctor of Philosophy

University of York

Biology

September 2019

Abstract

The bacterial butanol hyper-producer *Clostridium saccharoperbutylacetonicum* is of interest to industrial biotechnology and efforts are underway to engineer strains of this species and related solventogenic Clostridia species for traits such as improved tolerance and production capabilities. Recent synthetic biology engineering efforts in model organisms have been underpinned by the availability of genetic tools such as characterized biological parts. Intrinsic transcriptional terminators are crucial to the normal functioning of both natural and synthetic genetic systems but have not been thoroughly investigated in *C. saccharoperbutylacetonicum*. In order to study terminators in this species, two pairs of matched reporters of gene expression were explored and tested – two enzymatic (GusA and LacZ) and two fluorescent (mCherry and phiLOV2.1Opt), with the dual enzymatic system emerging as the most robust for medium-throughput use. Computational terminator prediction approaches and a terminator strength prediction model developed were extensively utilized and evaluated to construct a library of *Clostridium* terminators, which was combined with a library of known terminators used in other model systems. The broad-host-range dual enzymatic reporter vectors were used to assess the strength of these terminators in two model organism bacterial species - *Escherichia coli* and *Bacillus subtilis* - as well as in *C. saccharoperbutylacetonicum.* Terminators that function effectively in *C. saccharoperbutylacetonicum* were discovered; some terminators exhibited species-specific efficiency while others had broad-host-range activity. Together with the low predictive power of the strength prediction model, these observations suggest that important determinants of terminator activity are still to be determined. The dual reporter system has the potential for use in the screening of larger libraries of terminators to investigate the underlying mechanistic features for future applications in synthetic biology.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my primary academic supervisor Prof. Gavin Thomas for his help, guidance and support throughout this work and for the opportunity to conduct research under his supervision. I would also like to thank my other academic supervisor Prof James Chong and my industrial supervisor Dr Liz Jenkinson for their help and guidance. I would also like to thank my thesis advisory panel – Dr Christoph Baumann and Prof Maggie Smith for their guidance and encouragement.

I would also like to thank all of the past and present laboratory members of the Thomas, Chong and Green Biologics Laboratories and L1 Laboratories that I have had the pleasure to interact with over the years for their support and help.

I would also like to thank my family for their continued love and support. And to all my friends, thank you for your support.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. Part of this work was published as review article on which the author of this thesis was a joint first author. All sources are acknowledged as References.

Chapter I

Introduction

1. Introduction

1.1 Clostridia in industrial biotechnology

While microbial fermentation of carbohydrates to ethanol and lactic acid for the production of beverages and food processing has been used by humans since antiquity, the history of industrial biotechnology stretches back to the late 19th and early 20th centuries, where microbes were identified and then utilised specifically for the production of small molecule chemicals. One of the early examples relates to the production of acetone by bacteria from the genus *Clostridium* [1]. The biological production of acetone was used at the onset of WWI for chordite production, the process was developed by Chaim Weizmann in Manchester, UK[1].

These Gram-positive, spore forming, obligately anaerobic organisms have remarkable capabilities to produce a range of solvents that are useful to man[2]. In addition to acetone, these processes produced butanol and ethanol, leading to the fermentation being given the name 'ABE' for 'acetone-butanol-ethanol'[3].

The *Clostridium* cell wall contains peptidoglycan, in *C. acetobutylicum* it stains Gram-positive during exponential phase but during stationary phase it stains Gram-negative[4]. The membrane undergoes changes during ABE production – a decrease in the unsaturated/saturated fatty-acid ratio as well as changes in the phospholipid composition have been reported during solvenogenesis[4]. Changes in the membrane composition are expected as butanol is known to cause toxicity by disrupting the cell membrane[5]. Active efflux of solvents has not been identified to date in *Clostridium* species – the fermentation products are thought to exit the cell by diffusion.

Since this early work of Weizmann and colleagues, and driven by the demands of industry, many different ABE-producing clostridial isolates were identified; those strains that were used in industry are primarily represented by four species[6,7], namely *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharobutylicum*, and *Clostridium saccharoperbutylacetonicum*, although ABE fermentation has been observed in other clostridia[8–10] . Furthermore, not all butanol-producing clostridia produce the canonical trio of solvent products; for example, *Clostridium pasteurianum* produces butanol and 1,3-propanediol (instead of acetone) from glycerol[11]. While the ABE process had until recently fallen out of favour due to competition from the petrochemical industry, the necessity of identifying alternative fuels due to increased carbon emissions from fossil fuel use has renewed interest in the production of butanol and acetone as a potential biofuel candidate and sustainable commodity chemical, respectively[12].

1.1 Solvent production across the Clostridia

The particular metabolic features of solventogenic Clostridia that distinguish them from many other bacteria are the biochemical pathways that result in high levels of solvent production. ABE-producing clostridia typically show a biphasic growth pattern, producing acetic and butyric acids during the early stages of growth, and then undergoing a metabolic 'switch', assimilating the produced acids and producing solvents[13]. During acidogenesis carbohydrates are oxidized to pyruvate through glycolysis – the process generates NADH and ATP. After pyruvate decarboxylation, acetyl-CoA is further metabolized to either acetate or butyrate, generating more ATP. Notable differences are that the butyrate pathway uses 2 more molecules of NADH than the acetate pathway (while generating the same amount of ATP) possibly helping to recycle the NAD^+ co-factor. If NAD^+ is depleted glycolysis cannot proceed. When the

concentration of acids increases and lowers the extracellular pH, the acids are re-assimilated and reduced to solvents (acetone, butanol and ethanol). This process temporarily lowers the toxicity of the environment (as solvent accumulation is also detrimental) and recycles NAD^+ co-factor as it consumes NADH.

A recent comparative genomics study of many industrial saccharolytic strains (those preferring sugar as a carbon source) revealed that the known solventogens fall within two sister clades: one exemplified by *C. acetobutylicum* and one by *C. beijerinckii*[14] (Figure 1.1). Interestingly, another comparative genomics paper which included more diverse species from the genus, but fewer industrial solventogens, supported the split, with the genus' type species *Clostridium butyricum* being more closely related to *C. beijerinckii* and the pathogen *Clostridium tetani* clustering closer to *C. acetobutylicum*[15]. Altogether, these findings serve to re-iterate that complex traits within the Clostridia such as pathogenicity are paraphyletic (also see Fig. 1). On the other hand, solventogenesis (of ethanol, butanol or acetone) may be very widespread in the genus, but there has not yet been a definitive comparative study reporting the extent of its conservation to our knowledge, and species and strains certainly vary in their productivity[16]. Topics of engineering interest have included improved characteristics such as solvent production[17,18], sugar utilisation[19,20], growth on alternative feedstocks such as lignocellulose[21–24] and the production of alternative products[25–27].

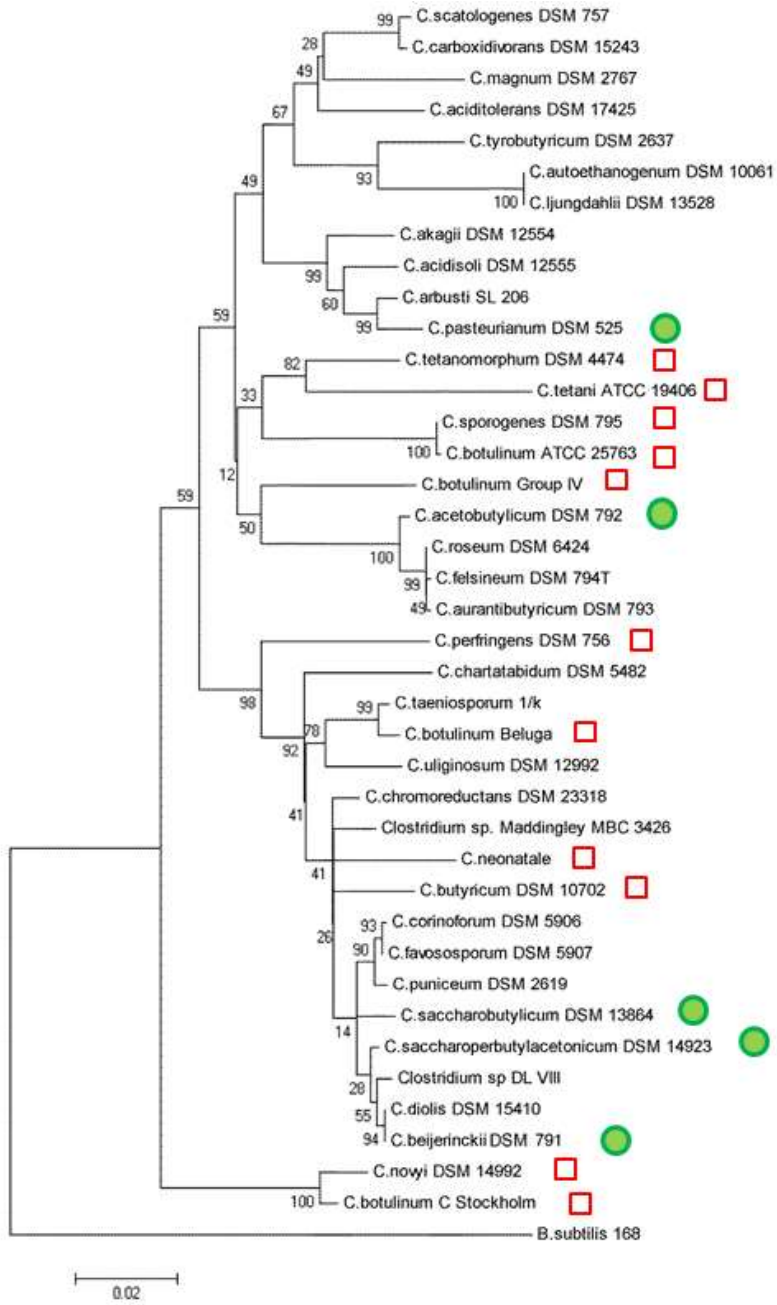


Figure 1. 1. Phylogenetic tree of selected organisms from the genus *Clostridium* based on 16S sequences.

The tree was built using Maximum Likelihood method based on the Tamura-Nei model[28]. The bootstrap consensus tree inferred from 1000 replicates[29] is taken to represent the evolutionary history of the taxa analyzed[29]. Evolutionary analyses were conducted in MEGA5[30]. Green circles denote butanol-producing species, strains of which have been used in industrial biotechnology (IB), red squares mark risk group 2 species[31] (risk group 3 being highest risk). Sequences from the species' type strains were used for the construction of this tree.

The phylogenetic tree in Figure 1.1 was constructed using 16S rRNA sequences from the type strains of the species displayed. The species highlighted in green contain strains of historical industrial biotechnological importance. It is worth noting that many of the high-solvent producers now classified as *C. beijerinckii*, *C. saccharobutylicum* and *C. saccharoperbutylacetonicum* were erroneously classified as *C. acetobutylicum* (for a list of corrected strain classifications see Keis, Shaheen and Jones (2001)[7]). It is also important to note that, of the existing body of genetics and molecular biology research in this genus, the majority has been carried out in *C. acetobutylicum* ATCC 824 (=DSM 792) which is considered a close relative to the original Weizmann strain[32]. Perhaps, second-most studied is *C. beijerinckii* NCIMB 8052 (note that this is not the species type strain and it was one of the miss-classified strains). These two strains are considered model organisms for ABE production[33]. Research in *C. acetobutylicum* ATCC 824 has further benefited from the early publication of the whole genome sequence in 2001, making it the first published clostridial genome[34], wheras the genomic sequence of *C. beijerinckii* NCIMB 8052 was made publically available in 2007[35]. Interestingly, while a transformation protocol for *C. acetobutylicum* ATCC 824 was published relatively early (1992 and 1993) (details of the protocol are discussed further below)[36,37] it was not the first solventogenic strain to be transformed with plasmid DNA – a *C. beijerinckii* NCIMB 8052 derivative (SA-1) protoplasts were transformed using heat treatment[38] in 1984. Subsequently, conjugation[39] and electroporation protocols[40] were also developed for *C. beijerinckii* NCIMB 8052 in 1987 and 1988, respectively. Additionally, PEG-mediated protoplast transformation of a *C. saccharoperbutylacetonicum* strain (also considered *C. acetobutylicum* at the time) was also reported early (1988), however the technique relied on an isolated mutant displaying an improved transformation efficiency[41], the mutant was not genetically mapped but a change in autolysin activity was observed which increased protoplast

regeneration rates. It is perhaps worth noting that the electroporation component of the aforementioned *C. acetobutylicum* ATCC 824 protocol[37] was found unreproducible in the strain obtained from the German strain collection - *C. acetobutylicum* DSM 792 (= ATCC 824) (though this may have been due to minor strain differences)[42]. To function in *C. acetobutylicum* DSM 792 (= ATCC 824), the electroporation component of the transformation protocol required modifications that made it similar to the one reported earlier for *C. beijerinckii* NCIMB 8052[40]. Another unexpected difference in the transformability of strains considered identical was reported when an electroporation protocol for *C. saccharoperbutylacetonicum* N1-4 (HMT) was published, the authors identified three distinct colony morphologies with markedly different transformation efficiencies, they also found that stocks distributed from ATCC yielded variable combinations of the three types including pure low transformability type[43]. The mechanistic and genetic basis of this variability in morphology and transformability of this strain has not been investigated.

In several cases, development of transformation protocols has required the circumvention of restriction systems which degrade incorrectly methylated DNA. Indeed, electroporation of *C. acetobutylicum* only became viable with the expression of the *B. subtilis* phage ϕ3T I methylase in the *E. coli* cloning host; this methylates the sequence GCNGC, which would otherwise be cleaved by the *C. acetobutylicum* Cac824I type II restriction enzyme[37]. Likewise, transformation of *C. pasteurianum* ATCC 6013 (DSM 525) requires the methylation of CGCG sequences, which has been accomplished by the use of the M.FnuDII[44] or M.BepI[45] methylases. Other organisms have more complex restriction systems; in *C. saccharobutylicum* NCP 262, which has two type I restriction systems, expression of the methylation and specificity domains of these systems on a plasmid in *E. coli* was sufficient to allow transformation by

conjugation[46]. Some developments have also been made in transforming non-type strains, which may have restriction patterns which differ from those of the type strains. For example, *C. pasteurianum* NRRL B-598 is part of the *C. beijerinckii* cluster[14] but requires the use of a *dam*$^-$/*dcm*$^-$ strain of *E. coli* for successful transformation, suggesting that the type IV system of this strain is particularly important[47]; conversely, *C. beijerinckii* NCIMB 8052 can be transformed with much greater efficiency even with DNA from a *dam*$^+$/*dcm*$^+$ host[48]. While an analysis of such developments in the entire genus *Clostridium* would be beyond the scope of this chapter, recent publications by Pyne et al.[49] and Minton et al.[50] provide a comprehensive review of the development of *Clostridium* strains for genetic engineering. Although the engineering of some of these species has historically been hindered by a lack of transformation protocols (as described above), these are now available for the transformation of the major industrial strains from the aforementioned five species[37,40,43,44,46]. However, it is certain that the range of genetically tractable *Clostridium* species and strains will expand with future research.

The resurgent interest in solventogenic *Clostridium* species suggests that synthetic biology tools are needed and there are clear shortages of particular biological parts with characterized activity available for the engineering of these organisms. We believe this warrants a concerted effort to address the limitations of the current toolbox. Furthermore, with the establishment of transformation protocols throughout the genus *Clostridium*, it is also important to consider the potential of adapting existing parts and tools for use in other, less well-developed solventogens.

1.2 Development of synthetic biology tools for use in solventogenic Clostridia

Traditional genetic methods for adding recombinant DNA into Clostridial cells and ultimately stably onto the chromosome can be massively improved by using synthetic biology methods to create gene cassettes that can controllably alter the genotype and resulting phenotype of the cells in ways to improve their function in industrial biotechnology. To develop a reliable set of tools for the solventogenic Clostridia, the creation of standard genetic parts is essential. The use of this single set of parts to assemble synthetic gene cassettes and larger gene clusters means that they must work together in a consistent and predictable manner.

The basic parts of synthetic biology are the minimal sequence elements with biological function in gene expression (***Figure 1.2***), including promoters, ribosome binding sites, transcriptional terminators and other factors in the mRNA that affect stability.

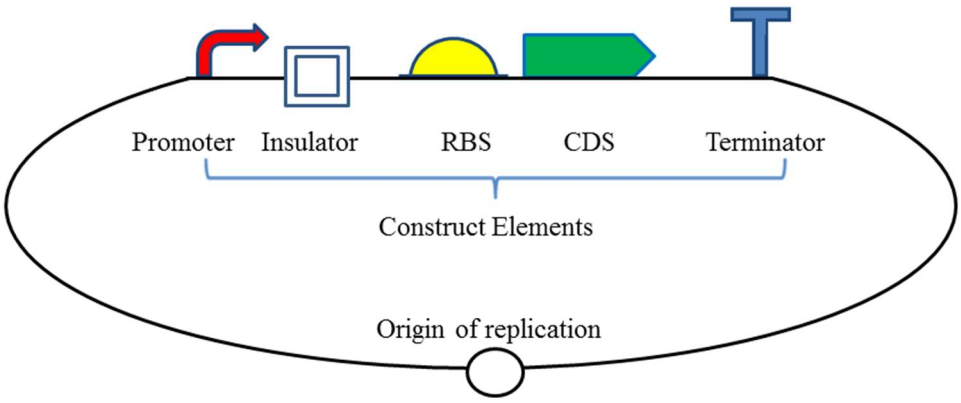


Figure 1.2 Features of genetic constructs depicted using symbols from the SBOLv[334]. Examples of promoters, ribosomal binding sites, reporters, terminators and replicons are described in the main text.

Also, synthetic biology tools rely on gene reporters to measure levels of gene expression, and these elements are often combined into plasmids which need

an origin of replication. Work in *Escherichia coli*, *Saccharomyces cerevisiae* and other model organisms has been transformed in the past decade by the rapid development, characterization and standardization of parts. However, work in these organisms benefits from a legacy of biological knowledge that is not necessarily available when working in non-model organisms, making the engineering of the latter even more challenging.

The remainder of this introduction focusses in detail on the nature of the required parts and the current state of the art in their development and use. We start by discussing origins of replication and chromosomal integration techniques. Any experiment to assess the function of a part in a particular genetic host requires ways to measure the function of this part and reporter genes and their use is considered next. Then we consider the fundamental genetic elements for transcription and translation, first starting with the concept of insulation, translation, post-transcriptonal control, transcription initiation and promoters and finally ending with transcriptional termination and terminators – the least systematically studied and understood component of the system, which then forms the focus of this thesis.

1.3 Replication

1.3.1 Plasmid origins of replication

There are three main mechanisms of plasmid replication – theta, strand displacement and rolling circle[51]. Theta replication is similar to chromosomal replication – replication is initiated at one or more origins (*ori*) by a Rep initiator protein and proceeds uni- or bidirectionally (most plasmids replicating in this way do so unidirectionally) – the name of this mechanism is derived from the resemblance of replication intermediates under electron microscopy to the Greek letter θ. Theta replication proceeds through melting of

the DNA, RNA primer synthesis and DNA synthesis initiation by extension of the RNA primer[51]. In plasmids theta-type DNA synthesis on the leading strand is continuous whereas on the lagging strand it is discontinuous and the synthesis of the strands is coupled[52].

Strand-displacement replication is initiated at symmetrical origins of replication independently of host factors when the origins are exposed as ssDNA by plasmid-encoded factors[51]. Replication then proceeds through a strand-displacement mechanism, if the origins do not initiate simultaneously the end-products of replication are circular dsDNA (consisting of the parental minus strand and newly synthesized plus strand) and circular ssDNA representing the displaced parental minus strand[51]. Whereas if the two origins initiate at the same time a theta structure is formed[51].

Rolling circle replication is unidirectional and replication is initiated at a double-stranded origin *dso* by a Rep protein that nicks the DNA at a specific site, the 3'-OH end is used as a primer[51]. The nicked parental plus strand is displaced as replication of the leading strand proceeds until the *dso* and a DNA transfer reaction terminates replication (the ssDNA is circularized in this reaction)[51]. The leading strand replication end products are circular dsDNA consisting of a parental and newly synthesized strand and circular ssDNA. Replication on the displaced ssDNA is initiated at the single strand origin *sso* by host factors[51].

Four replicons are in routine usage in solventogens: pCB102 (from *C. butyricum*), pBP1 (from *C. botulinum*), pCD6 (from *C. difficile*) and pIM13 (from *B. subtilis*)[53]. The pIM13 replication origin is thought to replicate via rolling-circle replication[54] while there is evidence that pCD6 replicates in similar fashion to pIP404[55,56] (which is either theta or strand-displacement).

The replication mechanisms of pCB102 and pBP1 are unknown. In *C. acetobutylicum*, two other replicons that have been used are the pAMβ1 replicon and the pUB110 replicon[57]; the pUB110 replicon was found to be somewhat more stable than pIM13, whereas the pAMβ1 replicon was highly unstable. The low stability of pIM13 has been taken advantage of in homologous recombination-mediated genome editing as a pseudo-suicide vector[58] whose integration in the chromosome is indicated by increased colony size[59]. The *C. beijerinckii* filamentous phage CAK1's origin of replication has been used in *C. beijerinckii* strains[60]. Additionally, the development of a replicon specific for *C. saccharoperbutylacetonicum* N1-4 was reported in 2007[61], this replicon is identical to the origin of the endogenous plasmid from *C. saccharoperbutylacetonicum* N1-504[14] . A thermosensitive origin pWV01ts derived from *Lactococcus lactis cremoris*[62] has been shown to work in both *C. ljungdahlii* and *C. acetobutylicum*[63]. Segregation and transformation frequencies are available; however, more work is needed to determine copy number and compatibility groups.

Efforts directed at improving the transformation by electroporation efficiency of *C. pasteurianum* using a transformation-curing-transformation screen with a plasmid based on the pIM13 replicon resulted in the isolation of several mutants with improved transformability[45]. This screening approach took advantage of the aforementioned low segregational stability of pIM13. One of the single nucleotide polymorphisms (SNPs) isolated was located in the CLPA_c30550 gene which has homology to structural maintenance of chromosome (SMC) proteins[45]. In *B. subtilis* SMC proteins are known to affect supercoiling *in vivo*, particularly suppressing positive supercoiling[64]. Interestingly, a historical transformable mutant of *Mycobacterium smegmatis*[65] was characterized and transformability was found to be due to a loss-of-function mutation in a non-canonical SMC protein which (as *wt)* inhibits the segregation of the heterologous

Mycobacterium fortuitum plasmids, which were used for transformations, by modulating their supercoiling levels[66]. Additionally, a mutant of *C. saccharoperbutylacetonicum* N1-4 which displayed improved segregational stability of pIM13 was isolated but was not genetically mapped[67]. More recently, FACS-sorting of a high level Flavin-binding Fluorescent Protein(FbFP)-expressing sub-population of *C. beijerinckii* cells lead to the isolation of a pIM13 mutant that contained a silent mutation and a non-synonymous substitution in the replication protein coding gene *repL* (along with inverted repeats in the 3' UTR of the FbFP gene)[68], the improved expression levels could be attributed to mRNA stability (3' UTR of FbFP gene) and higher copy number levels (*repL* mutations). However, this hypothesis needs to be validated. Nevertheless, it is clear that mutations in the pIM13 replicon and the chromosome (at least in some *Clostridium* species) could lead to improved properties of this plasmid.

1.3.2 The chromosome as a replicon and chromosomal integration

The integration of DNA into the genome, while not a 'part' in itself, is an important consideration for synthetic biology projects. Genomic integration has several advantages over plasmid-based expression strategies, including increased stability, removal of the requirement for antibiotic selection, and standardisation of copy number[69–71]. However, there are other factors that must be considered when using chromosomal integration. One implication of the integration position is the copy number effect – genes closer to the origin have a higher copy number than ones near the terminus in exponentially dividing cells due to the mechanism of DNA replication. Apart from the copy number effect, there is contradictory evidence and disagreement as to the effects of chromosome location on levels of expression – whether there is positional independence of gene transcription[72–74]. The nucleoid-associated protein HU was found to contribute to the observed positional independence of the transcription of

physically linked chromosomal genes (and thus promoter independence) in *E. coli*, the mechanism was speculated to be via constraining of transcription-induced supercoiling by HU[73]. In contrast, *E. coli hupA/B* mutants (non-producers of HU subunits) displayed drastic positional effects[73]. The HU protein that is also present in Firmicutes[75]. Genomic engineering in *E. coli* has been carried out extensively; lambda-Red recombineering[76] is well-established, and the utility of the newly developed CRISPR technique has been demonstrated in this species[77,78]. Examples of genomic integration in the solventogenic clostridia had been still somewhat limited until recently. An early enabling technology was ClosTron which adapted the Ll.LtrB intron for use in *Clostridium*[79]. Another method for genomic integration is Allele-Coupled Exchange (ACE), as demonstrated in *C. acetobutylicum*[23,24,69]. This is a homologous recombination-based method, where homology arms with different lengths are used to control the sequence of recombination events, and the second recombination leads to the generation of a selectable phenotype; Any selectable marker can be activated by coupling to a promoter, examples include the truncation or repair of the *pyrE* gene or the activation of a promoterless antibiotic resistance gene by integration downstream of a strong chromosomal promoter such as *thlA*[69]. Thus, one potential drawback to this method is that it only allows integration into a limited selection of loci. This drawback can be mitigated by the ability to carry out multiple rounds of iterative ACE, thereby making further genomic integrations into the same locus.

Nevertheless, many new developments have been made regarding the genetic manipulation of solventogenic clostridia. A variety of different allelic exchange-based strategies have been exemplified in *C. acetobutylicum*[59,80–82] and *C. beijerinckii*[83]; while most of these studies have focused on the generation of in-frame deletions and

subsequent complementation, Al-Hinai et al.[82] demonstrate the integration of a heterologous gene through gene replacement. Furthermore, the generation of point mutations through recombineering has been demonstrated in *C. acetobutylicum*[84], suggesting that the integration of DNA through this method may be feasible. Finally, mutant selection via CRISPR has been established in almost all of the main solventogenic species, with published examples in *C. beijerinckii*[85,86], *C. acetobutylicum*[87], *C. pasteurianum*[88], and *C. saccharoperbutylacetonicum*[89]. Future developments in gene editing are certain to improve the facility and range of genomic modifications that can be made in these organisms – some of the current bottlenecks stem from the reliance on native factors for homologous recombination – there is a need for relatively long homology arms and transformation with plasmid recombination templates as opposed to PCR products with short homology arms. The use of heterologous Non-Homologous End Joining (NHEJ) enzymes could enable the quick generation of imprecise but large genomic deletions[90].

1.4 Reporters of gene expression

Genetically-encoded reporter systems are the major *in vivo* gene expression measurement techniques available and are required to measure the activity of different parts being tested. Such reporter systems complement reporter-independent gene expression measurement techniques - Northern blots (mRNA) and Western blots (proteins) and modern techniques. Modern methods include the PCR-based RT-qPCR (quantitative reverse transcription PCR) and sequencing strategies such as RNA-seq and others.

The oxygen sensitivity of obligate anaerobes such as the clostridia limits the ease of use of many reporter systems, including some popular fluorescence-based and

enzymatically-based reporters, and has led to the development of some novel reporters which are slowly being adopted.

1.4.1 Fluorescent Reporter Proteins

The use of fluorescent reporter proteins is now widespread in biology. Successful use requires the correct folding and maturation of the fluorescent protein to enable detection. The level of signal for the fluorescent protein must be sufficiently high to enable accurate detection as there is no signal amplification as seen in enzymatic reporters.

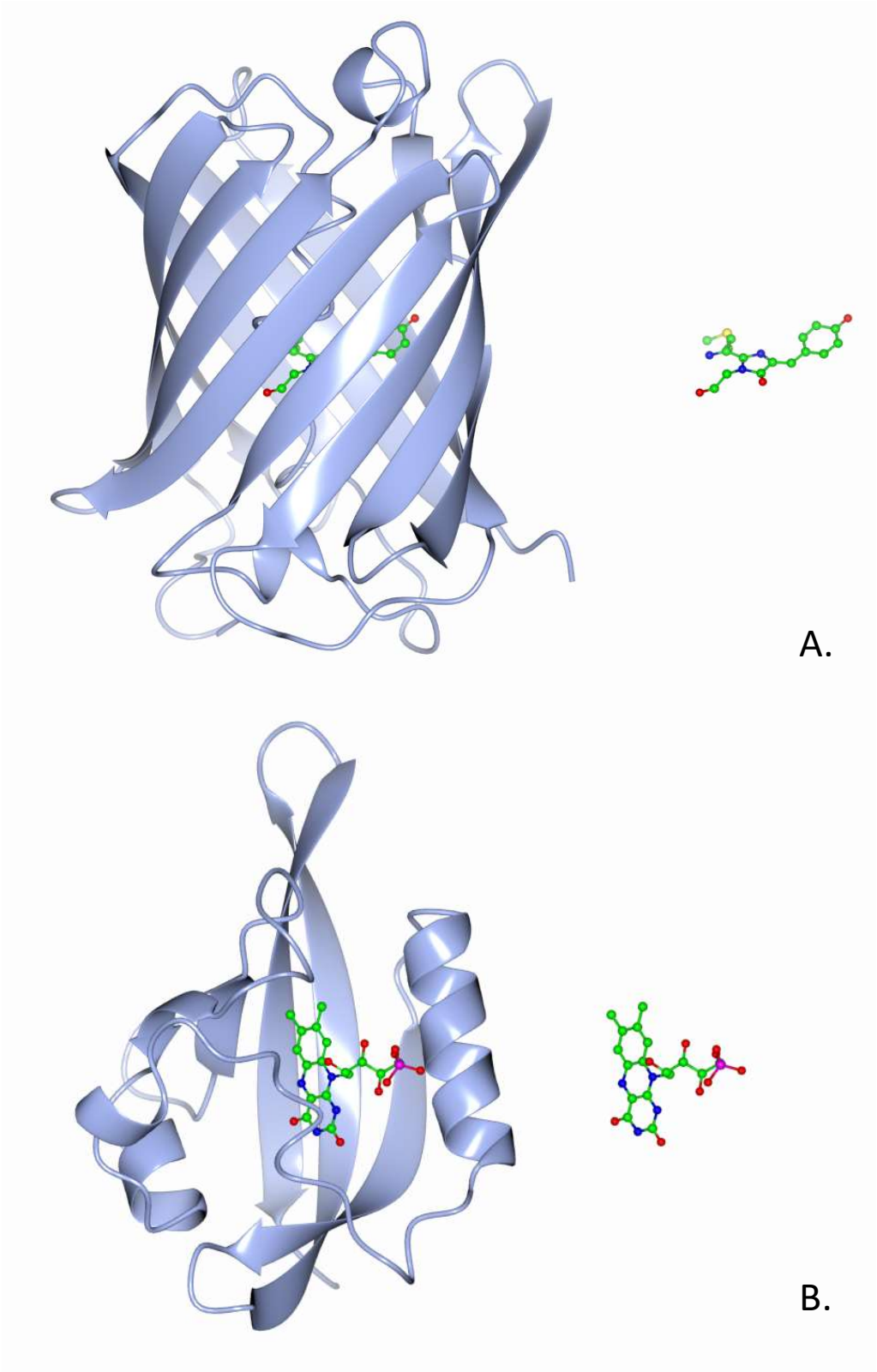


Figure 1.3 Crystal structures of fluorescent proteins mCherry and phiLOV2.1.

A. Crystal structure of mCherry (2H5Q) to the right and chromophore CH6 to the left.
B. Crystal structure of phiLOV2.1 (4EEU) and the chromophore FMN to the left.

1.4.1.1 GFP-family fluorescent proteins

The highly engineered family of Green Fluorescent Proteins (GFPs) now includes variants with improved brightness and photostability and with a range of different colours (different excitation and emission maxima)[91]. The overall structure of the GFP-family proteins is a beta barrel, an example, mCherry is shown in **Figure 1.3A**. The GFP-family fluorescent proteins are unique because their chromophore is not a separately synthesized chemical group but is formed of modified amino acids post-translationally[92]. During maturation, the chromophore amino acids undergo cyclization and incorporate an oxygen atom[92]. For example, the mCherry chromophore CH6 (**Figure 1.3A**) is generated from methionine, tyrosine, and glycine (MYG) in positions 71, 72 and 73[93]. This process doesn't occur under anaerobic conditions and therefore, the major limitation of the GFP-like proteins for anaerobes is the requirement of molecular oxygen for chromophore maturation and fluorescence[94]. This excludes the possibility of real-time gene expression monitoring; however, *in vivo* observations are still possible in some oxygen-tolerant clostridia such as *Clostridium perfringens*[95], where yellow fluorescent protein (YFP) was synthesized and fluorescence was developed after exposure of live cells to atmospheric oxygen[96]. Futhermore, EGFP was used successfully to monitor gene expression in *C. perfringens*[97].

Using a similar principle, it is possible to obtain a snapshot of the protein levels in whole cells by exposing fixed anaerobically grown bacteria to atmospheric O_2 in a process termed 'aerobic fluorescence recovery', enabled by the discovery that GFP-family proteins can undergo maturation in fixed cells. Two studies have demonstrated the technique in *C. difficile* using paraformaldehyde and glutaraldehyde to fix cells

expressing codon-optimized variants of CFP (cyan)[98] and mCherry (red)[99] and also GFPmut3b (green)[99]. While all three constructs generated an increase in fluorescence in their respective channels, mCherry had the least background (in the red channel), followed by CFP (in the blue channel) with GFPmut3b being difficult to distinguish from the high level of background in the green channel[99]. This approach works as the fixatives cross-link primary amines (especially lysine residues) which are absent from the GFP-family chromophores. Fixed cells offer an advantage to the imaging of dying live cells as they more accurately represent normal protein localization (gene expression changes are also likely to occur in dying or metabolically stressed cells)[99]. So far, the technique has not been used to systematically quantify gene expression but rather to label proteins and track their intracellular localization[99,100].

1.4.1.2 Flavin-binding Fluorescent proteins

The flavin-binding fluorescent proteins (FbFPs) are a class of alternative fluorescent reporters capable of maturation in anaerobic conditions. The FbFPs are small proteins (~11-15kDa) that have oxygen-independent fluorescent properties, using a flavin-mononucleotide (FMN) cofactor as the chromophore. An example of a FbFP (phiLOV2.1) is shown in **Figure 1.3B**.

A pioneering study in 2007 reported the development of three anaerobic fluorescent reporters derived from the LOV (Light-, Oxygen- or Voltage-sensing) domains of bacterial proteins: BsFbFP derived from YtvA of *Bacillus subtilis,* PpFbFP from SB2 of *Pseudomonas putida* plus an *E. coli* codon optimized variant of BsFbFP called EcFbFP[94]. This domain was engineered to emit fluorescence by mutagenesis of a reactive cysteine in the FMN-binding pocket[94].

These proteins are now commercially available from Evocatal Gmbh under the trademark name 'evoglow'. The three currently listed variants marketed for use in *Clostridium* species are Bs1 (monomeric BsFbFP), Bs2 (dimeric BsFbFP) and Pp1 (PpFbFP).

The reporters' functionality in *C. acetobutylicum* was first demonstrated by Schulz in 2013; the highest fluorescence levels were observed for a *C. acetobutylicum* codon-optimized Pp1 (referred to in text as "Pp2"), followed by codon-optimized Bs2 (referred to as "Bs3"). Interestingly the same constructs gave inverse results in *E. coli*[101]. While Evocatal GmbH offers C-Pp1 and C-Bs2 *Clostridium* codon-optimized reporters with publicly available nucleotide sequences, it is not clear whether the nucleotide sequences (reflective of the codon optimization approach) are the same as Pp2 and Bs3. A study in 2014 reported the placement of the evoglow Pp1 reporter downstream of the *cipP* promoter to monitor growth of *Clostridium cellulolyticum* on cellulose[102]. A more recent study detected increased fluorescence in *C. acetobutylicum* and *Clostridium ljungdahlii* by using Pp1 and Bs2[63]. Crucially, the fluorescence increases were only observable in PETC medium without yeast extract (which is known to contribute to high green background fluorescence)[63]. The functionality of the evoglow Pp1 variant was also demonstrated in *C. pasteurianum*[103]. A Bs1-coding gene variant codon-optimized for *C. beijerinckii* was reportedly used in that organism successfully, the organism was grown in P2 (no yeast extract) and P2YE medium (1g/L of yeast extract)[68]. Furthermore, Bs2 expression resulted in increased fluorescence in *Clostridium tyrobutyricum* (it is unclear whether fluorescence studies were performed on RCM or CGM-grown cultures)[104].

Another FbFP example, developed from the *Arabidopsis thaliana* LOV2 domain of the blue-light receptor protein Phot2, is the improved LOV (iLOV) FbFP[105]. The iLOV

FbFP has been further modified for enhanced properties, generating variants such as photostable iLOV 2.1 (phiLOV 2.1 (shown in **Figure 1.3B**) and others[106,107]. While phiLOV 2.1 FbFP has improved photostability, it is still substantially lower than that of GFP. A recent study demonstrated the utility of a *C. difficile* codon optimized phiLOV 2.1 in three *Clostridium* species (expression was driven from an engineered anhydrotetracycline-inducible promoter)[108]. Under the test conditions, the three species exhibited varying levels of fold-increase of fluorescence over their background autofluorescence level: 3.2-fold increase in *C. difficile* R20291 (not a *Clostridium sensu stricto* species), 5.6-fold in the pathogenic toxin-producer *Clostridium sordellii* ATCC 9714 and 4.5-fold in the solventogen *C. acetobutylicum* ATCC 824[108]. While the FbFP reporters should be detectable when expression is driven from very strong promoters, further improvements in brightness and photostability will make them more generally useful for multiple applications. CreiLOV is another FbFP variant derived from blue light photoreceptor Phot1 of *Chlamydomonas reinhardtii*[109] with reported improved thermostability, photostability, maturation and brightness[110]. Indeed, studies have reported lack of detection of phiLOV2.1 and CreiLOV in *C. acetobutylicum*, highlighting the problems associated with their expression. Furthermore, a study in *C. perfringens* found Bs1's rapid photobleaching during microscopy and its low brightness undesirable and in comparison to EGFP[97].

One of the drawbacks of the FbFPs is that all of the available variants are of the same colour; additionally, unfolding caused by Sec-mediated translocation to the periplasm has been reported to result in loss of fluorescence due to the loss of the bound flavin cofactor[107]. These limitations prevent the use of FbFPs as the sole reporters for multi-output circuits or secretion. Engineering of FbFPs for different emission spectra has proven to be somewhat complicated; while a simulation study of a theoretical iLOV

mutant predicted a red shift increase[111], experimental results showed that the mutation resulted in an overall brightness decrease and a blue shift of emission[112]. Research is underway to develop more fluorescent derivatives of the LOV domains and to explore the potential for diversifying their colour spectrum[111,112]. The other limitation of FbFPs is their relative dimness, which can decrease the sensitivity of a reporter system; depending on the measurement instrument used, a promoter's expression level might be below the detection limit. Photostability improvements (as in phiLOV) have resulted in a dimmer mutant[106]; for applications where a longer exposure time is crucial, the dimmer, more stable protein is preferable, but for an automated single measurement application such as flow cytometry, the brighter variant may be more suitable. Overall, fluorescent protein reporters provide the ability to quantify cell-to-cell heterogeneity of gene expression (when used in flow cytometry or microscopy), which can be very useful, and while there are still some limitations with FbFPs, they have the potential to be a route to reliable *in vivo* real-time expression monitoring in *Clostridium*.

1.4.1.3 Self-labelling proteins and fluorogen-activating protein labels

An alternative approach in imaging to using autofluorescent proteins such as the GFP-family proteins and flavin-binding fluorescent proteins is the use of self-labelling proteins or fluorogen-activating protein labels[113]. Self-labelling proteins catalyse the covalent auto-attachment of an organic fluorescent dye[113]. Fluorescent dyes have the following advantages over fluorescent proteins – higher brightness, higher photostability, broader colour range and more useful chemical properties such as photo-activation and photoswitchability[113]. Disadvantages of using fluorescent dyes over fluorescent proteins include higher background fluorescence and reduced probe stability[113]. Two examples of self-labelling enzyme tags encoded on a multicopy plasmid that have been applied to *C. difficile* are the CLIP-tag and SNAP-tag

systems[114]. SNAP-tag is based on the human DNA-repair enzyme O^6-alkylguanine-DNA alkyltransferase (hAGT)[115], whereas CLIP-tag was generated from SNAP-tag by mutagenesis for the purpose of orthogonal substrate specificity[116]. While SNAP-tag reacts with O^6-benzylguanine fluorescent dye conjugates[115], CLIP-tag is preferentially labelled by O^6-benzylcytosine[116].

Recently, the use of the SNAP-tag label as reporter for gene expression in *C. acetobutylicum* ATCC 824 was demonstrated[117]. However, in that study the SNAP-tag achieved a maximum of only 1.5-fold mean fluorescence intensity increase over the control using the red dye SNAP®-Cell TMR-Star while the use of the green dye SNAP®-Cell 505-Star resulted in a smaller increase[117].

The yellow fluorescence-activating and absorption-shifting tag protein (Y-FAST) was also used as a reporter in *C. acetobutylicum* ATCC 824[118]. The construct using the *thl_ac* promoter resulted in nearly 2-fold increase in fluorescence. The use of a mutant *thl_ac* produced an over 4-fold increase in fluorescence. Similarly to SNAP and CLIP tags, the Y-FAST tag is not fluorescent itself but binds the ligand 4-hydroxy-3-methylbenzylidene-rhodanine (HMBR)[119]. The fluorogen (HMBR) is also non-fluorescent by itself. A notable difference between Y-FAST and SNAP or CLIP is the full reversibility of the binding.

1.4.2 Enzymatic Reporter Proteins

Enzymatic reporters catalyse a (preferably) unique reaction either *in vivo* or *in vitro* that is readily measurable and distinguishable from the background level. Specific activity is calculated to estimate protein levels and thus gene expression. Enzymatic reporters frequently require the addition of substrates and cofactors to the assay reaction mixture

as well as the production of cell lysate. This often means that enzymatic assays involve more preparation steps than fluorescent proteins but can have adjustable sensitivity by varying substrate levels. Also, *in vitro* lysate assays are, in typical practice, bulk population measurements.

1.4.2.1 Chloramphenicol acetyltransferase

The first reporter which has been extensively used in the *Clostridium* genus, both *in vivo* and, more quantitatively, *in vitro*, is the chloramphenicol acetyltransferase (CAT) reporter (encoded by *catP* in the commonly used modular pMTL80000 vector series[53]). The system was first developed for use in *C. perfringens* (a medically relevant non-solventogenic bacterium)[120] and has since been used in *C. acetobutylicum*[121]. Chloramphenicol acetyltransferase is an enzyme that catalyses the covalent attachment of an acetyl group from acetyl-CoA to chloramphenicol[122], and is the basis of the chloramphenicol resistance marker found in many bacterial vectors. Transfer of the acetyl group exposes the thiol group of CoA, allowing the progress of the reaction to be observed by addition of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, also known as Ellman's reagent); this compound reacts with the free thiol, releasing 5-thio-2-nitrobenzoate, which can be detected by measurement of absorbance at 412 nm[122, 123]. The assay relies on a continuous spectrophotometric rate determination to calculate specific activity. Potential drawbacks include endogenous activity in chloramphenicol resistant strains (which may be resolved by disruption of the resistance gene) as well as high levels of endogenous non-specific coenzyme A transferases[124] (knockouts of which would be more laborious and would likely have growth and phenotypic effects) and the cost of the substrate acetyl-CoA.

1.4.2.2 Carbohydrate hydrolases: β-glucuronidase, β-galactosidase, amylase, endoglucanase

The CAT assay's drawbacks led to the adaptation of a classical *E. coli* reporter in *Clostridium*: the β-galactosidase enzyme, encoded by *lacZ* (the gene was derived from *Thermoanaerobacter thermosulfurigenes*)[124]. To quantify enzyme activity spectrophotometrically, ortho-nitrophenyl-β-galactoside (ONPG) is used in an *in vitro* assay and an increase in absorbance at 420 nm due to the release of ortho-nitrophenol is measured. Similarly, the GUS reporter system, which utilizes β-glucuronidase (*E. coli gusA*), has been used in *C. acetobutylicum* in a fluorimetric assay with a cell lysate[125]. The fluorimetric assay measures the release of 4-methylumbelliferone (4-MU) after cleavage of 4-MU-β-galactoside or 4-MU-β-glucuronide (by a β-galactosidase or β-glucuronidase, respectively), 4-MU emits light at 460 nm when excited by 365 nm light[126]. Both the β-galactosidase and β-glucuronidase reporter systems benefit from the commercially available range of fluorometric, spectrophotometric and histochemical substrates and kits, making them an improvement over the CAT assay, although they are still not inexpensive.

The endogenous *amyP* gene (encoding an amylase expressed during solventogenesis) has been used in *C. acetobutylicum*[127] as a reporter to study the phenomenon of strain degeneration (loss of solventogenesis) which is often caused by loss of the pSOL1 megaplasmid on which *amyP* is located. A codon-optimized amylase ($AmyE^{opt}$) has been used successfully as a secreted reporter in *C. difficile* by addition of a zinc metalloprotease PPEP-1 signal sequence[128]. It is noteworthy that the strain used in the above study was not capable of degrading starch under laboratory conditions; use of amylase as a reporter in amylolytic strains (such as *C. acetobutylicum* ATCC 824) may require knockout of endogenous amylases to increase signal to background ratio.

In *C. beijerinckii*, a secreted endoglucanase (*eglA*) was cloned from *C. saccharobutylicum* NCP 262 and used as a reporter for gene expression[129]; assays used were agar plate and cell lysate assays which measured substrate (carboxymethylcellulose, CMC) clearance on plates and product (p-nitrophenol from cleavage of *p*-nitrophenyl cellobioside) accumulation in lysate, respectively.

1.4.2.3 Lipase and Alkaline phosphatase

The lipase encoded by *tliA*, from *Pseudomonas fluorescens* SIK W1, has also been used as a reporter in *C. beijerinckii* NCIMB 8052[130]. Lipases are enzymes which hydrolyse the ester bonds found in long-chain acylglycerols, releasing fatty acids[131]. Activity can thus be assessed by an enzymatic assay measuring the cleavage of *p*-nitrophenyl-esters[131] such as *p*-nitrophenyl decanoate[130]. However, this reporter was only useable in *C. beijerinckii* as this species has a very low endogenous lipase activity. As such, it may not be suitable in other species, such as *C. acetobutylicum*, which has been observed to show inducible lipase activity[132].

A colorimetric alkaline phosphatase assay was developed and used in *C. difficile* in 2015 based on the *phoZ* gene product from *Enterococcus faecalis*[133]. Activity of this gene can be determined by a colorimetric assay with *p*-nitrophenyl phosphate as the substrate. Use in solventogenic clostridia could be limited by native phosphatase activity. In order to examine the suitability of *C. difficile* as a host for this reporter, BLAST analysis was used to screen for *phoZ* homologues, and activity towards 5-bromo-4-chloro-3-indolyl phosphate was tested[133]. While the four main industrial strains do not have a *phoZ* homologue, homologues can be found in other species such as *C. pasteurianum* (CLPA_RS02340, with 29% identity to *phoZ*).

1.4.2.4 Bioluminescent Reporters

The luciferase (*lucB*) reporter was also used in *C. acetobutylicum* successfully, allowing luminescence detection[134]. However, it also requires oxygen, ATP and luciferin. Notably, the cells used in the luciferase assay were neither lysed nor fixed, but were live cells that were exposed to atmospheric conditions, washed, and kept on ice. This treatment could conceivably introduce changes in gene expression levels prior to measurement. This assay has the lowest background signal level but the requirement for live cell exposure to oxygen may introduce variability. A codon-optimized luciferase ($sLuc^{opt}$) was also successfully secreted in *C. difficile* using the aforementioned zinc metalloprotease signal peptide[128].

Bacterial luciferase reporter systems based on the entire lux operon such as *luxCDABE* from *Aliivibrio fischeri* or *luxABCDE* from *Photorhabdus luminescens* do not require the addition of exogenous substrates for the development of bioluminescence[135]. However, only the heterodimeric luciferase (*luxAB*) from the *A. fischeri* operon has been used (with the exogenous addition of substrates flavin mononucleotide and decyl aldehyde) in *C. perfringens*[136].

1.4.3 Conclusion on Reporters

While there are many reporter choices available to clostridial researchers, we would argue that the multiplicity of reporters used has not helped ease the comparison of data obtained by different laboratories. Altogether, a single reporter has not been established as a community standard; given the drawbacks of each particular system, it is difficult to identify one standout reporter, although our hope would be that improved fluorescence reporters with increased brightness and photostability would be the most useful and enable single cell studies in live cells. Currently, we would advocate the

continued use of enzymatic reporters such as GusA and LacZ, coupled with the adoption of fluorescent reporters in the *Clostridium* genus as the two reporter types offer distinct advantages. The latter, importantly, enable single-cell resolution. While novel fluorescence-based techniques have been reported (SNAP-tag, Y-FAST) in the genus, they are not as bright as GFP-family proteins such as mCherry. Comparing results obtained with different reporters can give increased confidence in the results obtained. An interesting approach to direct reporter comparison is the use of translational fusions between reporters; examples from *E. coli* include a FRET pair YFP-FbFP fusion[137] and Gemini (lacZα-GFP) fusion[138]. If successfully applied to the *Clostridium* species, such bi-functional reporters have the potential to become a single standard reporter for part characterization. It has been argued that comparability of measurements of activity of genetic elements for the purposes of part characterization is hindered by differing reporters, activity calculations and conditions used[139]. However, this approach hasn't been widely adopted. To evaluate SNAP-tag, researchers produced a fusion between *gusA* and SNAP-tag and tested it in *C. acetobutylicum,* ultimately determining that SNAP-tag did not produce a sufficiently large fluorescence increase for the purpose of characterizing promoters weaker than wild-type *thl_ac*[117]. Additionally, α-peptide complementation strategies using the *E. coli lacZ* gene in *B. subtilis*[140] and *Mycobacterium smegmatis*[141] have been reported but a similar approach in Clostridia has not been attempted. Interestingly, the *T. thermosulfurigenes*-derived LacZ protein appears to lack the α-complementation region[142] and is a dimer[143] (rather than a tetramer like the *E. coli* LacZ protein).

1.5 Insulation

A biological part's adjacent sequences can have a profound effect on its behaviour compared to the sequence context in which it was characterized, a property referred to

as context dependence[144]. This poses a fundamental challenge to the synthetic biology principle of part creation and characterization[145]. Examples include the activity of a minimal promoter being altered by downstream sequence that doesn't have promoter activity[146]. To counteract this issue in reproducibility genetic engineers have started utilizing a new class of parts called insulators[147]. Several strategies can be undertaken to insulate a part from its genetic context, and these can be split into two main categories: DNA-level insulators (such as simply using flanking buffer zones of sequence without a biological function or secondary structure) and RNA-level insulators. The latter includes post-transcriptional modification of RNA as well as the commonly used flanking double terminators[148] that prevent read-through transcription into synthetic gene constructs. Post-transcriptional insulators consist of inclusion of ribozyme-based insulators or using CRISPR-RNA-processing to decouple the 5' UTR from the coding sequence (CDS)[145,149].

Different parts require different types of insulation in order to achieve maximum reproducibility without compromising features such as strength. Promoters used in synthetic biology projects are often minimal (-35 to -10 region) and without characterized transcriptional start sites; the inclusion of important functional elements such as the UP element has been previously recommended.[150] Accordingly, the addition of an upstream and downstream insulating sequence has increased reproducibility in different genetic contexts[151]. It is worth noting that the strong *Clostridium* promoters that are in widespread use in the *Clostridium* community, *fdx* and *thlA*, are 200 and 150bp respectively. Even though this recombinant *thlA* is longer than a usual minimal promoter, a long 5'UTR contributes to this part's length (recombinant *thlA* is 59bp from its 5' end to the transcriptional start site), while the recombinant 5' end is slightly truncated to exclude a Rex NADH-dependent regulator binding site (as mentioned

before). A study that reported the use of a minimal *thlA* promoter in *C. acetobutylicum*, *miniPthl* which has a truncated 5' UTR, did not test activity variation in different genetic contexts[152].

Work in *E. coli* has demonstrated the utility of 'bicistronic' ribosome binding site (RBS), where a leader peptide is translationally coupled with the CDS of interest, in improving reliability and context independence (downstream gene sequence) when a particular 5' UTR is combined with a new coding sequence[151]. An upstream RBS and start codon initiate translation of the leader peptide, the stop codon of which overlaps with the start of the downstream CDS, while the latter's RBS is positioned within the leader peptide. The ribosomes translating the leader peptide unfold the 5' UTR (of the downstream CDS), preventing it from forming secondary structures with the mRNA of the downstream CDS and thus influencing translation. The same study employed standard transcriptional start site ("+1 promoter"): 5' UTR junctions to minimize (or insulate against) unforeseen effects of combining promoters with new 5' UTRs.

As mentioned above, the observation of ribosomal repression of transcription termination also necessitates the more widespread use of 'distance' insulators of a sequence without emergent function and secondary structure to separate the stop codon and the stem-loop hairpin[153]. Such strategies are yet to be implemented in *Clostridium* engineering projects.

1.6 Translation

1.6.1 Translation initiation elements

Bacterial ribosome-binding sites (RBS) are short sequences located in the 5' untranslated region of messenger RNA (mRNA) transcripts, consisting of a Shine-Dalgarno sequence (SD), polynucleotide spacer, and a translation initiation codon[154].

Commonly, native promoter-RBS combinations have been used (such as in the pMTL80000 vectors). An alternative strategy is to use a native RBS (such as that of the *C. acetobutylicum thlA* gene) fused to a new promoter, as in a study from 2016[24] that generated several new hybrid promoters. Others have experimented with the length of the spacer[155,156]. Yet, in contrast to model organisms (*E. coli* and even *B. subtilis*), there are few published comparisons of modified RBSs for use in solventogenic clostridia. Hence the generation and screening of synthetic RBS libraries could be a promising route for optimising the expression of synthetic gene constructs in clostridia.

SD sequences provide sequence complementarity for the 3' terminus of the 16S rRNA (known as the anti-Shine-Dalgarno or aSD[157]) which acts as a guide for the ribosome complex enabling mRNA recognition by the translation machinery and translation initiation. The consensus hexamer SD sequence is AGGAGG and the optimal aligned spacing from the start codon in *E. coli* was reported to be 5nt[157]. While most of the knowledge on bacterial translation initiation comes from *E. coli* work, early studies indicated that the Firmicute *B. subtilis* requires a longer complementary region between the SD and the 16S rRNA to achieve comparable expression levels[158,159]. An early study on translation initiation revealed that Firmicutes have, on average, a higher complementarity of the predicted SD region to the 16S rRNA 3' terminus than *E. coli* does[160]. Recently, a systematic analysis of SD-aSD pairings in *B. subtilis* and in *E. coli* confirmed this trend[161]. Replicating these studies in solventogens would provide useful information.

The spacer is the mRNA region between the SD and the start codon. The aligned spacer (which is the distance between the start codon and the 5' end of the aSD, base-paired to

a SD[157,162,163] as revealed by sequence alignment) is of particular importance for translation initiation[157]. Defining the length of the aligned spacer precisely is difficult, as interpretations of the extent of the aSD region (starting from the 3' 16S rRNA terminus) vary. The 3' end of the 16S rRNA gene (aSDs) of *B. subtilis* is identical to that of *C. perfringens*[164] (as well as identical to those of solventogenic clostridia, based on our sequence analysis of published genomes) but to our knowledge there is no reported experimental validation of the clostridial mature 16S rRNA 3' ends *in vivo*. Spacers in different species may have different optimal lengths; for example, spacers of *Pyrococcus abyssi* are, on average, roughly 3 nucleotides longer than those of *E. coli*[165], whereas the spacers in *Bifidobacterium longum*[166] would be considered shorter. According to our definition of the putative clostridial aSD (5' GAUCACCUCCUUUCU 3'), in *C. acetobutylicum*, the native RBS of the *thlA* promoter has an 'aligned spacer' of 4 bases.

The effect of altering the length of the *thlA* spacer was recently investigated in two studies conducted in *C. acetobutylicum*[155,156]. Interestingly Yang *et al.,* 2016 showed that a lengthened *thlA* RBS spacer with an extra 6 nucleotides (to a total of 14 bases-aligned spacer of 10bp), encoding a SalI restriction site, did not significantly alter reporter expression in comparison to the WT *thlA* RBS. Shortening the spacer below the WT length resulted in a decrease in expression, while further increases over an aligned spacer length of 10 bases (by the addition of a XbaI site) removed almost all of the expression[155].

Yang *et al.* (2016) demonstrated the potential benefits of utilising modified spacers in *C. acetobutylicum*[155]. Overexpression of the biotin synthesis genes *bioY*, *bioD*, *bioA*, and *bioB* was observed to provide an improvement in growth phenotype and solvent production. The *thlA* promoter was used to drive expression of *bioY*, *bioD* and *bioA*;

replacement of the *thlA* RBS spacer with a shortened, less 'effective' variant resulted in a further improvement in growth characteristics. These results demonstrate not only the ability of modified RBS to optimize expression of synthetic pathways, but also the necessity of considering the effects of any alterations to the spacer, e.g. the introduction of restriction sites, when generating synthetic constructs. In 2017, Yang *et al.* generated a library of spacers (a BamHI site preceded the start codon in all cases) by starting with an aligned spacer of 2 bases and progressively increasing its length by 2 bases. After testing 11 variants the authors found that an aligned spacer of 4 bases gave the strongest levels of expression followed by 8, 6 and 10 bases in that order. Not only the length but also the sequence of the spacer influences translation initiation, for example the introduction of a secondary SD within the primary SD's spacer region[167].

Computational tools to design RBSs exist, such as the RBS calculator, which use biophysical models of RBS recognition and translation initiation, including RNA base-pairing between the aSD and SD, spacer length and messenger secondary structure[168,169]. There are reports that *de novo* design produces more accurate results than translation initiation rate prediction of natural sequences for Gram-positives[170]. There are few reports of the RBS calculator in *Clostridium* species – such as in the thermophilic *Clostridium thermocellum*[171].

1.6.2 Codon usage during translation

While codon usage is not formally a 'biological part', it is an important feature of coding sequences, the differential frequency of synonymous codons amongst genomes, referred to as codon usage bias[172,173], has been shown to strongly influence heterologous protein expression levels[174,175].

High genomic AT-content is characteristic of the Firmicutes and is reflected in the nucleotide composition of coding sequences. The *Clostridium* species' codon usage differs from that of other Firmicutes as well as the Proteobacterium *E. coli*[176] and there are also bioinformatically observable variations within the *Clostridium* genus itself[177] but the significance of the latter in influencing gene expression has not been experimentally verified to our knowledge. Genetically encoded reporters have been used heterologously within the Firmicutes phylum without codon-optimization, for example *Staphylococci* have been sources of reporters and antibiotic resistance genes for Clostridia. On several occasions researchers have successfully used native reporter genes from *E. coli* (*gusA*)[125], *T. thermosulfurigenes* (*lacZ*)[134] and the firefly *Photinus pyralis* (*lucB*)[134] (we describe these reporters in more detail in the Enzymatic Reporter Protein section). Codon-optimized genes for *C. difficile* have been used in *C. acetobutylicum* [108], as well as bespoke *C. acetobutylicum* codon-optimized ones[23,101]. Researchers have also codon-optimized several GFP-like proteins for use in Firmicutes[178] (with *B. subtilis* in mind) but these have not been used in solventogens to our knowledge.

Codon optimization is not a trivial problem and codon optimization strategies vary considerably. The codon-adaptation index (CAI)[173] has been the historical measure of codon usage bias in an organism while there are others such as the codon bias index and the effective number of codons[179]. Interestingly, simply improving a heterologous gene's CAI (making it more like a native gene or a highly expressed native gene) has not been found to correlate with expression levels[180,181]. Efforts have been made to improve the indices describing codon usage bias and translational efficiency (which codon usage bias is thought to reflect) by studying endogenous gene expression[182,183]. In addition, condition-specific usage tables have also been reported[184]. In an alternative

strategy, Welch and colleagues developed genetic algorithms to select partial least squares regression models which revealed that codons predominantly read by tRNAs that are most highly charged during amino acid starvation were good predictors of expression levels. Based on these results the researchers developed proprietary codon optimization algorithms to maximize protein expression[180,185] which allowed them to predict expression levels in *E. coli*.

While it is clear that there is room for improvement in the heterologous protein expression strategies used in *Clostridium* solventogens, codon optimization strategies themselves are still being developed and the underlying principles are not yet fully understood; progress and existing approaches in the field have been reviewed elsewhere[186]. A good starting point is for researchers to report the details of the codon optimization strategy undertaken when publishing work containing codon-optimized genes. This way data from heterologous protein expression in solventogens can be compared more reliably.

1.7 Post-transcriptional control of gene expression

Tuning gene expression levels in *Clostridium* species has been achieved using control at the RNA level – by either influencing translation or changing RNA degradation rates.

1.7.1 Control of mRNA stability

Another potential avenue for optimisation of expression levels is the adjustment of mRNA stability. Altering the stability of an mRNA transcript influences the number of transcripts in the cell, thereby affecting the overall rate of translation. In bacteria, a number of factors are associated with mRNA stability, such as secondary structures, RNase recognition sites and polyadenylation, amongst others[187]. The presence of secondary structures at the 5' end of the mRNA has been observed to provide an

increase in mRNA stability[187]. In *E. coli*, these structures prevent the binding of RNase E, an endonuclease which binds single-stranded RNA at the 5' end and then scans for cleavage sites. This property has been exploited in *E. coli* by the generation of libraries of synthetic hairpins for introduction into the 5' untranslated region (UTR)[188]. While *C. acetobutylicum* has an RNase E homolog, RNase E/G, it is not certain if this behaves in the same way as *E. coli* RNase E, due to having a different domain organisation[189]. Nevertheless, *C. acetobutylicum* also has a homolog of RNase Y[190], which fulfils the role of RNase E in *B. subtilis*[191]. Correspondingly, the utility of 5' hairpins for protection of mRNA has been demonstrated in clostridia; the introduction of 5' stem-loop sequences was confirmed to increase mRNA stability, reporter expression, and expression of the genes *adhE1* and *adhE2* in both *C. acetobutylicum* and *C. beijerinckii*[192]. This effect was much more pronounced during solventogenesis than during acidogenesis. Similarly, the introduction of a terminator hairpin in the 3' UTR can result in improved mRNA stability via inhibition of nuclease activity[187]. Although not fully explored, this principle has been demonstrated in solventogenic clostridia; the expression of a *cat* reporter gene in *C. acetobutylicum* was observed to increase by approximately 36% when the downstream *adc* terminator was replaced by a synthetic terminator, BBa_B1010, from the iGEM registry[192]. Additionally, a terminator with activity in the reverse orientation prevents the formation of antisense transcripts which are known to reduce protein expression levels[193].

A completely opposite approach is to reduce mRNA stability by introducing RNase sites into the 3' UTR. This may be useful in the case of proteins which form inclusion bodies when overexpressed. In *E. coli*, sequences derived from the *cat* gene, which contains 28 RNase E sites, have been shown to reduce mRNA stability[194]. When

combined with two poorly soluble heterologous enzymes, the *cat*-derived 3' UTRs were shown to result in an increase in soluble protein, with concomitant increases in enzyme activity. The authors noted that this improvement could not be observed simply by using a weaker promoter; it was proposed that by limiting the stability of the mRNA, the number of proteins that could be produced from a single transcript was decreased, thus limiting the local concentrations of protein during translation.

1.7.2 Antisense RNAs and Riboswitches

Bacterial antisense RNAs (asRNAs) are transcripts with complementarity to another RNA. Several studies have used asRNAs to reduce native gene expression in *C. acetobutylicum*[193,195] and in *C. pasteurianum*[196], demonstrating that multiple routes to regulate gene expression are available in the solventogenic clostridia.

Riboregulators are another class of naturally-occuring and generally trans-activating asRNA elements that respond to a signal nucleic acid by Watson-Crick base pairing[197]. Riboregulators can regulate transcription (in the case of 6S RNA possibly by binding RNAP and acting as a DNA mimic), RNA stability (for example. they can increase stability by basepairing with the 3' end) and mRNA translation (basepairing with translation initiation region can prevent its recognition by the ribosome)[197]. They have defined sensor and effector domains and have been rationally designed to repress[198] and activate gene expression[199] in *E. coli*. Toehold switches (a synthetic cis-regulatory subgroup of riboregulators) that activate gene expression in the presence of cognate RNAs rely on sequestering the RBS and start codon[200]. Like riboswitches, riboregulators are known to be present in *Clostridium* genomes.

Riboswitches are RNA sequences that are able to bind to a soluble ligand, influencing the properties of the RNA. In nature, riboswitches typically contain a binding domain, or 'aptamer', and an 'expression platform' which mediates the effect[201]. Binding leads to a change in conformation of the RNA, leading to formation of a secondary structure which can act as a terminator or an anti-terminator. While riboswitches are typically found in the 5' UTRs of mRNAs, some have been recently determined to control the expression of antisense RNAs or protein-sequestering small RNAs, while others have been shown to control access to recognition sequences such as RNase sites[202]. The range of applications has been further developed by the creation of synthetic riboswitches. In bacteria, riboswitches have been developed that can influence translation initiation by inhibiting access to the RBS; ligand binding leads to a conformational change or even to self-cleavage, revealing the RBS and allowing translation[203]. Riboswitches are found in all taxa and a number have been characterized in the solventogenic clostridia[204–206]. However, riboswitches have not yet been used in the engineering of these organisms. As discussed earlier, only a limited range of inducible systems is available for the engineering of clostridia; the use of synthetic riboswitches could be a promising alternative for the creation of controlled promoters.

1.8 Transcription

Transcription is the first stage of gene expression and the main stage for regulation of gene expression. Most existing biological parts in *Clostridium* fall within the transcription category and are mostly promoters (summarized in Table 1.1). In contrast, the termination of transcription has been studied relatively little and we present some analysis on the potential to study and improve the parts available for reliable termination.

1.8.1 Promoters

Promoters are DNA elements that are capable of driving transcription by RNA polymerase (RNAP) of downstream regions. Core promoter architecture in bacteria includes the -10 region or TATAAT box (Pribnow sequence), the -35 region and a spacer (with an optimal length of 17bp in *E. coli*) between the two conserved regions[207]. This motif is recognized by the housekeeping sigma factor that provides sequence specificity to RNAP (called RpoD in *E. coli* and SigA in *B. subtilis*, *C. acetobutylicum* and other solventogens)[208,209]. Both the consensus sequence and sigma factors bear very significant similarities between Firmicutes and *Escherichia*.

In addition, some *E. coli* promoters contain AT-rich UP elements (upstream of -35 region) that are responsible for recognition by the carboxy-terminal domain of the RNAP α-subunit[210], an additional *E. coli* promoter feature is the 'extended -10' region[211] (upstream of the -10 region and within the spacer). Interestingly, near-consensus promoters (including ones with UP elements) have been found to be significantly more common in Firmicutes (including Clostridia) than in other bacteria[212], a feature that was not explicable through higher AT-content alone. Recently, researchers demonstrated that random mutagenesis of *wt* clostridial promoters outside of the -10 and -35 boxes can also result in unexpected loss of activity, providing further evidence for a requirement for more extensive promoter sequence conservation in *Clostridium* than in *Escherichia*[117]. In Clostridia additional sequences with resemblances to UP elements (termed phased A-tracts) have been described[213] and a conserved extended -10 region that differs from its *E. coli* analogue has also been proposed[214]. The majority of promoters used in the genetic engineering of clostridia have been identified from the transcriptional units of important metabolic genes and most have been shown to be able to drive strong constitutive expression of a gene of

interest[215]. The most commonly used constitutive promoter in *C. acetobutylicum* (Table 1) is that of the native thiolase gene (*thlA*)[216,217], which has been used for the expression of genes involved in production of solvents such as butanol[218] and isopropanol[25], the transcriptional regulator gene *tetR*[152] (used a minimal promoter variant- *miniPthl*), and, in a modified form, for the expression of cellulosomal scaffoldins[219] and glycoside hydrolases[24,220]. The *C. acetobutylicum* thiolase enzyme catalyses the condensation of acetoacetyl-CoA from acetyl-CoA which are central metabolic intermediates[216]. The *C. acetobutylicum thlA* promoter (hereinafter referred to as *thlA*) is a SigA-dependent promoter (as evidenced by its near-consensus -35 and -10 regions[217]); however, in its normally chromosomal context it is also the subject of regulation by the redox-sensing transcriptional repressor Rex[221]. Rex inactivation was found to increase native *thlA* activity in *C. acetobutylicum* about 12-fold[222]. The Rex-binding site has been omitted from the core promoter in commonly used synthetic constructs[53]; yet this binding site is conserved in the promoters of orthologous genes from other solventogens (according to the RegPrecise database[223] and our promoter region alignments), suggesting that there are additional levels of control for this promoter that could be easily added by addition of the Rex sites if desirable. Indeed, some studies using similar promoters such as *thlA* and *ptb* have reported conflicting strength/activity findings, which may well be due to the cloning of regions of varying length (that likely include regulator binding sites such as Rex) and choosing to include the native RBSs or not (such as the ones reported between promoters in the Schulz 2013[101] and Girbal 2003[125] studies).

Thiolase promoters have also been used for expression in other clostridia such as *C. beijerinckii*[224] and *C. pasteurianum*[225]. However, gene expression data (RNAseq from *C. beijerinckii*) indicates that there are other genes that have higher expression levels

than the ones currently used as promoter sources; many of those genes encoded are hypothetical proteins[35]. Recently, Yang *et al.,* 2017 constructed a sequence logo of 18 promoters previously identified in *C. acetobutylicum* which revealed a strongly conserved -10 region. Using degenerate oligos to mutagenize the core *thlA* promoter elements, the authors were able to generate a promoter library of variable strengths, including mutations that increased the strength of the promoter, suggesting that *thlA* can be improved further using synthetic biology approaches.

The use of a constitutive promoter may not always be desirable; it may be preferable to use an inducible promoter, allowing controlled expression of a gene of interest. So far, the only naturally inducible promoters exemplified in a solventogenic clostridium are the *Staphylococcus xylosus xylA* promoter, which is repressed by XylR in the absence of D-xylose[125], and the *C. perfringens bgaL* promoter, which is repressed by BglR in the absence of lactose[82]. However, constitutive promoters can be made inducible by addition of operator sites (for transcriptional repressors) or by the addition of binding sites for activators, and several such promoters have been developed for use in solventogenic clostridia. LacI-repressible versions of thiolase and ferredoxin promoters have allowed the construction of clostridial expression constructs for genes where expression in *E. coli* would be toxic[23,24,220]. In *C. acetobutylicum*, the *fac* promoter (*Clostridium pasteurianum* ferredoxin promoter with *lac* operator) is able to function as an IPTG-inducible system in the presence of LacI, allowing approximately 10-fold induction[121]. A TetR-repressed, anhydrotetracycline-inducible promoter has also been used in *C. acetobutylicum*[152], generated by the fusion of the chloramphenicol acetyltransferase promoter *pcm* with the tetracycline operator *tetO*. This promoter could achieve up to 313-fold induction, although high levels of anhydrotetracycline were

inhibitory to growth. Another TetR-repressed promoter used in *Clostridia* is P*tet* (referred to in this work as *Pxyl/tetO* to reflect its origin and to distinguish it from other promoters known as P*tet*), it was generated by adding a tetO operator between the -35 and -10 boxes of a minimal version of the strong P*xyl* promoter of *B. subtilis* (excluding its native XylR operator)[226] and was further improved by mutagenesis[227]. This promoter has been used in *C. difficile*[228] and *C. acetobutylicum*[108]. However, operators are known to influence the basal activity of the promoter[229]. In fact, a recent study in *E. coli* found the core RpoD promoter too sensitive to sequence context and operator insertions to be a suitable target for forward engineering efforts and turned to extracellular sigma factors (ECFs) with T7 RNAP whose promoter core sequences they found to be more insensitive to operator addition[230]. To generate many novel synthetic hybrid regulated promoters, the researchers sought to combine the operators of various transcription factor repressors with core promoter elements and they found that the RpoD core promoter was more sensitive to sequence (the resulting promoters varied in activity – became stronger or weaker rather than merely regulated by the addition of operators) context than ECFs and T7 RNAP. An alternative strategy - CRISPR-mediated repression of transcription - has been demonstrated in several solventogenic species, namely C. *acetobutylicum*[87], *C. beijerinckii*[224] and *C. pasteurianum*[103],

Orthogonal expression systems, e.g. promoters that require other non-native elements for activity (commonly an alternative sigma factor or phage polymerase) are uncoupled to a certain degree from the evolutionary constraints and cellular regulation[231] and have also found widespread use in driving high levels of expression for heterologous protein production[232] and also the cloning of genes with toxic products[233]. This approach allows the total repression of genes until they are introduced into the organism of interest. A

commonly used example in *E. coli* is the T7 promoter, which requires the phage T7 polymerase for activity[234]. An example from Clostridia is expression driven from the *C. difficile tcdB* promoter that is dependent on the native sigma factor TcdR. The *tcdB* promoter is highly active in *C. acetobutylicum*, but only when *tcdR* has been integrated into the genome[121]. This enabled the high-level expression of a *mariner* transposon in *C. acetobutylicum* without negative effects on the *E. coli* cloning host. There is a strong case for designing these orthologonal-type systems as simply using a clostridial housekeeping promoter (such as the *thlA* promoter) with a clostridial ribosome binding site would not provide orthogonality in commonly used cloning hosts such as *E. coli* or *B. subtilis* due to the degree of conservation between the transcription and translation initiation systems amongst these organisms; on the contrary, as a rule strong clostridial promoters and RBSs (see translation initiation section) often retain their strength in *E. coli* whereas the opposite is observed more rarely, this is likely due to the on average higher similarity to the Bacteria-wide consensus translation and transcription initiation signals in *Clostridium*[160,214]. Interestingly, AT-rich DNA can be toxic to *E. coli* (clostridial genomes are very AT-rich) due to transcriptional activity from spurious intragenic promoters and RNAP titration[235]. This is exacerbated by the fact that the *E. coli* extended -10 region, which is sufficient alone for transcription initiation in *E. coli*, could also be present by chance in Clostridium-derived sequences, increasing the likelihood of spurious and unpredictable transcription during cloning in *E. coli*.

Table 1. 1 Promoters used in engineering of *Clostridium* species

Promoter	Comments
Constitutive	
P*thlA_Cac* (*C. acetobutylicum*)	Widely used for constitutive gene expression in *C. acetobutylicum,* for example expression of *C. beijerinckii ald*, *C. ljungdahlii bdh*[218], *C. acetobutylicum adc*, *ctfA*, *ctfB*,[25] and *tetR* (*miniPthl* variant)[152]. Activity analyzed using GusA[125] and several FbFP[101]reporters ; Two mutagenesis libraries based on this promoter were generated and produced a wide range of strengths, the mutants were active in *C. acetobutylicum*[117,156]*, C. ljungdahlii*[156] and *C. sporogenes*[117]
P*thl_Cperf* (*C. perfringens*)	Used for expression of *srtA* genes from *C. acetobutylicum*, *L. monocytogenes*, and *B. cereus*[24].
P*fdx_Cspo* (*C. sporogenes*)	Activity analyzed using CatP reporter[121]. Used in the ClosTron system for expression of the Ll.LtrB intron[236]. Used for expression of *spoA* integrated into the chromosome for complementation of a *spoA* mutant[59].
P*thl_Cbeij* (*C. beijerinckii*)	Used for dCas9 expression in *C. beijreinckii*[224]
P_{J23119} (synthetic)	Used to drive gRNA expression in *C. acetobutylicum*[87], *C. beijerinckii*[87] and *C. saccharoperbutylacetonicum*[89]
Regulated	
P*thl_lacO1:O1* (*thl* with double *lac* operator)	*lac*-repressed version of *thlA*. Used to express a miniscaffoldin *cipc1*[219] and weakened version for expression of mannanase *man5K*[220].
P*thlOid* (*thl* with single *lac* operator)	A *lac*-repressible version, used to drive chromosomal expression of the *C. cellulolyticum* glycoside hydrolase *xyn10A*[24].
P*fac* (single *lac* operator, derived from *C. pasteurianum*)	Activity analyzed using CatP reporter[79,121] and Pp2 FbFP[101]. Formerly used in the ClosTron system for expression of the Ll.LtrB intron[79]. Used for expression of *codA* in the *C. acetobutylicum* knockout vector pMTL-SC7515[59].
P*fdxOid* (*fdx* with single *lac* operator)	Used for chromosomal expression of *C. cellulolyticum* glycoside hydrolase *cel9G*[24]. Has RBS from *C. acetobutylicum thlA* promoter.

P*facOid* (*fac* with single *lac* operator)	Used for chromosomal expression of *C. cellulolyticum* glycoside hydrolase *cel48F*[24]. Has RBS from *C. acetobutylicum thlA* promoter.
P*tcdB* (*C. difficile*)	Developed for an orthogonal expression system, requires the exogenous sigma factor TcdR for function[121].
P*xylA* (*S. xylosus*)	Repressed by XylR and inducible by D-xylose; assessed with β-glucuronidase reporter giving 17-fold induction with D-xylose[125].
P*pcm* (*tetO1*-containing variants)	Repressed by TetR; inducible in the presence of anhydrotetracycline[152].
P*bgaL* (*C. perfringens*)	Activated by BgaR and inducible by lactose[82].
P*xyl/tetO* (*B. subtilis*)	Repressed by TetR[226,227]; activity in *C. acetobutylicum* shown by phiLOV 2.1 Opt FbFP expression[108]
Fermentation phase-specific	
P*adc* (*C. acetobutylicum*)	Activity assessed using β-glucuronidase[125], β-galactosidase[134] and Pp2 FbFP[101]. Primarily active after onset of solventogenesis.
P*ptb* (*C. acetobutylicum*)	Activity assessed using β-glucuronidase[125], luciferase[134] and Pp2 FbFP[101]. Active during acidogenesis (not solventogenesis)
P*ptb* (*C. beijerinckii*)	Activity assessed through expression of *lacI*[121].
P*sol* (*C. acetobutylicum*)	Activity assessed using β-galactosidase and luciferase reporters[134]. Weak expression, primarily active during late exponential phase.
P*bdhA* (*C. acetobutylicum*)	Analyzed using β-galactosidase reporter[134]. Comparatively weak expression, primarily active in early exponential phase.
P*bdhB* (*C. acetobutylicum*)	Analyzed using β-galactosidase reporter[134]. Primarily active until onset of solventogenesis.
P*hydA* (*C. acetobutylicum*)	Activity analyzed using β-glucuronidase[125] and Pp2 FbFP[101]. High activity during acidogenesis, decreases to low after

P*adhe2* (*C. acetobutylicum*)	phase shift Activity analyzed using Pp2 FbFP[101]. Strong expression during solventogenesis[101,237]

1.8.2 Transcriptional terminators

Bacteria have two distinct mechanisms that function in transcription termination. Both types of terminators are usually located in the 3' end of transcriptional units. Rho-dependent terminators rely on the Rho protein that recognizes a target sequence and causes RNA polymerase to fall off of the template DNA[238]. The specific DNA recognition sites, *rut* sites, have been used before in genetic circuits[239], but not extensively, perhaps due to the relatively poor mechanistic understanding of the process[240]. The Rho factor is a hexamer that recognizes C-over-G-enriched stretches of mRNA and contributes to transcriptional polarity by terminating coding transcripts that are not being actively translated[238]. Rho-dependent terminators appear rare in *Clostridium* species. The second mechanism, which also has been reported to be more widespread[241], is referred to as Rho-independent or intrinsic termination. Intrinsic terminators possess a conserved structure - a short GC-rich hairpin followed by a poly-U transcribed sequence. Several models of intrinsic termination have been proposed – amongst them are the thermodynamic/allosteric, RNA pullout/hybrid shearing and forward translocation models. In the thermodynamic/allosteric model, the hairpin stem loop sterically clashes with the RNAP exit pore - this is thought to lead to melting of the upstream end of the RNA:DNA hybrid and conformational changes in RNAP resulting in termination[242]. According to the hybrid shearing model (also referred to as the RNA pullout model), the hairpin folding and subsequent steric clash cause slippage of the RNA in the hybrid[243], transcript dissociation from the template is also facilitated by the fact that ribo-uracil-deoxyribo-adenine is the weakest nucleotide base pair[244]. In the forward translocation model, the hairpin formation moves RNAP forward up the template DNA (in the absence of further transcript elongation), this places the RNAP in

a ‘hypertranslocated’ state and shortens the hybrid[245]. Transcription termination models are not always mutually exclusive and some modes of termination are proposed to occur at certain types of intrinsic terminators, for example – hybrid shearing was proposed to occur at terminators with perfect or near perfect U-tracts whereas forward translocation would occur at ones with imperfect U-tracts[243,246]. Intrinsic terminators are often found downstream of operons; however, they are also involved in transcription attenuation when present within coding regions or downstream of promoters[247].

While a number of terminators have been used in the construction of clostridial expression constructs, often derived from clostridial genes such as *adc*[125,248], *fdx*[59], CD0164[59] and *thlA*[85,218], there have been few published analyses of terminator strengths in clostridia (**Table 1.2**). The *C. pasteurianum fdx* terminator has been shown to be highly effective at preventing read-through inhibition of the replicon from the *fac* promoter in a clostridial vector[121], and screening of a selection of terminators in *C. acetobutylicum*[24] showed that the *E. coli rrnB* terminator T1 loop was able to function as an efficient terminator, reducing expression of a downstream gene. However, these analyses only took into consideration the effect of the terminator on a downstream target under the influence of a single promoter. As previously mentioned, the introduction of a terminator may have an effect on expression of an upstream gene by influencing mRNA stability, which is not a desirable feature of a standard part[187]. However, this effect can be quantified in an appropriately designed assay[148]. Furthermore, the efficiency of termination may increase or decrease depending on promoter activity; this has been recently exploited for the development of a genetic band-pass filter in *E. coli*[249]. The band-pass filter is a genetic device that allows the induction of a target gene to be confined to a specific concentration range of an inducer molecule. In this case, this was achieved using a synthetic gene circuit which consisted

of the gene coding for the LuxI enzyme under the control of the *Ptet* promoter and the promoter-strength regulated terminator. LuxI synthesizes AHL which together with LuxR induces the *Plux* promoter which controls the synthesis of the target gene. TetR and LuxR are constitutively expressed.

Terminator strength has also been shown to be influenced by the hairpin's proximity to a stop codon or when present within a coding region[250]. A recent study demonstrated the gradual increase in termination efficiency as distance between the stop codon and hairpin increased[153]. The ribosome was found to repress transcriptional termination when the stop codon and hairpin were in close proximity – termination repression was relieved when insulator sequences of approximately 30 bp were used[153].

Given the low number of characterized terminators (**Table 1.2**), their arguably incomplete characterization and critically the observation that only one of those terminators has been found to be highly efficient (>90%) in a solventogen[24] (one other was qualitatively shown to reduce read-through[121]), expanding the range of characterized clostridial terminators appears necessary for the further development of a clostridial synthetic biology toolbox. This is because the availability of only two characterized high efficiency terminators constrains researchers to either designing single transcription unit (TU) constructs or risk introducing unexpected effects via read-through as uncharacterized terminators may or may not be strong or may not function as terminators at all (as shown later in this work). Even for simple shuttle vectors a single terminator is rarely sufficient as the vector backbones themselves already contain at least one transcription unit (that of the resistance gene) and in many cases a replication initiation protein as it the case with all the *Clostridium* replicons to the author's knowledge. The presence of a coding gene is not necessary for the existence of a transcription unit that requires to be insulated from the rest of the plasmid – the ColE1-

derived replicons do not require a plasmid-encoded replication factor for replication but initiation is controlled by two plasmid origin-encoded overlapping transcripts[51]. Arguably, the insertion of an independently transcribed gene/TU in the chromosome should warrant the availability of at least two characterized highly efficient terminators – to insulate from both upstream[251] and downstream readthrough (relative to direction of transcription from the promoter in the TU)[74]. In practice, double terminators (constructed of two individual terminators) have been used by some researchers[252] – increasing the demand to four individual terminators per TU integration. Reasons to construct a synthetic double terminator might be to lower transcriptional readthrough into and/or out of the TU and/or to create a synthetic bi-directional double terminator by combining two uni-directional terminators. A bidirectional terminator is defined as a terminator that functions when transcribed in either direction[253].

To discover new terminators, one approach that can be undertaken is the use of algorithms to extract putative terminator sequences from genomes, also known as 'part mining'. Several such bioinformatics tools exist; they rely on seed sequences, secondary RNA structure features or both[241,254–256]. These bioinformatics tools were developed to aid genome annotations by identifying the 3' ends of operons and may not detect all the features required by a functional terminator. Three of the more cited search algorithms are TransTermHP (TTHP)[255], RNIE[241] and WebGeSTer (WG)[257]. TTHP is a widely used tool, while RNIE's authors state that their tool eliminates false positives in comparison to TTHP[241]. Finally, WebGeSTer classifies the results into different types of intrinsic terminators based on overall secondary structure. They discovered that canonical U-tract containing intrinsic terminators (termed 'L-shaped') form the majority of structures found within Firmicutes, whereas they are a minority in the *E. coli* model where they were first identified[256]. One possibility is that this distribution is the

consequence of the low GC-content of Firmicute genomes. In addition, Actinobacteria (high GC-content) seem to have an inverse distribution. However, inspection of prediction data from Firmicute genomes with comparable GC-content to *E. coli* (50%) such as *Geobacilli* shows that predicted L-shaped terminators continue to be more numerous than I-shaped ones in the Firmicutes even at the higher genomic GC-content. Whether the differences between Firmicutes and Proteobacteria are statistically significant and whether there is a relationship between relative number of L-shaped terminators and GC-content within each clade should be a focus of future statistical analysis. A comparison of the GC-content distribution within the genomes would also be needed. It appears that GC-content plays a role but may not the sole factor in determining terminator type distribution – phylogeny may also play a role.

Table 1. 2 Terminators used in *Clostridium* species

Terminator Name	Organism of Origin	Characterization and source
T*thlA_Cac*	***C. acetobutylicum* ATCC 824**	Predicted[218]
T*thl_Cbeij*	***C. beijerinckii* NCIMB 8052**	Predicted – used in *C. beijerinckii*[85]
T*tracRNA*	***Streptomyces pyogenes***	Predicted[258] – used in *C. beijerinckii*[85]
T*thl_Cspba*	***C. saccharoperbutylacetonicum* N1-4 (HMT)**	Predicted (Hennessy *et al.*, unpublished)
T*adc*	***C. acetobutylicum* ATCC 824**	Predicted[259]
T*fdx*	***C. pasteurianum***	Autoradiogram of S1 nuclease mapping of the 3' end of *in vitro* and *in vivo* (both *E. coli* and *B. subtilis*-derived) produced *fdx* mRNA[214], evidence for reduced read-through in *C. acetobutylicum*[121]
T_{CD0164}	***Clostridioles difficile* 630**	Predicted[53]
T*slpA_Cdiff*_ext	***Clostridioles difficile* 630**	Predicted[228]
T*slpA_Cdiff*	***Clostridioles difficile* DSM 27639**	Western blot (of CipA2) in *C. acetobutylicum*[24] – poor efficiency
T*slpA_Lac*	***Lactobacillus acidophilus***	Western blot (of CipA2) in *C. acetobutylicum*[24] – poor efficiency
T*tyrS*	***B. subtilis***	Western blot (of CipA2) in *C. acetobutylicum*[24] – 50% efficiency
T*gyrA*	***B. subtilis***	Western blot (of CipA2) in *C. acetobutylicum*[24] – poor efficiency
T*pepN*	***B. subtilis***	Western blot (of CipA2) in *C. acetobutylicum*[24] - ~50% efficiency
T_1*rrnB* (BBa_B0010)	***Escherichia coli***	Multiple GFP/RFP ratios in *E. coli*[148]; Western blot (of CipA2) in *C. acetobutylicum*[24] - ~90%

		efficiency
T*phiTD1*	**Phage *φ*29 of *B. subtilis***	Autoradiogram of S1 nuclease mapping of the 3' end of *in vivo* produced mRNA *(B. subtilis*-derived)[260] ;Reduction in Km^R and reduction in APH-II activity in *S. lividans*[261]; Western blot (of CipA2) in *C. acetobutylicum*[24] – poor efficiency
BBa_B1010	***E. coli*** (modified T*thr* by replacing A-T bp with G-C in stem of hairpin)	GFP/RFP ratio in *E. coli*; improvement in upstream CAT activity in *C. acetobutylicum*[192]
T*hydA* (BBa_K2715014)	*C. acetobutylicum* ATCC 824	Predicted[262]

1.9 Discussion

While the development of biorenewables as an alternative to petroleum-derived commodity chemicals and fuels has resulted in the emergence of new markets[263], the success of industrial ABE fermentation relies on increasing productivity, broadening the range of feedstocks and improving tolerance to solvents and by-products are all existing challenges to achieving higher sustainability and ensuring the economic viability of *Clostridium*-derived biorenewables[264]. *Clostridium* species remain important hosts for the biological production of solvents and their further development relies on the adaptation of novel methodologies such as synthetic biology and metabolic engineering. Much progress has been made in the latter with several projects improving industrially relevant strains; however, the availability of biological parts with known behaviour is one of the limiting factors for the rate and scale of genetic designs with multiple

coordinated components such as genetic circuits[148]. The reasons for our anticipation are twofold: first, the physical availability of biological parts streamlines assembly of genetic pathways, and second, knowledge about parts' behaviour is crucial in predicting and analysing the behaviour of pathways and genetic circuits. This chapter provides a summary of the multiple areas where improved knowledge of parts would provide a better toolkit for synthetic biologists using these organisms and improve the rate at which genetic designs can be created, tested and improved.

In light of the small number of characterized or tested terminators in the *Clostridium* toolbox and specifically the industrially-relevant *C. saccharoperbutylacetonicum*, the aim of this research was to first identify suitable reporters of gene expression (I), to develop a sensitive transcriptional terminator reporter system (II), to test identify suitable terminator candidates for the *Clostridium* toolbox (III), to use a comparative characterization approach across several bacterial species in order to broaden basic knowledge about transcription terminaton (IV) and to identify both strong and unusual transcriptional terminators for improved genetic designs in *Clostridium* (V).

Chapter II

Materials and Methods

2. Materials and Methods

2.1 Growth media, antibiotics and additives

2.1.1 Growth media for *Escherichia coli*

Escherichia coli cultures were routinely grown in liquid Lysogeny Broth (LB)-Miller medium or on solid LB-Miller agar plates (prepared with the addition of 1.5% (w/v) agar) at 37°C, unless otherwise specified. The pH was adjusted to 7.5 prior to sterilization via autoclaving at 121°C for 15-20min. Double-distilled water (ddH_2O) was used for all media preparation. The resistivity of the double-distilled water was between 15MΩ × cm and 18MΩ × cm.

Lysogeny Broth-Miller

Tryptone		10g
NaCl		10g
Yeast extract		5g
ddH_2O	up to	1000ml

C Minimal Medium

NH_4Cl	10g
Na_2HPO_4	10g
KH_2PO_4	5g
NaCl,	3g
$MgSO_4$	0.25g
Casamino acids	0.002% (w/v)
Thiamine-HCl	1mg

Glucose	optional	30g
Glycerol		40g
ddH_2O	up to	1000ml

2.1.2 Growth media for *Bacillus subtilis*

B. subtilis was routinely cultured in liquid LB medium and on solid Lysogeny Agar (LA) medium. For inducing natural competency the following media were used.

SpC		**Volume Added**
10 x T-base		10ml
1M $MgSO_4$		0.1ml
50% (w/v) D-Glucose		1ml
10% (w/v) yeast extract		2ml
10% (w/v) casamino acids		0.25ml
Tryptophan (5mg/ml)		1 ml
ddH_2O	up to	100ml

SpII		**Volume Added**
10 x T-base		10ml
1M $MgSO_4$		0.35ml
50% (w/v) D-Glucose		1ml
10% (w/v) yeast		1ml

extract		
10% (w/v) casamino acids		0.1ml
Tryptophan (5mg/ml)		1 ml
ddH_2O	up to	100ml

10x T-Base	**Final Concentration**
$(NH_4)_2SO_4$	0.15M
K_2HPO_4	0.8M
KH_2PO_4	0.44M
$Na_3C_6H_5O_7$ (sodium citrate)	34mM
ddH_2O up to	1000ml

2.1.3 Growth media for *Clostridium saccharoperbutylacetonicum*

Cultures of *C. saccharoperbutylacetonicum* were grown in liquid Reinforced Clostridial Medium (RCM) or on solid RCM agar plates (prepared with the addition of 1.5% (w/v) agar) at 32°C, unless otherwise specified. RCM was purchased as ready-to-use powder from Merck. A modified version of RCM (RCM_m) was prepared from individual components, omitting agar from the liquid medium composition. RCM_m contained soluble starch that was pre-dissolved in boiling water as a stock of 20g/L.

Reinforced Clostridial Medium (modified)	
Meat extract	10g
Proteose Peptone	10g

Yeast extract		3g
NaCl		5g
D-Glucose		5g
Sodium acetate		3g
Cysteine-HCl		0.5g
Soluble Starch		1g
ddH_2O	up to	1000ml

Cultures of *C. saccharoperbutylacetonicum* were also grown in liquid *Clostridium* Growth Medium (CGM) or on solid CGM agar plates (prepared with the addition of 1.5% (w/v) agar) at 32°C. The original CGM recipe[216] was modified by the addition of sodium citrate and the omission of antifoam.

<u>*Clostridium* Growth Medium (modified)</u>

Yeast extract		5g
Asparagine		2g
NaCl		1g
K_2HPO_4-$3H_2O$		0.75g
KH_2PO_4		0.75g
$(NH_4)_2SO_4$		2g
$MgSO_4$-$7H_2O$		0.4g
$FeSO_4$-$7H_2O$		0.01g
$MnSO_4$-H_2O		0.01g
D-Glucose		5g
$Na_3C_6H_5O_7$ (sodium citrate)		0.5g
Cysteine-HCl		0.5g
ddH_2O	up to	1000ml

Cultures of *C. saccharoperbutylacetonicum* were also grown in liquid 2x Yeast-Tryptone (YT) or on solid 2xYT agar plates (prepared with the addition of 1.5% (w/v) agar) at 32°C.

2x Yeast-Tryptone

Tryptone		16g
NaCl		5g
Yeast extract		10g
ddH_2O	up to	1000ml

Cultures of *C. saccharoperbutylacetonicum* were also grown in liquid Minimal *Clostridium* Medium (MCM) at 32°C. The original recipe[265] was modified by the addition of sodium citrate, thiamine, biotin and p-amino benzoic acid.

Minimal *Clostridium* Medium (modified)

KH_2PO_4	0.5g
$MgSO_4$.7H2O	0.3g
$FeSO_4.7H_2O$	0.01g
$NH_4CH_3CO_2$ (ammonium acetate)	3g
$Na_3C_6H_5O_7$ (sodium citrate)	0.5g
Thiamine-HCl	1mg
Biotin	10µg
p-amino benzoic acid (PABA)	1mg

D-Glucose		20-40g
ddH_2O	up to	1000ml

2.1.4 Preparation of anaerobic media, storage of plates and use of oxygen indicators

Dissolved liquid medium components were placed in a round bottom flask with a rubber stopper and brought to a boil on a Bunsen burner flame while being sparged with anoxic N_2. When the redox/O_2 indicator resazurin was used (1mg/L final concentration), the solution was heated until the indicator turned colourless. During medium preparation the medium is reduced by heating; stoppering the bottle and sparging with N_2 to lower the influx of atmospheric gas into the vessel while the increased temperature of the solution prevents more atmospheric O_2 of dissolving into the solution. During the reduction of the medium the blue-coloured resazurin phenoxazine dye undergoes irreversible reduction to the pink-coloured (at neutral pH) and fluorescent resorufin dye[266]. At a reduction potential of $E_h \leq -110$ resorufin is reversibly reduced to the colorless and non-fluorescent dihydroresorufin, the reducing conditions correlate with low amounts of dissolved O_2. If O_2 is re-introduced and thus the redox potential rises - $E_h > -51$ the solution turns pink (dihydroresorufin is oxidized back to resorufin). The medium was then placed inside the anaerobic workstation where it was aliquoted into glass bottles or vials which were then sealed with aluminium crimp tops and butyl rubber stoppers prior to autoclaving. Solid medium was brought to a boil outside the anaerobic workstation and poured inside the anaerobic workstation where it was allowed to set before being stored at 4°C in sealed steel custom-made incubators, pressurized with anoxic N_2.

Heat-sensitive media and compounds were placed in the anaerobic workstation in containers with loosened caps at least 48 hours in advance of their use and filter sterilized prior to usage.

The aforementioned steel incubators were made by Mr Mark Bentley of the Biology Research Workshop, University of York. They were also used for culturing anaerobic strains at 32°C outside the anaerobic workstation as the anaerobic workstation was not temperature controlled.

The use of resazurin was avoided for fluorescence experiments as a precaution, except for the first microscopy experiments in Chapter 3 due to resorufin's fluorescence overlap with mCherry properties although autofluorescence fluorescence indicative of resorufin fluorescence was not observed in the controls in the aforementioned experiment.

2.1.5 Antibiotics

Antibiotics were kept as frozen stocks at -20°C. Prior to culture inoculation in liquid growth media, aliquots were thawed and appropriate amounts were added. The addition of antibiotics to solid media was done after autoclaving or melting in a microwave and once the temperature of the molten agar had decreased to approximately 55°C.

Antibiotic	**Stock concentration** [mg/mL] (solvent)	**Working concentration *(E. coli)*** [µg/mL]	**Working concentration *(B. subtilis)*** [µg/mL]	**Working concentration *(C. saccharoper-butylacetonicum)*** [µg/mL]
Ampicillin	50 (ddH_2O)	100	N.A.	N.A
Chloramphenicol	30-34 (EtOH)	15-30	5	N.A
Colistin	10 (ddH_2O)	50	N.A.	N.A.
Carbenicllin	50 (ddH_2O)	100	N.A.	N.A.
Erythromycin	100 (EtOH)	500	Not used	100 / 40 (solid)
Kanamycin	30 (ddH_2O)	60	N.A.	N.A.

Gentamycin	20 (ddH_2O)	20	Not used	Not used
Tetracycline	5-10 ($EtOH:ddH_2O$)	10	Not used	10
Spectinomycin	50 (ddH_2O)	100	Not used	Resistant
Thiamphenicol	25 (DMSO)	N.A	N.A	75

2.1.6 Additives

Additives were kept as frozen stocks at -20ºC. Prior to culture inoculation in liquid growth media, aliquots were thawed and appropriate amounts were added. Additives that serve as inducers of gene expression were added either at onset of culture growth or at a particular time point of growth. The addition of additives to solid media was done after autoclaving or melting in a microwave and once the temperature of the molten agar had decreased to approximately 55ºC.

Additive	**Stock concentration** [mg/mL unless otherwise specified] (solvent)	**Working concentration *(E. coli)*** [µg/mL unless otherwise specified]	**Working concentration *(B. subtilis)*** [µg/mL]	**Working concentration *(C. saccharoper-butylacetonicum)*** [µg/mL]
IPTG	1M (ddH_2O)	1mM	N.A.	N.A
X-Gal	50 (DMSO:EtoH)	20	5	N.A
X-Gluc	50 (DMSO:EtoH)	20	N.A.	N.A.
Magenta-Gluc	200 (DMSO:EtoH)	80	N.A.	N.A.
Anhydrotetracycline	0.1 (EtOH)	0.1-0.4	Not used	0.4
Doxycycline	0.4 (EtOH)	0.4	Not used	0.4

2.1.7 Bacterial culture preservation and revival

Bacterial strains were maintained as 12.5% (w/v) glycerol stocks at -80ºC in screw-top plastic tubes and cryovials.

Wild-type *Clostridium saccharoperbutylacetonicum* N1-4 (HMT) was received as lyophilized stocks (produced by Green Biologics, Ltd) in crimp top glass vials. To reconstitute the stocks, 2mL of RCM was used to aseptically resuspend the cell cake. These manipulations were performed in a laminar flow hood outside the anaerobic workstation. Cultures were inoculated with several different volumes of the suspension (usual inoculation volumes were between 0.5mL and 1mL) and incubated in air-tight serum bottles overnight at 32ºC. After overnight growth the cultures were sampled for OD_{600} and checked under microscope for motility and cell morphology (thin, motile rods – single cells and in chains) –phenotypes associated with both culture viability and pre-stationary phase cultures. Culture(s) displaying pronounced motility and the correct cell morphology were used for the preparation of glycerol stocks or sub-culturing.

For long term storage of *C. saccharoperbutylacetonicum* anaerobic glycerol stocks were also prepared in crimp top glass vials.

2.2 DNA cloning

2.2.1 Agarose gel electrophoresis

Tris-Borate-EDTA (TBE) (10x) was prepared with the following compositon – 108g/L of Tris, 55g/L orthoboric acid and 9.3g/L sodium EDTA (ethylenediaminetetraacetic acid) in ddH_2O. The stock solution was diluted 10-fold with ddH_2O to produce a working solution for agarose gel electrophoresis of DNA samples. Agarose gels were prepared by dissolving agarose (0.5-2g/L) in 1x TBE buffer by heating in a microwave

and poured directly into electrophoresis tank and allowed to set. SybrSafe stock solution (Invitrogen) was added 1 in 10 000 to the molten agarose before it began setting. Solid agarose gels were then submerged in 1x TBE buffer. DNA samples were mixed 1in 6 with 6x NEB Purple Loading Dye following the manufacturer's instructions.

2.2.2 Restriction digestion

Restriction digests were performed with enzymes purchased from NEB and Thermo Fischer, following the manufacturer's instructions unless otherwise stated.

2.2.3 Polymerase chain reaction (PCR)

PCR was used for amplification of DNA fragments for cloning. For cloning purposes – a high fidelity polymerase was used – Q5 High-Fidelity DNA polymerase was purchased from NEB and used according to the manufacturer's instructions. Primers were added to a final concentration of 0.5µM, dNTPs were added to a final concentration of 2mM and Q5 Reaction Buffer was added to a 1x final concentration, nuclease-free or double-distilled water was used to top up reactions. Q5 DNA polymerase Master Mix was also used for cloning where the dNTPs and buffer are pre-mixed.

PCR was also used for routine screening of strains for the transformation with DNA constructs in the form of colony PCR. Several DNA polymerases were used for this application - MyTaq DNA Polymerase and MyTaq DNA Polymerase Master Mix (Bioline Reagents), GoTaq DNA polymerase and GoTaq Green Master Mix (Promega) and PCRBIO HS Taq DNA Polymerase Master Mix Red and PCRBIO Ultra Mix Red (PCR Biosystems), according to the manufacturers' instructions.

2.2.4 Preparation of Plasmid DNA

Plasmid DNA was purified from bacterial strains using several commercially available kits – Miniprep kit (Qiagen), NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel) and

EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic), Wizard® Plus SV Minipreps DNA Purification Kit (Promega) and Monarch Plasmid Miniprep Kit (NEB), following the manufacturers' recommended protocols.

2.2.5 DNA purification from agarose gels

Gel extraction was the preferred method of DNA purification when more than one molecular species was present in the sample. DNA was purified from agarose gel slices using several kits – EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic), Zymoclean Gel DNA Recovery Kit (Zymo Research), Monarch DNA Gel Extraction Kit (NEB), QIAquick Gel Extraction Kit (Qiagen) and Wizard® SV Gel and PCR Clean-Up (Promega), following the manufacturers' recommended protocols. For gel extractions that were deemed more difficult due to low starting DNA concentrations, DNA electro-elution was performed. Briefly, the gel slice was placed in dialysis tubing, 1x TBE was added to tube and then the tube was clamped shut and placed in an electrophoresis tank. The tank was then ran at standard electrophoresis settings, the liquid content of the tube was aspirated and placed into an Eppendorf tube. The DNA was precipitated and purified using the ethanol precipitation method.

2.2.6 Ethanol Precipitation of DNA

Briefly, ethanol precipitation of DNA was performed using the following protocol – 0.1 volumes of $C_2H_3NaO_2$ (sodium acetate) were mixed with 2.5-3 volumes of ice-cold absolute ethanol and added to 1 volume of sample. The mixture was then placed at -20°C for 1 hour or overnight or at -80°C for 1 hour – the longer the incubation the more precipitate would form. The sample was then centrifuged at $2 \times 10^4 \times g$ for 30min at 4°C. The pellet was then washed twice with 0.5mL of ice-cold 75% (v/v) ethanol, centrifugation steps were 10min at 4°C. After the final wash step centrifugation, the

ethanol supernatant was discarded. The pellet was then air dried and resuspended in nuclease free water.

2.2.7 DNA purification from enzymatic reactions

DNA from enzymatic reactions was purified using DNA purification kits from the following manufacturers - DNA Clean & Concentrator Kits (Zymo Research) and DNA Gel Extraction Kit (Bio Basic), following the manufacturers' recommended protocols.

2.2.8 DNA purification from PCR reactions

DNA from polymerase chain reactions was purified using DNA Clean & Concentrator Kits (Zymo Research) and EZ-10 Spin Column PCR Products Purification Kit (Bio Basic), following the manufacturers' recommended protocols.

2.2.9 Genomic DNA purification from bacterial culture

Genomic DNA extraction from bacterial strains was performed using Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's recommended protocol.

2.2.10 Oligonucleotide cloning by partial annealing of ssDNA oligonucloetides

Reverse complimentary oligonucleotides were designed with 4 base overhangs. Oligonucleotides were re-suspended in ddH_2O to a concentration of 100µM. Phosphorylation reactions were set up using 0.8µL of each oligonucleotide, 1µL Thermofisher Polynucleotide Kinase (PNK), 2µL NEB T4 DNA Ligase Buffer (or 2µL PNK Buffer A and 2µL 10mM ATP) and ddH_2O up to 20µL. Reactions were incubated at 37ºC for 30min. One volume of 6xSSC Buffer and half a volume of 100mM EDTA were added to the reaction. The reaction was then heated up to 95ºC for 5min and was then cooled at 1ºC/s until the temperature reached 12ºC. This procedure was performed in a thermocycler. Alternatively, a heating block was kept at 95ºC for 5min and then

unplugged, the samples were kept on the heating block while it cooled down to room temperature (20ºC). Samples were then kept at 4ºC or -20ºC for longer term storage.

2.2.11 Polymerase Cycling Assembly (PCA) of dsDNA gBlocks into larger fragments

A polymerase chain assembly reaction was performed with overlapping dsDNA g Block fragments (supplied by IDT DNA). G block fragments were resuspended to a concentration of 50ng/µL. The fragments were cycled without the addition of primers in a thermocycler with normal PCR settings with an annealing temperature corresponding to the overlap regions of the fragments. Then that reaction was used as a template for PCR with a pair of primers unique to the ends of the assembled product.

2.2.12 Overlap extension for oligonucleotide cloning

An overlap extension PCR reaction was performed with overlapping ssDNA oligonucleotides designed to serve as both template and primer for amplification in order to clone short DNA fragments. The reaction conditions were the same as a normal PCR with an annealing temperature corresponding to the overlap region of the oligonucleotides. This was the preferred method for cloning terminators longer than 60bp, the assembled dsDNA contained restriction sites for Golden Gate assembly.

2.2.13 Golden Gate assembly

Golden Gate assembly reactions were performed with Type IIS restriction enzymes purchased from NEB or Thermo Fischer together with T4 DNA Ligase from NEB. The buffer used in all reactions was T4 DNA Ligase Buffer from NEB. The reaction was cycled between 2.5 min at 37ºC and 2.5 min at 16ºC for 30 cycles, followed by 10min at 55ºC and 20min at 80ºC in a thermocycler when BpiI, Esp3I or BsaI-HFv2 were used. When BsaI was used the incubation times at 37ºC and 16ºC were increased to

5min. If an internal restriction was present in the DNA fragment being cloned, the 55ºC step was omitted.

2.3 Bacterial transformation

2.3.1 Competent cell preparation and transformation of *Escherichia coli* by heat shock

Strains of *E. coli* were made competent using several protocols for chemically competent cells. The $CaCl_2$ protocol and the Hanahan competent cell protocol were used. The Mix and Go! E.coli Transformation Kit (Zymo Research) was used according to the manufacturer's protocol.

Briefly, the $CaCl_2$ protocol was performed as follows for same day transformations– an overnight of the desired strain was inoculated and grown overnight (preferably from a single colony, streaked onto a plate from a frozen stock). A flask with 20mL of LB was inoculated with 1mL of overnight culture. The culture was then grown at 37ºC until the OD_{600} reached 0.3-0.4. The culture was then transferred into 1mL aliquots and incubated on ice for 0.5-1 hour. The aliquots were then centrifuged for 10min at $3.5 \times 10^3 \times g$ at 4ºC. The supernatant was removed and the pellets were resuspended in 100µL of ice cold 100mM $CaCl_2$and incubated on ice for 20-60min. The suspensions were then centrifuged for 10min at $3.5 \times 10^3 \times g$ at 4ºC. The supernatant was removed and the pellets were resuspended in 50µL of ice-cold $CaCl_2$.

For larger batches of competent cells, the culture volume was increased to 250mL of LB which was inoculated with 2.5mL of overnight culture. After the second cell centrifugation step the pellet was resuspended in 8mL of ice-cold $CaCl_2$ with 20% (w/v) glycerol. Cells were aliquoted to into 100µL aliquots and snap-frozen in liquid N_2.

Briefly, the Hanahan method was used as follows, 200ml of LB culture in a large flask were inoculated to a starting OD_{600} of 0.1 with an overnight culture. The culture was

grown at 37ºC until an OD_{600} of 0.6 was reached when the culture was placed on ice. The cells were then aliquoted in two flasks and harvested by centrifugation for 15min at $4 \times 10^3 \times g$ at 4ºC. The supernatant was removed and the pellets were resuspended in 11mL of ice cold RF1 Buffer (100 mM KCl, 50 mM $MnCl_2.4H_2O$, 30 mM $C_2H_3KO_2$, 10 mM $CaCl_2.2H_2O$, 15% (w/v) Glycerol) and incubated on ice for 15-30min. The cells were then harvested by centrifugation for 15min at $3.5 \times 10^3 \times g$ at 4ºC. The supernatant was removed and the pellets were resuspended in 7.5mL ice-cold RF2 buffer (10 mM MOPS, pH 6.8, 10 mM KCl, 75 mM $CaCl_2.2H2O$, 15% (w/v) Glycerol). After incubation on ice for 15min, aliquots of 100µL were frozen at -80ºC.

All chemically competent *E. coli* strains were transformed using the heat shock procedure. Briefly, an appropriate amount of DNA was added to the competent cell aliquot (after thawing on ice for 15min) and the mixture was incubated on ice for 30min. Then the mixture was placed at 42ºC for 30s and then back on ice for 2min. Then 500µL of LB were added to the cells and they were incubated at 37ºC for 1-2 hours before plating either an aliquot or the entire suspension (after concentrating by centrifugation) on selective solid medium. The plates were then incubated at 37ºC overnight.

2.3.2 Electrocompetent cell preparation and electroporation of *Escherichia coli*

Strains of *E. coli* were made electrocompetent using the following protocol. Briefly, a culture of the desired *E. coli* strain was grown at 37ºC until an OD_{600} of 0.5-0.7. The cells were then harvested by centrifugation at $4 \times 10^3 \times g$ at 4ºC for 15min and washed in ice cold ddH_2O. The wash step was repeated twice and then the pellet was resuspended in 1 to 250 final volume of ice cold 12.5% (w/v) glycerol. The cells were then frozen at -80ºC. For electroporation, the cells were thawed at room temperature,

DNA was added to the cells and mixture was transferred to an electroporation cuvette. The cells were then pulsed with 1.8kV in a BioRad MicroPulser electroporator and immediately recovered in 1ml of LB medium. After 1 hour incubation at 37ºC with shaking (180rpm), an aliquot or the entire suspension (after concentrating by centrifugation) was plated on selective solid medium. The plates were then incubated at 37ºC overnight.

2.3.3 Conjugative transfer of plasmids from *Escherichia coli* to *Clostridium saccharoperbutylacetonicum*

The *E. coli* donor strain CA434 (HB101 carrying IncP-1β conjugative plasmid R702, Tet^R) was transformed with plasmids of interest. The transformants were selected on LB agar with the appropriate antiobiotic and tetracycline (10μg/mL). The tetracycline was later replaced for kanamycin (60 μg/mL). A 1mL aliquot of an overnight culture of the transformed donor strain was centrifuged at $6 \times 10^3 \times g$. The cell pellet was resuspended in 1mL of PBS buffer. The resuspension was then centrifuged at $6 \times 10^3 \times g$ again and the supernatant was removed. The cells were transferred to the anaerobic cabinet. The cell pellet was then resuspended in 200μL of overnight culture of *Clostridium saccharoperbutylacetonicum* N1-4 (HMT). The mixture was pipetted onto a non-selective agar plate and incubated for 4-8 hours at 32ºC. The cell mixture was then washed off the plate with 1mL of Phosphate-buffered Saline (PBS) buffer and a spreader. The mixture was then plated onto selective plates that included the appropriate antibiotic for plasmid selection and colistin at 50μg/mL for donor strain counter-selection. The plates were incubated at 32º for 24-72 hours until the appearance of single colonies. Strains were checked for the presence of plasmid and for species identity by colony PCR using plasmid and chromosome-specific primers.

Using this technique it was possible to use the IncP-1α mobilization plasmid pTA-Mob (in strain DH10B) for conjugative transfer to *Clostridium saccharoperbutylacetonicum* N1-4 (HMT). The major differences are that pTA-Mob is engineered from RK2, lacks an OriT so cannot be co-transferred and carries only gentamycin resistance while R702 is multi-drug resistant (kanamycin, streptomycin, sulphonamide and tetracycline). Plasmid pTA-Mob was a kind donation by Dr Rahmi Lale of the Norwegian University of Science and Technology.

2.3.4 Competent cell preparation of *Bacillus subtilis*

A culture of *B. subtilis* was streaked onto LB agar from a frozen glycerol stock and incubated overnight at 37ºC. The next day a colony was picked and re-suspended in 100µL of LB medium. Two dilutions of the re-suspension were prepared – 1:10 and 1:100 and 80µL of each was plated and incubated overnight at room temperature. The next day, the thinner of the two lawn plate cultures was washed off of the plate using 5ml of wash buffer and OD_{600} was recorded. The desired OD_{600} was less than 1.0. A liquid culture was inoculated using 25mL of SpC (Spizizen Competence) medium at a starting OD_{600} of 0.01. The culture was incubated at 37ºC with shaking in a sterile 250mL flask. Growth was monitored at OD_{600} and plotted, when the culture reached stationary phase, incubation was continued for a further 2 hours when pre-warmed SpII (Spizizen II) medium was inoculated 1 in 10, preferably to 25mL final volume of culture. The SpII medium culture was then incubated at 37ºC for 90min, with shaking, in 250mL flasks. The cells were collected by centrifugation and the supernatant was retained. The cells were resuspended in 1/10 the original culture volume in retained supernatant. The cells were diluted with sterile 50% (w/v) glycerol to a final 12.5% (w/v) glycerol concentration and aliquoted into 400µL aliquots and stored at -80ºC until needed.

Wash Buffer		**Volume Added**
10 x T-base		10 ml
1M $MgSO_4$		100 µl
ddH_2O	up to	100ml

2.3.5 Transformation of *Bacillus subtilis*

Aliquots prepared as described in Section 2.3.4 were thawed at 37ºC and then 100µL aliquots were mixed with an equal amount of 1x T-Base 2mM EGTA in 2mL micro-centrifuge tubes. Plasmid DNA was added to the mixture and the tubes were incubated at 37ºC with shaking at 180 rpm for 20min. The mixture was then plated onto LB agar plus appropriate antibiotic. The plates were then incubated at 37ºC overnight until single colonies appeared. Negative controls without the addition of DNA were also always performed in parallel.

2.3.6 Electrocompetent cell preparation and electroporation of *Clostridium saccharoperbutylacetonicum*

A starter culture was inoculated from a -80ºC 16.6% (w/v) glycerol stock or from a lyophilized bacterial pellet (kept at 4ºC in the dark) in RCM or RCM_m. The starter culture was grown overnight at 32ºC. The starter culture was then used to inoculate CGM or CGM_m medium in a 1 to 10 ratio. The OD_{600} of the CGM culture was monitored every 40 – 60min. When the OD_{600} reached 1.2, the culture was transferred into an anaerobic workstation and then pelleted by centrifugation at 4ºC and $4 \times 10^3 \times g$

for 10min. In the anaerobic workstation, the supernatant was removed and the pellet was re-suspended in 25% of the original volume in freshly mixed Electroporation Wash Buffer (EWB) (294mM sucrose, 0.6mM Na_2PO_4, 4.4mM NaH_2PO_4, 10mM $MgCl_2$, pH 6.0). The $MgCl_2$ and $NaPO_4$ buffer were kept as 1M and 0.5M stocks, respectively, inside the anaerobic workstation and were mixed with anaerobic 300mM sucrose and anaerobic ddH_2O which were also kept in the workstation, prior to the wash steps. After centrifugation at 4°C and $4 \times 10^3 \times g$ for 10min, the pellet was re-suspended in 1.66% of the original culture volume of Electroporation Buffer (EB) (294mM sucrose, 0.6mM Na_2PO_4, 4.4mM NaH_2PO_4, pH 6.0), most commonly 1ml EB for 60ml culture or 2ml for 120ml culture. Aliquots of 200µL were transferred into 2mm gap electroporation cuvettes into which plasmid DNA had been pipetted prior and which were transferred to the anaerobic workstation prior to the start of cell washing. After 5min on ice, the cuvettes were electroporated with settings of 1.5kV voltage, 25µF capacitance and 200Ω resistance (a Bio-Rad Gene Pulser electroporation system with a Pulse Controller attachment plugged in was used at the University of York, alternatively, a Bio-Rad Micropulser electroporation system which has a pre-set 10µF capacitor and two resistors at 600Ω and 30Ω, respectively, was used at Green Biologics Ltd(GBL)). Following electroporation, the cells were re-suspended in 1ml CGM as quickly as possible (after transfer back through the airlock in the anaerobic workstation (approx. 5min) at the University of York or immediately at GBL). After 18h of incubation at 32°C, 200µL of the cultures were plated onto selective CGM plates and incubated at 32°C for 1-3 days.

2.4 Oligonucleotides used in this study.

The oligonucleotides used in this study are listed in Table 2.1.

Table 2.1 Oligonucleotides used in this study (for PCR and terminator library assembly).

Oligo Code	Oligo Verbose Name	Sequence (5' to 3')
IG0001	>35_CATERM_3401-REV_F	AGGTTAAGTTGCATAAAAAGAAGCTGTCCAGAGTAGGACAGCTTCTTTTTAATTAAATAT
IG0002	>35_CATERM_3401-REV_R	AAGCATATTTAATTAAAAAGAAGCTGTCCTACTCTGGACAGCTTCTTTTTATGCAACTTA
IG0003	>38_CbeijTERM_4660-REV_F	AGGTATGACCTTTGTTAAATATAAAAATGCCCCAATAAAGGGGCATTTTTATGTATGTGATAATTAAT
IG0004	>38_CbeijTERM_4660-REV_R	AAGCATTAATTATCACATACATAAAAATGCCCCTTTATTGGGGCATTTTTATATTTAACAAAGGTCAT
IG0005	>40_CsbuTERM_20-REV_F	AGGTGAATTTTTTATTCTATAAATAAAGGGATATATCATTTTAAAGATATATCCCTTTATTAATTATATTAACAAGAA
IG0006	>40_CsbuTERM_20-REV_R	AAGCTTCTTGTTAATATAATTAATAAAGGGATATATCTTTAAAATGATATATCCCTTTATTTATAGAATAAAAAATTC
IG0007	>42_CsbuTERM_339-REV_F	AGGTCTTGTTTATCTATAAGAAGAGGCTATTTCAAATGAAATAGCCTCTTCTTATAGATATTAAAA
IG0008	>42_CsbuTERM_339-REV_R	AAGCTTTTAATATCTATAAGAAGAGGCTATTTCATTTGAAATAGCCTCTTCTTATAGATAAACAAG
IG0009	>55_cwpV_F	AGGTAGTATAAATAATAAAATTTATTTGTAAAAAGAGTAGCACTTTTAATTCTAAAGGCTACTTTTTTTAT
IG0010	>55_cwpV_R	AAGCATAAAAAAAGTAGCCTTTAGAATTAAAAGTGCTACTCTTTTTACAAATAAATTTTATTATTTTATACT
IG0011	0_Vector phiLOV_2.1 FWD	TCACAGGGAAGACTGTCCGCTTCCTCGCTCACT
IG0012	0_Vector phiLOV_2.1 REV	TCACAGGGAAGACTGATAGGATCCTTAAACATGATCTGAACC
IG0013	1_Insertion Site Oligo FW	TCACAGGGAAGACCACTATAGTATCTGAATCGAAAGGTTGAGACCAACTGACGCAGTAGGTCTCCGCTTTAGTGGTACTAGAGATTTGGTCTTCCCTGTG
IG0014	1_Insertion Site Oligo REV	TCACAGGGAAGACCAAATCTCTAGTACCACTAAAGCGGAGACCTACTGCGTCAGTTGGTCTCAACCTTTCGATTCAGATACTATAGTGGTCTTCCCTGTG

IG0015	1st Mut FW FULL	AGAAGGTAGACCTTATGAAGGTACTCAAACAGCTAAATTAAAAG
IG0016	1st Mut REV FULL	TTCATAAGGTCTACCTTCTCCTTCTCCTTCTATTTCAAATTCAT
IG0017	2_CatP Plus FWD	TCACAGGGAAGACACGATTTGAGAGGGAACTTAGATGG
IG0018	2_CatP Plus REV	TCACAGGGAAGACACTAACTTAACTATTTATCAATTCCTGCAATTC
IG0019	2_mCherry_FW	TCACAGGGAAGACACGATTAAAGAGGAGAAATACTAGATGGTGAGC
IG0020	2_mCherry_REV	TCACAGGGAAGACACTAACGTATTATTACTTGTACAGCTCGTCCATGC
IG0021	2nd Mut FW FULL	AAGGTAGACATTCAACAGGAGGTATGGATGAATTATATAAATAAGGA
IG0022	2nd Mut REV FULL	CTGTTGAATGTCTACCTTCTGCTCTTTCATATTGTTCAACTAT
IG0023	3_traJ Module FWD ACTUAL	TCACAGGGAAGACGTGTTAATTTTAATAAAACTTTAAATAGAAAAAGGCTTCTCTCATGAGAAG
IG0024	3_traJ Module REV	TCACAGGGAAGACGTGCCTTCGCCCTTCCTGCTTCG
IG0025	34_tThi_spa_FW	ATGAAGACACAGGTGTATACAAGTTCACATTCGCAACAAGTTACTATGATAAGATATATTATCATAGTAACTTTTTTATATAATA
IG0026	34_tThi_spa_REV	ACGAAGACACAAGCCAATCACTCACTAATTCAAATACTATTTTCCTCTTTAGTATACTCTTTTATACCGAATCAATTAAAATTTTTATTATATAAAAAAG
IG0027	35_CATERM_3401_FW	AGGTATATTTAATTAAAAAGAAGCTGTCCTACTCTGGACAGCTTCTTTTTATGCAACTTA
IG0028	35_CATERM_3401_REV	AAGCTAAGTTGCATAAAAAGAAGCTGTCCAGAGTAGGACAGCTTCTTTTTAATTAAATAT
IG0029	36_CATERM_42_FW	AGGTGAGTATGGAGTAAAATAGTAAAATAAATGTGCCTCAACTTAGATGTTAAGGCACATTTATTTTTATATATTATTCATGTTTT
IG0030	36_CATERM_42_REV	AAGCAAAACATGAATAATATATAAAATAAATGTGCCTTAACATCTAAGTTGAGGCACATTTATTTTACTATTTTACTCCATACTC
IG0031	37_CspaTERM_560_FW	AGGTTTTTGTAACAGCTCATTTTTTTATATAATTAATGTGAGTTCTTAATTGAATTCACATTAATTATATAAGATAATATATTTGTATTAAG
IG0032	37_CspaTERM_560_REV	AAGCCTTAATACAAATATATTATCTTATATAATTAATGTGAATTCAATTAAGAACTCACATTAATTATATAAAAAAATGAGCTGTTACAAAA
IG0033	38_CbeijTERM_4660_FW	AGGTATTAATTATCACATACATAAAAATGCCCCTTTATTGGGGCATTTTTATATTTAACAAAGGTCAT
IG0034	38_CbeijTERM_4660_REV	AAGCATGACCTTTGTTAAATATAAAAATGCCCCAATAAAGGGGCATTTTTATGTATGTGATAATTAAT

IG0035	39_CsbuTERM_1093_FW	AGGTATTTAAAATATATTTTAGAAAATAAAAGTACTCTTATTGCTACATTGCAATAAGAGTACTTTTATTTTCGGTTAGTTATAATATAA
IG0036	39_CsbuTERM_1093_REV	AAGCTTATATTATAACTAACCGAAAATAAAAGTACTCTTATTGCAATGTAGCAATAAGAGTACTTTTATTTTCTAAAATATATTTTAAAT
IG0037	4_ErmB Module FWD	TCACAGGGAAGACACAGGCCGGCCGAAGCAAACT
IG0038	4_ErmB Module REV	TCACAGGGAAGACACCGGACAAGCAGCAGATTACGCG
IG0039	40_CsbuTERM_20_FW	AGGTTTCTTGTTAATATAATTAATAAAGGGATATATCTTTAAAATGATATATCCCTTTATTTATAGAATAAAAAATTC
IG0040	40_CsbuTERM_20_REV	AAGCGAATTTTTTATTCTATAAATAAAGGGATATATCATTTTAAAGATATATCCCTTTATTAATTATATTAACAAGAA
IG0041	41_CsbuTERM_674_FW	AGGTTGACATTATCCTTAAGAAGCGCTAAAATTTGCGCTTCTTTCTCATTTGTATG
IG0042	41_CsbuTERM_674_REV	AAGCCATACAAATGAGAAAGAAGCGCAAATTTTAGCGCTTCTTAAGGATAATGTCA
IG0043	42_CsbuTERM_339_FW	AGGTTTTTAATATCTATAAGAAGAGGCTATTTCATTTGAAATAGCCTCTTCTTATAGATAAACAAG
IG0044	42_CsbuTERM_339_REV	AAGCCTTGTTTATCTATAAGAAGAGGCTATTTCAAATGAAATAGCCTCTTCTTATAGATATTAAAA
IG0045	43_CATERM_3875_FW	AGGTTGATAAAATCTCGTAATTATATAAAGAAAAAGCAAGAGAACTCTCTTGCTTTTTCTTTATATAGCTTCATATCCTATGGC
IG0046	43_CATERM_3875_REV	AAGCGCCATAGGATATGAAGCTATATAAAGAAAAAGCAAGAGAGTTCTCTTGCTTTTTCTTTATATAATTACGAGATTTTATCA
IG0047	44_CsbuTERM_2221_FW	AGGTTTTCCTTTACAAAAATGAAGACTCTAGTGAAATCACTAAAGTCTTCATTTTTAAATGTGTAT
IG0048	44_CsbuTERM_2221_REV	AAGCATACACATTTAAAAATGAAGACTTTAGTGATTTCACTAGAGTCTTCATTTTTGTAAAGGAAA
IG0049	45_CATERM_1157_FW	AGGTTAAAAATAAGAACCTATTAAAATAAATGGGCGCTGCTTTAATGCAGTGTCCATTTATTTTAAGTTTAAATAGAAACGT
IG0050	45_CATERM_1157_REV	AAGCACGTTTCTATTTAAACTTAAAATAAATGGACACTGCATTAAAGCAGCGCCCATTTATTTTAATAGGTTCTTATTTTTA

IG0051	46_CspaTERM_850_FW	AGGTCCAAATTTACAAGGAGTTATAAAGTTAAACCACTACGTTTGTAGTGGTTTAACTTTATAATAATGAATAAGCTTAT
IG0052	46_CspaTERM_850_REV	AAGCATAAGCTTATTCATTATTATAAAGTTAAACCACTACAAACGTAGTGGTTTAACTTTATAACTCCTTGTAAATTTGG
IG0053	47_CA_plasTERM_14_FW	AGGTTTGGATGATGATGAATGATAAAAAAATAGGACCTTTTGGTCCTATTTTTTTATATAAACTTATATAATT
IG0054	47_CA_plasTERM_14_REV	AAGCAATTATATAAGTTTATATAAAAAAATAGGACCAAAAGGTCCTATTTTTTTATCATTCATCATCATCCAA
IG0055	48_CspaTERM_1466_FW	AGGTTATTGGTTAATAAAATAGGTACCTTAAAACTACTATCTTGTAATTTTAAGATACCTATTTTGAAATTTACAA
IG0056	48_CspaTERM_1466_REV	AAGCTTGTAAATTTCAAAATAGGTATCTTAAAATTACAAGATAGTAGTTTTAAGGTACCTATTTTATTAACCAATA
IG0057	49_CspaTERM_2009_FW	AGGTCCATTTATGATAAAATAAGCACTAACTTTTGTTAGTGCTTATTTTATCATAAATAG
IG0058	49_CspaTERM_2009_REV	AAGCCTATTTATGATAAAATAAGCACTAACAAAAGTTAGTGCTTATTTTATCATAAATGG
IG0059	50_CPATERM_1752_FW	AGGTCTTTGGATAGAAGAATAAAGTGCATCTGTAATTTGATGCACTTTATTTTTAATTTATTTA
IG0060	50_CPATERM_1752_REV	AAGCTAAATAAATTAAAAATAAAGTGCATCAAATTACAGATGCACTTTATTCTTCTATCCAAAG
IG0061	51_CATERM_293_FW	AGGTTGCTGTCATAAAAAAGTACCAGCCATTAATTAGGCTGGTACTTTTTATCCATCATT
IG0062	51_CATERM_293_REV	AAGCAATGATGGATAAAAAGTACCAGCCTAATTAATGGCTGGTACTTTTTTATGACAGCA
IG0063	52_CPATERM_1881_FW	AGGTAATTTATTACCAAAATTTAAAAGTACATTCTTACTAAATTGTAAGAATGTACTTTTAAATTGTAAGGAAATAGA
IG0064	52_CPATERM_1881_REV	AAGCTCTATTTCCTTACAATTTAAAAGTACATTCTTACAATTTAGTAAGAATGTACTTTTAAATTTTTGGTAATAAATT
IG0065	53_CbeijTERM_773_FW	AGGTAGGAACCATATAAAAGAAAGGATACTTAGTTGCTTGATATTGCAACTAAGTATCCTTTCTGTCTATTAACCGTT

IG0066	53_CbeijTERM_773_REV	AAGCAACGGTTAATAGACAGAAAGGATACTTAGTTGCAATATCAAGCAACTAAGTATCCTTTCTTTTATATGGTTCCT
IG0067	54_CsbuTERM_2075_FW	AGGTTTATCTATATAAAAATAAGGATGAACTATTGCTCATCCTTATTTTTATATTATAAT
IG0068	54_CsbuTERM_2075_REV	AAGCATTATAATATAAAAATAAGGATGAGCAATAGTTCATCCTTATTTTTATATAGATAA
IG0069	AmyE FW	ATGTTTGCAAAACGATTCAAAACCTC
IG0070	AmyE REV	TCAATGGGGAAGAGAACCGC
IG0071	BBa_B0010_FW	AGGTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTC
IG0072	BBa_B0010_REV	AAGCGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGG
IG0073	Bba_B0062m_FW	AGGTATTTCAGATAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCT
IG0074	Bba_B0062m_REV	AAGCAGAAATCATCCTTAGCGAAAGCTAAGGATTTTTTTTATCTGAAAT
IG0075	Bba_hisoperon*_FW	AGGTTCCGGCAAAAAAGGGCAAGGTGTCACCACCCTGCCCTTTTTCTTTAAAACCGAAAAGA
IG0076	Bba_hisoperon*_REV	AAGCTCTTTTCGGTTTTAAAGAAAAAGGGCAGGGTGGTGACACCTTGCCCTTTTTTGCCGGA
IG0077	C_term FW	ATAGCATTTTGGATTGAAACTATTAATGAGGCGGGTGTAATTAATATTAAAGTAATATCCACATCCTTGGGGCAAAGAATGTATGATAAAAGTTAA
IG0078	C_term REV	ACCTTTAACTTTTATCATACATTCTTTGCCCCAAGGATGTGGATATTACTTTAATATTAATTACACCCGCCTCATTAATAGTTTCAATCCAAAATG
IG0079	CD0164_FW	AGGTATAAAAAAATTGTAGATAAATTTTATAAAATAGTTTTATCTACAATTTTTTTAT
IG0080	CD0164_REV	AAGCATAAAAAAATTGTAGATAAAACTATTTTATAAAATTTATCTACAATTTTTTTAT
IG0081	Clo_Lac_DOWN_Seq	TTGGAGCCTGTGACATTTTG
IG0082	Clo_Lac_SacI-A	TTTGAGCTCCTGCAGTAAAGGAGAAAATTTTATGAGGAAGATTATTCCTA
IG0083	CLo_Lac_SeqFW	AGTTCAACACTTATTTAGACGCAG
IG0084	CLo_Lac_SeqREV	GTTCCATCTGCCTTTCTTCG
IG0085	Clo_Lac_UP_SEQ	TGCGTCTAAATAAGTGTTG
IG0086	Clo_Lac-GEN_BamHI_2	ATAGGATCCACCTTTATTCAACTTTTATCATTAC
IG0087	Clo_LacZ FW	AGGGAAGACACAATGAGGAAGATTATTCCT

IG0088	Clo_LacZ REV	TCCGAAGACACACCTCTATGCTCCA
IG0089	Clo_LacZ REV+C-term	TCCGAAGACACCTATTGCTCCACCTATGAGTGCTA
IG0090	Clo_LacZ_BamHI	ATAGGATCCTTAACTTTTATCATACATTCTTT
IG0091	Clo_LacZ_D_full	AGGGAAGACACACCTTTAACTTTTATCATACATTCTTT
IG0092	Clo_LacZ_Gen_D-REV	AGGGAAGACACACCTTTATTCAACTTTTATCATTAC
IG0093	Clo_LacZ_SacI	TTTGAGCTCCTGCAGTAAAGGAGAAAATTTTAATGAGGAAGATTATTCCT
IG0094	Clo_LacZGEN_BamHI	ATAGGATCCAGATGAAATTCTCTTTCTGT
IG0095	Clo_LacZGEN_SacI	TTTGAGCTCCTGCAGTAAAGGAGAAAATTTTATGAGAAAGATTATTCCTAT
IG0096	Clo_LacZGEN_Seq_Fw	CACATACTTAGATGCAGTAAAAG
IG0097	Clo_LacZ-Gen-C_FW	AGGGAAGACACAATGAGAAAGATTATTCCTATTAATAATAATTG
IG0098	D122E FW	AAGATGGAGAATTTATATATAAAGTTAAATTAAGAGGTAC
IG0099	D122E REV	ATATAAATTCTCCATCTTGTAATGAACTATCTTGTGT
IG0100	DVA_A Forward	GTACTAGTGGGTCTCAGGAGATGTCTTCTGCACCATATGCGGT
IG0101	DVA_C Reverse	CTACTAGTAGGTCTCTCATTACGTCTTCCCCGCGCGTTGGCCGAT
IG0102	ECK120010783_FW	AGGTACGAGCCAATAAAAATACCGGCGTTATGCCGGTATTTTTTTACGAAAGA
IG0103	ECK120010783_REV	AAGCTCTTTCGTAAAAAAATACCGGCATAACGCCGGTATTTTTATTGGCTCGT
IG0104	ECK120010800_FW	AGGTAGTTTGTTCGCCCGGTAGTTGTGACGCTACCGGGTTCTTTTCGA
IG0105	ECK120010800_REV	AAGCTCGAAAAGAACCCGGTAGCGTCACAACTACCGGGCGAACAAACT
IG0106	ECK120015170_FW	AGGTACAATTTTCGAAAAAACCCGCTTCGGCGGGTTTTTTTATAGCTAAAA
IG0107	ECK120015170_REV	AAGCTTTTAGCTATAAAAAAACCCGCCGAAGCGGGTTTTTTCGAAAATTGT
IG0108	ECK120029600_FW	AGGTTTCAGCCAAAAAACTTAAGACCGCCGGTCTTGTCCACTACCTTGCAGTAATGCGGTGGA CAGGATCGGCGGTTTTCTTTTCTCTTCTCAA
IG0109	ECK120029600_REV	AAGCTTGAGAAGAGAAAAGAAAACCGCCGATCCTGTCCACCGCATTACTGCAAGGTAGTGGAC AAGACCGGCGGTCTTAAGTTTTTTGGCTGAA
IG0110	ECK120033127_FW	AGGTTACTTCTTACTCGCCCATCTGCAACGGATGGGCGAATTTATACCC
IG0111	ECK120033127_REV	AAGCGGGTATAAATTCGCCCATCCGTTGCAGATGGGCGAGTAAGAAGTA

IG0112	ECK120033737_FW	AGGTGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGG
IG0113	ECK120033737_REV	AAGCCCTTTGGTCGAAAAAAAAAGCCCGCACTGTCAGGTGCGGGCTTTTTTCTGTGTTTCC
IG0114	ECK125108723_FW	AGGTTAGCGTAAAAGCAAAACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCG GTTTTTTTTCGCCGTCATA
IG0115	ECK125108723_REV	AAGCTATGACGGCGAAAAAAAACCGCCAGTAAACCGGCGGTGAATGCTTGCATGGATAGATTT GTGTTTTGCTTTTACGCTA
IG0116	EcoT1_FW	AGGTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGT TGTTTGTCGGTGAACGCTCTC
IG0117	EcoT1_REV	AAGCGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCT TTCGTTTTATTTGATGCCTGG
IG0118	FOR_SEQ	CGGGTTGTTAAACCTTCG
IG0119	FOR2 SEQ	GACTTTAAGCCTACGAATACC
IG0120	Fw external Spacer BsaI	AGAAGACATTAAGGATCCTATAGTATCTGAATCGAAAGGTCGAGACCAACTGACGCAGTA
IG0121	Fw external Spacer Esp	AGAAGACATTAAGGATCCTATAGTATCTGAATCGAAAGGTCGAGACGAACTGACGCAGTA
IG0122	FW internal Spacer BsaI	GACCAACTGACGCAGTAGGTCTCCGCTTCGGCAAGGTCGTGAAACCAAGCCGCTAGGTCCGGT AATGCGGAATCGAC
IG0123	FW internal Spacer Esp	GACGAACTGACGCAGTACGTCTCCGCTTCGGCAAGGTCGTGAAACCAAGCCGCTAGGTCCGGT AATGCGGAATCGAC
IG0124	gus 2nd FW	AGAAGACATAATGTTACGTCCTGTAGAAACCCC
IG0125	gus 2nd REV	AGAAGACATACCTTTATTGTTTGCCTCCCTGCTGCG
IG0126	gus Vector REV	AGAAGACATCTTATTGTTTGCCTCCCTGCTGCG
IG0127	GusA_DOWN_Seq	CAACGCGTAAACTCGAC
IG0128	GusA_Seq_Fw	ACTTTGCAAGTGGTGAATC
IG0129	GusA_Seq_Rev	GCTTCGAAACCAATGCCTA
IG0130	GusA_UP_Seq	CGATGTATCCCGGCATAG
IG0131	GusA-REV-BamHI	ATAGGATCCACCTTTATTGTTTGCCTCCCTGCTGCG
IG0132	L1U1H08_FW	AGGTCCCGCATGTTCGCATGCGGGTTTTTTTTT
IG0133	L1U1H08_REV	AAGCAAAAAAAAACCCGCATGCGAACATGCGGG

IG0134	L1U1H09_FW	AGGTCGACGATGTTCGCATCGTCGTTTTTTTTT
IG0135	L1U1H09_REV	AAGCAAAAAAAAACGACGATGCGAACATCGTCG
IG0136	L2U2H09_FW	AGGTACGGCCCTCGCAAGGGCCGTTTTTTTGT
IG0137	L2U2H09_REV	AAGCACAAAAAAACGGCCCTTGCGAGGGCCGT
IG0138	L3S1P11_FW	AGGTGACGAACAATAAGGCCTCCCTTCGGGGGGGCCTTTTTTATTGATACAAAA
IG0139	L3S1P11_REV	AAGCTTTTGTATCAATAAAAAAGGCCCCCCCGAAGGGAGGCCTTATTGTTCGTC
IG0140	L3S1P32_FW	AGGTGACGAACAATAAGGCCTCCCAAATCGGGGGGCCTTTTTTATTTTTCAACAAAA
IG0141	L3S1P32_REV	AAGCTTTTGTTGAAAAATAAAAAGGCCCCCCGATTTGGGAGGCCTTATTGTTCGTC
IG0142	L3S1P51_FW	AGGTAAAAAAAAAAAAGGCCTCCCAAATCGGGGGGCCTTTTTTATTGATAACAAAA
IG0143	L3S1P51_REV	AAGCTTTTGTTATCAATAAAAAAGGCCCCCCGATTTGGGAGGCCTTTTTTTTTTTT
IG0144	L3S2P00 FW	TCACAGGGAAGACACAGGTACTCGGTACCAAATTCCAGAAAAGAGGGGAGCGGGAAACCGCTC CCCTTTTTTCGTTTTGGTCCGCTTGTGTCTTCAGTGTCC
IG0145	L3S2P00 REV	GGACACTGAAGACACAAGCGGACCAAAACGAAAAAAGGGGAGCGGTTTCCCGCTCCCCTCTTT TCTGGAATTTGGTACCGAGTACCTGTGTCTTCCCTGTGA
IG0146	L3S2P21 FW	TCACAGGGAAGACACAGGTACTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGC CTTTTTTCGTTTTGGTCCGCTTGTGTCTTCAGTGTCC
IG0147	L3S2P21 REV	GGACACTGAAGACACAAGCGGACCAAAACGAAAAAAGGCCCCCCTTTCGGGAGGCCTCTTTTC TGGAATTTGGTACCGAGTACCTGTGTCTTCCCTGTGA
IG0148	L3S2P36 FW	TCACAGGGAAGACACAGGTACTCGGTACCAAATTCCAGAAAAGAGACGCTGAAAAGCGTCTTT TTTATTGATGGTCCGCTTGTGTCTTCAGTGTCC
IG0149	L3S2P36 REV	GGACACTGAAGACACAAGCGGACCATCAATAAAAAAGACGCTTTTCAGCGTCTCTTTTCTGGA ATTTGGTACCGAGTACCTGTGTCTTCCCTGTGA
IG0150	L3S2P56m_FW	AGGTACTCGGTACCAAATTTTCGAAAAAAGACGCTGAAAAGCGTCTTTTTTCGTTTTGGTCC
IG0151	L3S2P56m_REV	AAGCGGACCAAAACGAAAAAAGACGCTTTTCAGCGTCTTTTTTCGAAAATTTGGTACCGAGT
IG0152	lac Vector REV	AGAAGACATCTTAACTTTTATCATACATTCTTTGC
IG0153	LacMut_FW	GAAATGATAGGTGCTTGGGGTATGGTGTGG
IG0154	LacMut_REV	AGCACCTATCATTTCAGGTTTAACTATATCATAATCAATAAT
IG0155	M13F	TGTAAAACGACGGCCAGT
IG0156	M13R	CAGGAAACAGCTATGACC

IG0157	mCH Vector REV	AGAAGACATCTTATTTATATAATTCATCCATACC
IG0158	MCHERRY FW C L1	TCACAGGGGTCTCCAATGGTATCTAAAGGAGAAGAAGATA
IG0159	MCHERRY REV D L1	TCACAGGGGTCTCCACCTTTATTTATATAATTCATCCATACCTCCT
IG0160	mCherryOpt FW C L0/2 corrected	TCACAGGGAAGACACAATGGTATCTAAAGGAGAAGAAGATA
IG0161	mCherryOpt RV D	TCACAGGGAAGACACACCTTTATTTATATAATTCATCCATACCTCCT
IG0162	mChOpt GG FW	TCACAGGGAAGACACGATTGCTCCTGCAGTAAAGGAG
IG0163	mChOpt GG REV	TCACAGGGAAGACACTAACGGATCCTTATTTATATAATTCATCCATAC
IG0164	NF_793_AscI	ATAGGCGCGCCCACCTCCTTTTTGACTTTAAGCCTACGAATACC
IG0165	NF_794_SbfI	ACACCTGCAGGCACCGACGAGCAAGGCAAGACCG
IG0166	pADAPT MoClo A FW	TACGGTACCGAAGACACGGAGTGAGACCGAATTCGCGGCCGCTT
IG0167	pADAPT MoClo D FW	TCACAGGGAAGACACAGGTTGAGACCGAATTCGCGGCCGCTT
IG0168	pADAPT MoClo E REV	TCACAGGGAAGACACAAGCGGAGACCTTTTTTGCCGGACTGCA
IG0169	pADAPT MoClo E REVdigest	TACGGATCCGAAGACACAAGCGGAGACCTTTTTTGCCGGACTGCA
IG0170	pFDX FW A	TCACAGGGGTCTCCGGAGGTGTAGTAGCCTGTGAAATAAGTAAG
IG0171	pFDX REV C	TCACAGGGGTCTCCCATTTGTAACACACCTCCTTAAAAATTACA
IG0172	phiLOV FW C	TCACAGGGAAGACACAATGATTGAAAAAAGTTTTGTTATTAC
IG0173	PHILOV FW C L1	TCACAGGGGTCTCCAATGATTGAAAAAAGTTTTGTTATTAC
IG0174	phiLOV REV D	TCACAGGGAAGACACACCTTTAAACATGATCTGAACCAACT
IG0175	PHILOV REV D L1	TCACAGGGGTCTCCACCTTTAAACATGATCTGAACCAACT
IG0176	phiLOV-SacI	TTTGAGCTCCTGCAGTAAAGGAGAAAATTTTATGATTGAAAAAAGTTTTGT
IG0177	phiTD1_FW	AGGTAACAATCAAAAGAAAAGCCTATCGTCTGAGGAACGGTAGGCTCTTTTGTAGCATATAGTTG
IG0178	phiTD1_REV	AAGCCAACTATATGCTACAAAAGAGCCTACCGTTCCTCAGACGATAGGCTTTTCTTTTGATTGTT
IG0179	pTHL FW A	TCACAGGGGTCTCCGGAGGCGGCCGCTTTTTAACAAAATATA

IG0180	pTHL REV C	TCACAGGGGTCTCCCATTTGAACTAACCTCCTAAATTTTGATACGG
IG0181	repL_GATC_seq_Rev	CCATCTAAGTTCCCTCTCAAATTC
IG0182	repLm FW	TGGTCTCTAGGTATTTTTGGTTTTGGTCGTCGCCT
IG0183	repLm REV	TGGTCTCTACCTCTTACTCGAATTTGGTAACTTTGAGCAAGAGGC
IG0184	REV external Spacer	TGGAAGACATCATTCTAACTAACCTCCTAAAGAGCAATCTCTAGTACCACTGGAAGTCG
IG0185	REV internal Spacer	AGTACCACTGGAAGTCGACCGATCTGGTTGTCACTCACTACATTCGTGTCGGTCGATTCCGCA TTACC
IG0186	REV_SEQ	CAAGACCGATCCCCTACTA
IG0187	REV-ext-spacer_C	TGGAAGACATCATTCTAACTAACCTCCTAAAG
IG0188	slpA*_ext_FW	AGGTCTTTAAATAGAAAAAGGCTTCTCTCATGAGAAGTCTTTTTTATTTAAAATA
IG0189	slpA*_ext_REV	AAGCTATTTTAAATAAAAAAGACTTCTCATGAGAGAAGCCTTTTTCTATTTAAAG
IG0190	Tadc_FW	AGGTTAATAAAAATAAGAGTTACCTTAAATGGTAACTCTTATTTTTTTAATATTGTTTCATAG TATTTCTTTCTAAACAGCCATGGG
IG0191	Tadc_REV	AAGCCCCATGGCTGTTTAGAAAGAAATACTATGAAACAATATTAAAAAAATAAGAGTTACCAT TTAAGGTAACTCTTATTTTTATTA
IG0192	Tfdx_FW	AGGTATAAAAATAAGAAGCCTGCATTTGCAGGCTTCTTATTTTTAT
IG0193	Tfdx_REV	AAGCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTAT
IG0194	TgyrA _FW	AGGTAAGAAGAAGTGTGAAAAAGCGCAGCTGAAATAGCTGCGCTTTTTTGTGTCATAA
IG0195	TgyrA _REV	AAGCTTATGACACAAAAAAGCGCAGCTATTTCAGCTGCGCTTTTTCACACTTCTTCTT
IG0196	thlA(ac)_KpnI	ACCGGTACCGCTTTTTAACAAAATATA
IG0197	thlA(ac)_SacI	TAAGAGCTCAAATTTTGATACGGGGTAACA
IG0198	TpepN _FW	AGGTTAATTTATAAATAAAAATCACCTTTTAGAGGTGGTTTTTTTATTTATAAATTA
IG0199	TpepN _REV	AAGCTAATTTATAAATAAAAAAACCACCTCTAAAAGGTGATTTTTATTTATAAATTA
IG0200	TslpA_CD_FW	AGGTAAATATAAAAAGACTTCTCAGATGAGAAGTCTTTTTTGTGAAA
IG0201	TslpA_CD_REV	AAGCTTTCACAAAAAAGACTTCTCATCTGAGAAGTCTTTTTATATTT
IG0202	TslpA_LA _FW	AGGTTGAAAAAGGCAGAGCGAAAGCTCTGTCTTTTTT
IG0203	TslpA_LA _REV	AAGCAAAAAAGACAGAGCTTTCGCTCTGCCTTTTTCA
IG0204	Tthl_FW	AGGTAAAAATAACTCTGTAGAATTATAAATTAGTTCTACAGAGTTATTTTT
IG0205	Tthl_REV	AAGCAAAAATAACTCTGTAGAACTAATTTATAATTCTACAGAGTTATTTTT
IG0206	TtyrS _FW	AGGTATAATCAATCGTCCCTTCGTGTAAACGAAGGGGCGTTTTTTATTT

IG0207	TtyrS _REV	AAGCAAATAAAAAACGCCCCTTCGTTTACACGAAGGGACGATTGATTAT
IG0208	Vector FW	AGAAGACATAGGTGGATCCTATAAGTTTTAATAAAACTTTAAATAG
IG0209	VF2	CCACCTGACGTCTAAGAAAC
IG0210	VR	GTATTACCGCCTTTGAGTGA

2.5 Plasmids and Strains used in this study.

The plasmids used in this study are listed in Table 2.2.

Table 1.2 Plasmids used in this study.

Name of Plasmid	Resistance	Source
pMTL85151	Cm15-25	Laboratory stocks of GBL[53]
pSB1A3:P*lac*:RBS:iLOV	Carb 100	This work.
pSB1A3:Plac:RBS:mCherry	Carb 100	This work.
REF0	Ery 500	This work.
pTREF1	Ery 500	This work.
pTREF2	Ery 500	This work.
pTREF3	Ery 500	This work.
pTREF1-Blue	Ery 500	This work.
pTREF2-Blue	Ery 500	This work.
pTREF3-Blue	Ery 500	This work.
pTREF1_T1	Ery 500	This work.
pTREF1_T2	Ery 500	This work.
pTREF1_T3	Ery 500	This work.
pTREF2_T1	Ery 500	This work.
pTREF2_T2	Ery 500	This work.
pTREF2_T3	Ery 500	This work.
pTREF3_T1	Ery 500	This work.
pTREF3_T2	Ery 500	This work.
pTREF3_T3	Ery 500	This work.
DVA_AC	Carb 100	This work.
DVA_PthlA_AC	Carb 100	This work.
DVA_Pfdx_AC	Carb 100	This work.
CLOLACZ_GEN_CD	Carb 100	This work.
GUSA_CD	Carb 100	This work.
CLOLACZ_SYN_CD	Carb 100	This work.
DVA_DE_T1	Carb 100	This work.
DVA_DE_T2	Carb 100	This work.
DVA_DE_T3	Carb 100	This work.
B0015_DE (DVA_DE_T4)	Carb 100	CIDAR MoClo Library from Addgene[267]
DVA_DE_T5	Carb 100	This work.
DVA_DE_T6	Carb 100	This work.
DVA_DE_T7	Carb 100	This work.
DVA_DE_T8	Carb 100	This work.
DVA_DE_T9	Carb 100	This work.
DVA_DE_T10	Carb 100	This work.

DVA_DE_T11	Carb 100	This work.
DVA_DE_T12	Carb 100	This work.
DVA_DE_T13	Carb 100	This work.
DVA_DE_T24	Carb 100	This work.
DVA_DE_T15	Carb 100	This work.
DVA_DE_T16	Carb 100	This work.
DVA_DE_T17	Carb 100	This work.
DVA_DE_T18	Carb 100	This work.
DVA_DE_T19	Carb 100	This work.
DVA_DE_T20	Carb 100	This work.
DVA_DE_T21	Carb 100	This work.
DVA_DE_T22	Carb 100	This work.
DVA_DE_T23	Carb 100	This work.
DVA_DE_T14	Carb 100	This work.
DVA_DE_T25	Carb 100	This work.
DVA_DE_T26	Carb 100	This work.
DVA_DE_T27	Carb 100	This work.
DVA_DE_T28	Carb 100	This work.
DVA_DE_T29	Carb 100	This work.
DVA_DE_T30	Carb 100	This work.
DVA_DE_T31	Carb 100	This work.
DVA_DE_T32	Carb 100	This work.
DVA_DE_T33	Carb 100	This work.
DVA_DE_T34	Carb 100	This work.
DVA_DE_T35	Carb 100	This work.
DVA_DE_T36	Carb 100	This work.
DVA_DE_T37	Carb 100	This work.
DVA_DE_T38	Carb 100	This work.
DVA_DE_T39	Carb 100	This work.
DVA_DE_T40	Carb 100	This work.
DVA_DE_T41	Carb 100	This work.
DVA_DE_T42	Carb 100	This work.
DVA_DE_T43	Carb 100	This work.
DVA_DE_T44	Carb 100	This work.
DVA_DE_T45	Carb 100	This work.
DVA_DE_T46	Carb 100	This work.
DVA_DE_T47	Carb 100	This work.
DVA_DE_T48	Carb 100	This work.
DVA_DE_T49	Carb 100	This work.
DVA_DE_T50	Carb 100	This work.
DVA_DE_T51	Carb 100	This work.
DVA_DE_T52	Carb 100	This work.
DVA_DE_T53	Carb 100	This work.
DVA_DE_T54	Carb 100	This work.

pTA-Mob	Gent 20	Donation by Dr Rahmi Lale[268]
pDSW1728	Cm15	Donation by Dr David S Weiss[99].
pDSW1728_SDM	Cm15	This work.
pDSW1728_SDM_D122E	Cm15	This work.
pDSW1728_SDM_erm_B	Ery 500	This work.
pDSW1728_SDM_D122E_erm_B	Ery 500	This work.
pDSW1728-thlA	Cm15	This work.
pDSW1728-thlA_D122E	Cm15	This work.
pMCH-85151	Cm15	This work.
pMCH-85151 D122E	Cm15	This work.
pRPF185	Cm15	Donation by Dr Rob Fagan[228]
pRPF185-thlA	Cm15	This work.
pGUS-85151	Cm15	This work.
pRPF185-phiLOV2.1	Cm15	Donation by Dr Gillian Douce[108]
pRPF185-phiLOV2.1 -thlA	Cm15	This work.
pLOV-85151	Cm15	This work.
pLOV-85151*RBS	Cm15	This work.
pRPF185-phiLOV2.1_erm_B	Ery 500	This work.
pLAC185-thlA_Syn	Cm15	This work.
pLAC-85151_Syn	Cm15	This work.
pLAC-85151_Syn_RBS	Cm15	This work.
pGUS-B-LAC_Syn	Cm15	This work.
pGUS-X-LAC_Syn	Cm15	This work.
pGUS-R-LAC_Syn	Cm15	This work.
pLAC-B-GUS_Syn	Cm15	This work.
pLAC-R-GUS_Syn	Cm15	This work.
pLAC-X-GUS_Syn	Cm15	This work.
pLAC-85151_GEN	Cm15	This work.
pLAC-85151_GEN_UTR	Cm15	This work.
pLAC-B-GUS_GEN	Cm15	This work.
pLAC-X-GUS_GEN	Cm15	This work.
pLAC-GUS_T1	Cm15	This work.
pLAC-GUS_T2	Cm15	This work.
pLAC-GUS_T3	Cm15	This work.
pLAC-GUS_T4	Cm15	This work.
pLAC-GUS_T5	Cm15	This work.
pLAC-GUS_T6	Cm15	This work.
pLAC-GUS_T7	Cm15	This work.

pLAC-GUS_T8	Cm15	This work.
pLAC-GUS_T9	Cm15	This work.
pLAC-GUS_T10	Cm15	This work.
pLAC-GUS_T11	Cm15	This work.
pLAC-GUS_T12	Cm15	This work.
pLAC-GUS_T13	Cm15	This work.
pLAC-GUS_T14	Cm15	This work.
pLAC-GUS_T15	Cm15	This work.
pLAC-GUS_T16	Cm15	This work.
pLAC-GUS_T17	Cm15	This work.
pLAC-GUS_T18	Cm15	This work.
pLAC-GUS_T19	Cm15	This work.
pLAC-GUS_T20	Cm15	This work.
pLAC-GUS_T21	Cm15	This work.
pLAC-GUS_T22	Cm15	This work.
pLAC-GUS_T23	Cm15	This work.
pLAC-GUS_T24	Cm15	This work.
pLAC-GUS_T25	Cm15	This work.
pLAC-GUS_T26	Cm15	This work.
pLAC-GUS_T27	Cm15	This work.
pLAC-GUS_T28	Cm15	This work.
pLAC-GUS_T29	Cm15	This work.
pLAC-GUS_T30	Cm15	This work.
pLAC-GUS_T31	Cm15	This work.
pLAC-GUS_T32	Cm15	This work.
pLAC-GUS_T33	Cm15	This work.
pLAC-GUS_T34	Cm15	This work.
pLAC-GUS_T35	Cm15	This work.
pLAC-GUS_T36	Cm15	This work.
pLAC-GUS_T37	Cm15	This work.
pLAC-GUS_T38	Cm15	This work.
pLAC-GUS_T39	Cm15	This work.
pLAC-GUS_T40	Cm15	This work.
pLAC-GUS_T41	Cm15	This work.
pLAC-GUS_T42	Cm15	This work.
pLAC-GUS_T43	Cm15	This work.
pLAC-GUS_T44	Cm15	This work.
pLAC-GUS_T45	Cm15	This work.
pLAC-GUS_T46	Cm15	This work.
pLAC-GUS_T47	Cm15	This work.
pLAC-GUS_T48	Cm15	This work.
pLAC-GUS_T49	Cm15	This work.
pLAC-GUS_T50	Cm15	This work.
pLAC-GUS_T51	Cm15	This work.

pLAC-GUS_T52	Cm15	This work.
pLAC-GUS_T53	Cm15	This work.
pLAC-GUS_T54	Cm15	This work.
pGUS-X-LAC_GEN	Cm15	This work.
pGUS-B-LAC_GEN	Cm15	This work.

The bacterial strains used in this work are listed in Table 2.3.

Table 2.2 Strains used in this work.

Name of Strain	Bacterial species	Genotype	Source
DH5α	*E. coli*	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17*(r_K^- m_K^+), λ$^-$	Laboratory Stocks
CA434 (HB101 + R702)	*E. coli*	F$^-$ *mcrB mrr hsdS20(*r_B^- m_B^-*) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(*Sm^R*) glnV44* λ^-	GBL Laboratory Stocks
MDS42 LowMut *ΔrecA*	*E. coli*	MG1655 multiple-deletion strain[269] Δ*dinB* Δ*polB* Δ*umuDC* [270]ΔIS609 Δ*patD* Δ*ydcV* Δ*ydcU* Δ*ydcT* Δ*ydcS* Δ*ydcR* Δ*hicA* Δ*hicB* Δ*yncJ* Δ*ydcP* Δ*ydcN* Δ*ydcO* Δ*ydcM* Δ*recA*(1819) The *recA* 1819 mutation is a deletion of *recA*	Scarab Genomics
XL-1 Blue	*E. coli*	*endA1 gyrA96(*nal^R*) thi-1 recA1 relA1 lac glnV44* F'[::*Tn10 proAB*$^+$ *lacI*^{q} *Δ(lacZ)M15] hsdR17*(r_K^- m_K^+)	Agilent

NEB Turbo	*E. coli*	*glnV44 thi-1 Δ(lac-proAB) galE15 galK16 R(zgb-210*::Tn*10*)Tet^{S} *endA1 f huA2 Δ(mcrB-hsdSM)5*, ($r_K^- m_K^-$) F′[*traD36 proAB*$^+$ *lacI*^{q} *lacZΔM15*]	GBL Laboratory Stocks
N1-4 (HMT) (=DSM 14923)	*C. saccharoper-butylacetonicum*	Wild-type solvent producer,	GBL Laboratory Stocks
168	*B. subtilis*		Laboratory Stocks

2.6 Aerobic fluorescence recovery (AFR)

A 500µL aliquot of a bacterial culture was mixed with 120µL of 5x fixation cocktail (composition per sample - 20µL 0.1M $NaPO_4$ buffer pH 7.0, 100µL of 16% (w/v) paraformaldehyde, optionally 4% of 25% (w/v) glutaraldehyde). The mixture was well mixed by pipetting up and down and incubated at room temperature for 30min. For anaerobic strains these steps was performed in the anaerobic workstation, after the 30min room temperature incubation step the tubes were taken out of the anaerobic workstation. The mixtures were then placed on ice and incubated for 60min. The cells were then washed with 500µL PBS buffer (NaCl - 8g/L, KCl – 0.2g/L, Na_2HPO_4 – 1.44g/L, KH_2PO_4 – 0.24g/L, pH – 7.4) or $NaPO_4$ buffer three times (by centrifugation at $1.5 \times 10^4 \times$ g and resuspension). The cells were then resuspended in 30-100µL of 1x PBS or 1x $NaPO_4$ buffer.

2.7 Epifluorescence microscopy

Epifluorescence microscopy was performed on an Axioskop 40 (Zeiss). Illumination was provided by a high intensity mercury lamp (Zeiss). Zeiss filter set 43 was used for detection of red fluorescence and Zeiss filter set 10 was used for detection of green fluorescence. Samples fixed with the AFR protocol were diluted in 1x PBS buffer 1 in 100 and 12µL of cell suspension was pipetted onto a cleaned glass slide. A cleaned glass no. 1.5 cover slip was placed on top of the cell droplet and affixed using nail varnish to prevent evaporation. Prior to taking an epifluorescence image a corresponding phase contrast image was also taken. Images shown were taken with 100x magnification objective and a CCD camera (CoolSnap EZ). Prior to using the 100x objective immersion oil was applied to the coverslip. Zeiss Filter Set 10 and Zeiss Filter Set 43 were used for green and red fluorescence detection, respectively. Exposure times for phase contrast microscopy were 0.01s, fluorescence microscopy exposure times ranged from 0.5s to 3s (specific exposure times are listed in the respective figure legends). The image acquisition software was Micro-Manager 1.14.6. Images were analyzed using ImageJ 1.48v - automatic brightness adjustment was performed, and scale bars were generated.

2.8 Flow cytometry

AFR samples were diluted 1 in 1000 in 1x PBS buffer or 1x $NaPO_4$ buffer in Falcon® round bottom polystyrene tubes and analysed on a BD Fortessa flow cytometer. Between 10000 and 50000 events were collected, depending on the experiment. Fluorescence, forward and side scatter values were recorded. Red fluorescence was measured using excitation at λ_{ex}=561nm and emission was recorded at λ_{em}=600-620nm. Green fluorescence was measured using excitation at λ_{ex}=488nm and emission was recorded at λ_{em}=515-545nm. Green fluorescence was also measured using excitation at λ_{ex}=405nm and emission was recorded at λ_{em}=500-550nm.

2.9 Bacterial crude lysate extraction

Bacterial cultures were harvested by centrifugation in a bench-top microcentrifuge at $1.5 \times 10^4 \times g$ for 10min. The supernatant was discarded and the cell pellet was resuspended in 1/2 of the original volume of Buffer Z (Na_2HPO_4 – 60mM, NaH_2PO_4 – 40mM, KCl – 10mM, $MgSO_4$ – 1mM, pH – 7.0). Toluene was added at 1/1000 of the resuspension volume. The suspension was vortexed for 1min and then incubated on ice for 10min. Then the suspension was placed at 37ºC in a well-ventilated area with open lids. The caps were then closed and the permeabilized cells were kept on ice or at 4ºC until assayed.

Bacterial cultures were harvested by centrifugation in a bench-top microcentrifuge at $1.5 \times 10^4 \times g$ for 10min. The supernatant was discarded and the cell pellet was resuspended in 1/5 of the original volume of Buffer Z. The samples were centrifuged at $1.5 \times 10^4 \times g$ for 3min and the supernatant removed. The pellet was resuspended in 300-500µL BugBuster®, vortexed and placed on a rotary shaker at room temperature for 20min with slow shaking at 15rpm. Samples were then centrifuged in a refrigerated microcentrifuge at $2 \times 10^4 \times g$ for 20min at 4ºC. The supernatant was transferred to a 1.5mL micro-centrifuge tube and kept on ice for use in enzymatic reporter assays or frozen at -80ºC for later use. The tubes containing cell pellets of insoluble protein were either discarded or used for SDS-PAGE (Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis) analysis.

2.10 Enzymatic Reporter Assays

2.10.1 β-D-Glucuronidase Assay

Using the lysates (Toluene or BugBuster®) described in Section 2.9, parallel assays of β-glucuronidase and β-galactosidase enzymatic activity were performed in 96 well

plates at 37ºC in an Epoch 2 Microplate reader. The reaction composition was 160µL Buffer Z + 20µL lysate + 20µL substrate (6mM p-nitrophenyl- β-D-glucuronide). Increases in absorbance at 405nm due to product release (p-nitrophenol) were recorded at regular intervals (every 166 seconds, starting at the 29th second mark) for 60 minutes.

An alternative assay was performed at 60ºC in an Epoch 2 Microplate reader. Briefly, half of the lysate, described in Section 2.9, was placed at 60ºC for 30min and then centrifuged in a refrigerated microcentrifuge at 2 × 10^4 × g for 20min at 4ºC. The supernatant was then used in an assay performed in 96 well plates at 60ºC. The reaction composition was 160µL Buffer Z + 20µL lysate + 20µL substrate (6mM p-nitrophenyl-β-D-glucuronide). As mentioned previously, increases in absorbance at 405nm due to product release (p-nitrophenol) were recorded at regular intervals (every 166 seconds, starting at the 29th second mark) for 60 minutes.

Activity calculations were performed by finding the slope of the change in absorbance (ΔA_{405}) over time (t_{min}) in the linear phase of the reaction (rate of reaction). The $\Delta A_{405}/t_{min}$ was then divided by the extinction coefficient ε=0.018µM^{-1}cm^{-1} of p-nitrophenol to give international units - IU (µmol/min) – amount of substrate released per minute. We then calculated the amount of substrate released per minute per unit of cell density (OD_{600}). The OD_{600} value of the culture at harvest was multiplied by the volume harvested (V_h), divided by the resuspension volume (V_r), and multiplied by the volume of resuspension used in the assay (V_u), divided by the total assay volume (V_t). This calculation results in an IU/OD_{600} value (shown in equation below):

$$\mathrm{IU}/OD_{600} = ((\Delta A_{405nm}/\mathrm{t}_{min}))/\varepsilon) \,/(OD_{600} * (V_h/V_r) * (V_u/V_t))$$

2.10.2 β-D-Galactosidase Assay

Using the lysates (Toluene or BugBuster®) described in Section 2.9, parallel assays of β-glucuronidase and β-galactosidase enzymatic activity were performed in 96 well plates at 37ºC in an Epoch 2 Microplate reader. The reaction composition was 160µL Buffer Z + 20µL lysate + 20µL substrate (6mM p-nitrophenyl- β-D-galactopyranoside or 13mM o-nitrophenyl- β-D-galactopyranoside or 6mM o-nitrophenyl- β-D-galactopyranoside, where specified). Increases in absorbance at 405nm due to product release (p-nitrophenol) or at 420nm (o-nitrophenol) were recorded every 166 seconds starting at the 29th second mark of incubation for p-nitrophenol and starting at the 74th second mark of incubation for o-nitrophenol, respectively, for 60 minutes. After switching β-galactosidase substrates, only p-nitrophenol release was monitored at the aforementioned time intervals (every 166 seconds, starting at the 29th second mark).

An alternative assay was performed at 60ºC in an Epoch 2 Microplate reader. Briefly, half of the lysate described in Section 2.9 was placed at 60ºC for 30min and then centrifuged in a refrigerated microcentrifuge at $2 \times 10^4 \times g$ for 20min at 4ºC. The supernatant was then used in an assay performed in 96 well plates at 60ºC. The reaction composition was 160µL Buffer Z + 20µL lysate + 20µL substrate (6mM p-nitrophenyl- β-D-galactopyranoside or 13mM o-nitrophenyl- β-D-galactopyranoside or 6mM o-nitrophenyl- β-D-galactopyranoside, where specified). Increases in absorbance at 405nm due to product release (p-nitrophenol) or at 420nm (o-nitrophenol) were recorded as described earlier in this section.

Activity calculations were performed by finding the slope of the change in absorbance (ΔA_{405}) over time (t (min)) in the linear phase of the reaction (rate of reaction). The $\Delta A_{405}/t$ was then divided by the extinction coefficient $\varepsilon=0.018\mu M^{-1}cm^{-1}$ of p-nitrophenol to give international units - IU (μmol/min) – amount of substrate released per minute. We then calculated the amount of substrate released per minute per unit of cell density (OD_{600}). The OD_{600} values of the culture at harvest was multiplied by the volume harvested, divided by the resuspension volume, and multiplied by the volume of resuspension used in the assay, divided by the total assay volume. This calculation results in an IU/OD_{600} value.

When o-nitrophenol was the product released the same calculations were performed by finding the slope of the change in absorbance (ΔA_{420}) over time (t_{min}) in the linear phase of the reaction (rate of reaction). The $\Delta A_{420}/t_{min}$ was then divided by the extinction coefficient $\varepsilon=0.0045\mu M^{-1}cm^{-1}$ of o-nitrophenol to give international units - IU (μmol/min) – amount of substrate released per minute. We then calculated the amount of substrate released per minute per unit of cell density (OD_{600}). The OD_{600} value of the culture at harvest was multiplied by the volume harvested (V_h), divided by the resuspension volume (V_r), and multiplied by the volume of resuspension used in the assay (V_u), divided by the total assay volume (V_t). This calculation results in an IU/OD_{600} value (shown in equation below):

$$IU/OD_{600} = ((\Delta A_{420nm}/t_{min}))/\varepsilon) \,/(OD_{600} * (V_h/V_r) * (V_u/V_t))$$

2.10.3 Algorithm for detection of the linear phase of the reaction for activity calculation

A simple algorithm (implemented in Excel) was used to detect the linear phase in the curve of absorbance increase against time. The algorithm consists of calculating the slope *m* of the line for 5 consecutive data points and calculating a Pearson coefficient *r* for the same 5 points. The slope *m* is multiplied by the modulus of the Pearson coefficient *r*. The modulus of *r* is taken in order to discount potential sharp negative slopes which may result from data noise as the product of a negative slope and a negative Pearson coefficient *r* is a positive number which may be larger than any other positive slope *m* in the data. Alternatively, r^2 can be used to the same effect. We found that this step (using absolute *r*) was only relevant for estimating activity of negative controls and very low activity samples. The slope *m* value corresponding to the largest aforementioned product ($m * |r|$) is taken as the input for IU/OD_{600} calculation. We found that this approach compared favourably to visual inspection of curves and selecting the linear portion of a curve for further analysis (a method which was also used for activity calculations reported in this work).

2.11 Bacterial Growth Measurement

Bacterial growth was monitored using optical density measurements in a bench-top UV/Vis spectrophotometer (Jenway 6305). Samples were blanked against uninoculated medium. Absorbance was measured at 600nm wavelength. Cultures were always diluted in uninoculated medium prior to measurement to $OD_{600} < 1.0$ and in practice often to $OD_{600} < 0.5$.

2.12 Statistical Analysis

Unpaired t-tests were performed to compare data obtained with different constructs and conditions. One-way ANOVA analysis followed by Dunnett's multiple comparisons test was used for comparing multiple sample values to negative and positive controls. *P*-values are reported in the main text and in figures. Simple linear regression analysis was performed to calculate reported coefficients of determination (R^2). Correlation analysis was performed to calculate Pearson correlation coefficient *r*.

2.13 Terminator Efficiency and Strength Calculations

Terminator efficiency was calculated using **Equation 4.3** (Figure 4.1). Terminator strength was calculated using **Equation 4.4** (Figure 4.3). The terminator reference constructs (control dual reporter plasmid without terminator inserted) used in the calculations were grown and assayed on the same day as most of the terminator samples. In the case of 96-well plate enzymatic assays, terminator reference construct aliquots were included on every plate and the terminator samples on the plate were normalized against the activities obtained from their respective plate terminator reference aliquots (plate control). In cases where the terminator reference construct was not assayed or grown at the same time as the sample (due to technical issues), the average of all control replicates (combining all plate control replicates and all replicates grown on different days) was used for normalization.

Chapter III

Evaluation of the suitability of gene expression reporters for measuring transcription termination in the solventogen *Clostridium saccharoperbutylacetonicum*

3. Evaluation of the suitability of gene expression reporters for measuring transcription termination in the solventogen *C. saccharoperbutylacetonicum*

To measure the strength of transcriptional terminators, we require robust reporters of gene expression that function in *C. saccharoperbutylacetonicum* and ideally a range of other related bacteria. A number of different reporters have been used in a range of Clostridia (see Introduction and Gyulev *et al.*, 2018), including recently developed fluorescent reporters and a range of more traditional enzymatic reporters. For a final reporter system measuring termination (Chapter 4), we ideally need two reporters that can be used simultaneously and be measured *in vivo*. Hence, in this chapter we outline extensive work testing the suitability of 5 different reporters; two fluorescent and three enzymatic for use in a final reporter system for measuring transcription termination.

For each reporter, a number of modifications were required to ensure compatibility with later Golden Gate cloning, to ensure high levels of expression *in vivo* and to remove the overlap of one of the reporters with a selectable marker (Cm^R).

3.1 Fluorescent reporter mCherry

For mCherry we started with a published plasmid containing the mCherry gene and engineered a variety of changes to this plasmid in anticipation of its compatibility for use in the final transcriptional termination reporter. These changes and plasmids are summarized in Figure 3.1 and listed in Table 2.2. Plasmid pDSW1728[99] was a kind donation from Dr David S. Weiss, University of Iowa and is an *E. coli-C.difficile* shuttle vector that contains a tetracycline-inducible *mCherryOpt* gene.

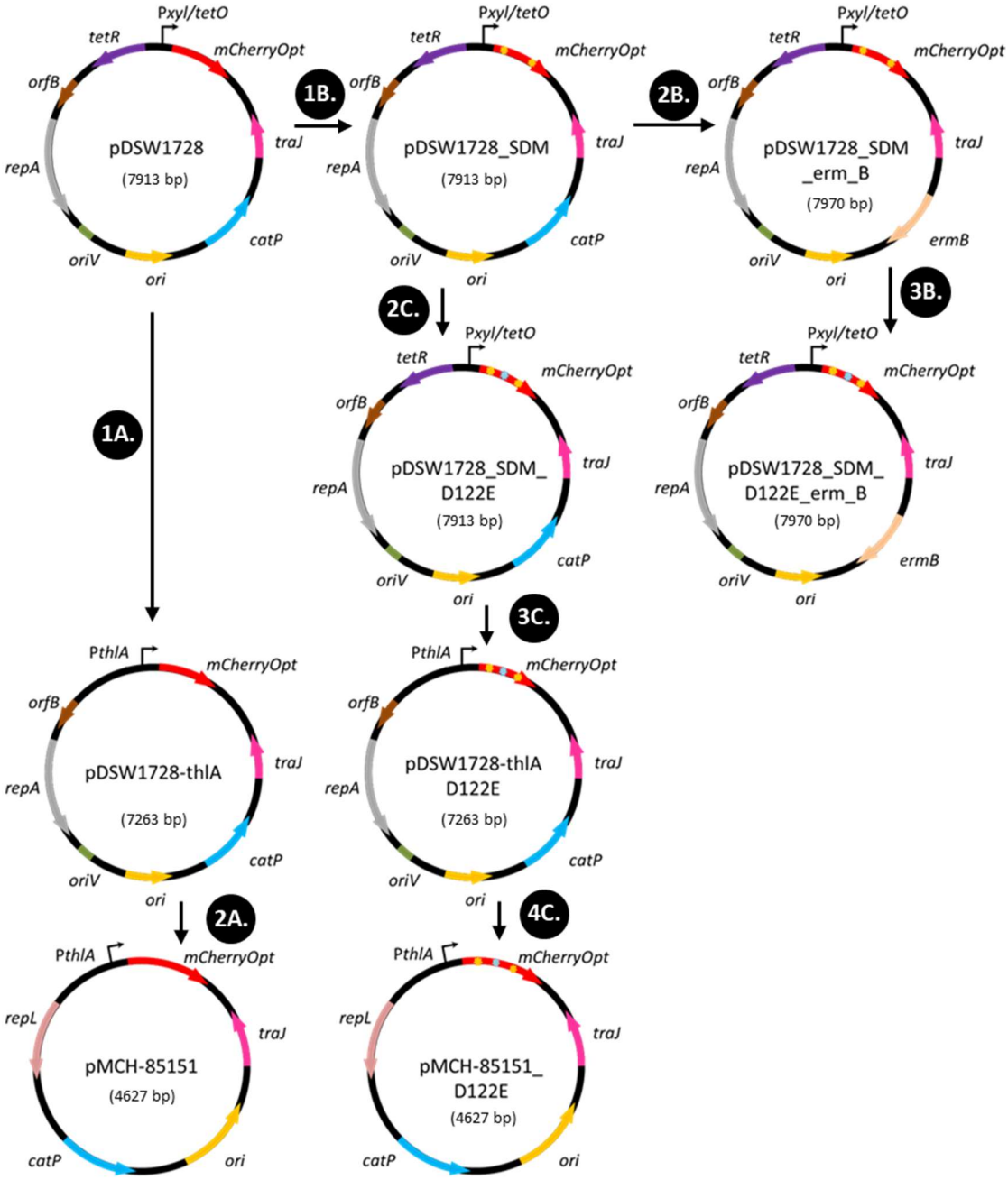


Figure 3.1 Schematic diagram of constructed plasmids of the mCherry series

1A. TetR-P*xyl/tetO* fragment is replaced with P*thlA*. **2A.** Backbone is exchanged for pMTL85151. **1B.** Site-directed mutagenesis for BpiI restriction sites removal in *mCherryOpt*. **2B.** Replacement of *catP* with *ermB* resistance marker. **2C and 3B.** Site-directed mutagenesis introducing D122E mutation in *mCherryOpt*. **3C.** TetR-P*xyl/tetO* fragment is replaced with P*thlA*. **4C.** Backbone is exchanged for pMTL85151.

3.1.1 Creating the mCherry plasmid series

3.1.1.1 Promoter replacement

The first potential issue in using pDSW1728 was that the tetracycline-derivative inducers of the promoter P*xyl/tetO* can inhibit bacterial growth[100,272]. Additionally, we wanted to use the mCherry reporter together with a green fluorescent FbFP phiLOV2.1 (see Chapter 3.2) and the tetracycline-derivative inducers are known to contribute to green background fluorescence[273,274]. Furthermore, the activity of the P*xyl/tetO* promoter had not been demonstrated in *C. saccharoperbutylacetonicum.*

With the goal of obtaining robust constitutive expression, the *tetR* gene and P*xyl/tetO* promoter were replaced with the strong constitutive thiolase promoter, P*thlA_Cac*, from *C. acetobutylicum* (see Introduction) to create pDSW1728-thlA. Briefly, plasmid pDSW1728 was digested with restriction enzymes SacI-HF and KpnI, dephosphorylated and the 7117bp restriction fragment gel extracted. Promoter P*thlA_Cac* was amplified using PCR from plasmid DVA_P*thlA*_AC (Table 2.3) using primers IG0196 and IG0197. The resulting 158bp PCR product was purified, digested with KpnI and SacI-HF, gel extracted and then ligated to the 7117bp pDSW1728 restriction fragment to form plasmid pDSW1728-thlA. *E. coli* colonies transformed with this ligation mix displayed a faint reddish pink colour that developed further when placed at 4ºC in the dark (to prevent bleaching), suggesting strong constitute expression of mCherry had been successful.

3.1.1.2 Site-directed mutagenesis of mCherryOpt

Plasmid pDSW1728 contains a *C. difficile* codon-optimized version of the gene coding for mCherry – *mCherryOpt*[99]. The sequence contains two BpiI (BbsI) restriction enzyme sites that would interfere with downstream uses of the gene that involve Golden

Gate cloning. The gene was modified using site-directed mutagenesis PCR using primer pairs IG0015 and IG0016 and IG0019 and IG0020 in order to remove two BbsI (BpiI) restriction sites. The silent mutations introduced were A to T at nucleotide position 120 of *mCherryOpt* and also A to T at nucleotide position 672. In both cases, the codon changed was GGA (Glycine) to GGT. The resulting plasmid was named pDSW1728_SDM.

Surprisingly, the *mCherryOpt* gene present in pDSW1728 contained a mutation that resulted in a Glu122Asp substitution in the mCherry. This change is contained with the reported sequence of the protein, but not discussed in the publication[99]. As there are reported instances where a single substitutions can alter the spectral properties of mCherry[275], this was reverted back to the original sequence. Plasmid pDSW1728 SDM was PCR amplified with primer pair IG0098 and IG0099 (see Methods and Materials). The resulting plasmid was named pDSW1728_SDM_D122E and was verified by Sanger sequencing using primers IG0119 and IG0186.

3.1.1.3 Vector exchange

As pDSW1728 (and the derivatives we constructed such as pDSW1728-thlA) is an *E. coli-C. difficile* shuttle vector, it contains a Gram-positive origin of replication, pCD6, that was isolated from *C. difficile* (see Introduction). As the copy number of the vector could significantly influence measurable reporter levels, we wished to compare this origin to another, pIM13, that has been used previously in *C. saccharoperbutylacetonicum*[276]. The copy number of pIM13 is high in *Bacillus subtilis*[54], and while not having being experimentally measured in *C. saccharoperbutylacetonicum*, is thought to be lower than in *B. subtilis*, closer to that reported in *Staphylococcus aureus*[276]. The use of the pIM13 origin also allows the use

of the reporter vector in the model Gram-positive bacterium *B. subtilis*, expanding the general utility of the reporters. In addition, the transfer of the reporter cassette to the modular pMTL80000-vector series that contains pIM13, would allow us to easily accomplish any further changes to the plasmid such as origin of replication and selectable marker replacement if required.

The P*thlA*::*mCherryOpt* transcription unit from pDSW1728-thlA was amplified using primers IG0164 and IG0165. The resulting 1316bp PCR product was then PCR purified, digested with AscI and SbfI and gel extracted. Plasmid pMTL85151 was digested with the same restriction enzymes, dephosphorylated and the 3329bp fragment was gel extracted and the two fragments were ligated to form plasmid pMCH-85151 (Fig. 3.1). When transformed into *E. coli,* colonies, containing pMCH-85151, had a faint reddish pink colour. Indeed the colour appeared more intense than with pDSW1728-thlA. This observation and higher DNA concentrations from plasmid extractions (data not shown) indicated that the Gram-negative origin of pMTL85151 (listed as ColE1) maintains a higher copy number than pDSW1728 (also listed as ColE1). The pMTL85151 ColE1 module's listed sequence matches that of pUC19, pSB1C3 and pSB1A3 (mentioned below) amongst others and is known to produce a very high copy number[277], the alignment of the reported sequences of pDSW1728 and pMTL85151 reveals a difference from the sequence in pDSW1728 at two positions.

To introduce the wild-type mCherry allele into pMCH-85151, it was digested with enzymes SacI-HF and BamHI and the restriction fragment of 3892bp was gel extracted. Plasmid pDSW1728_SDM_D122E was also digested with enzymes SacI-HF and BamHI and the restriction fragment of 743bp was gel extracted and ligated with the pMCH-85151 restriction fragment to form plasmid pMCH-85151 D122E.

3.1.1.4 High-copy E. coli plasmid expressing mCherry

Since we were using *mCherryOpt* and this gene has a 30.1% GC-content and a markedly different codon usage compared to *E. coli* with 15 rare codons, we anticipated that this may lead to low levels of gene expression in *E. coli*. In order to validate our pilot flow cytometry and fluorescence microscopy experiments, we needed a good positive control likely to express high levels of active mCherry in *E. coli.*

To that end, a plasmid constructed from commonly used biological parts from the Registry of Standard Parts was created. BioBrick BBa_J06702, which consists of RBS (B0034), mCherry CDS (J06504) and a double terminator (B0015) and is carried on plasmid backbone pSB1C3, was digested with XbaI and PstI. The 869bp fragment was gel extracted. Plasmid backbone pSB1A3 was digested with EcoRI and PstI and the 2155bp fragment was gel extracted and dephosphorylated. Promoter P*lacZYA*-RBS combination (BBa_J04500 (promoter BBa_R0010 + RBS BBa_B0034)) carried on plasmid pSB1C3 was digested with EcoRI and SpeI and the 220bp fragment was gel extracted. The three fragments were ligated together in a single ligation reaction to form plasmid pSB1A3:P*lac*:RBS:mCherry. Interestingly, by design this combination of BioBricks yields pSB1A3:P*lac*:RBS:RBS:mCherry with a duplicated BBa_B0034 RBS, however, selecting for a visibly strongly red colony post transformation yielded a construct with a single RBS copy upon sequencing. *E. coli* colonies derived from this clone, had a very strong reddish pink colour that developed further when placed at 4ºC in the dark (to prevent bleaching).

3.1.1.5 Resistance marker change

Finally, the pDSW1728 plasmid carries Cm^R and as one of the reporter genes that we were considering (see Introduction and Chapter 3.5) was chloramphenicol

acetyltransferase (CAT), we decided to replace the selectable marker of this plasmid and its derivatives in order to eliminate the selectable marker-reporter overlap.

To that end, plasmid pDSW1728_SDM was amplified with primers IG0011 and IG0024. This amplification encompassed most of the plasmid with the exception of the resistance gene *catP*. The PCR products were dephosphorylated using NEB Antarctic Phosphatase. The *ermB* gene was amplified from pMTL83251 with primers IG0037 and IG0038. The *ermB* PCR product was phosphorylated using T4 kinase and blunt-end ligated with the PCR products from pDSW1728_SDM to form plasmids pDSW1728_SDM_erm_A and pDSW1728_SDM_erm_B. Plasmids A and B differed in the orientation of the *ermB* gene with respect to the rest of the plasmid. Plasmid B in which *ermB* is in the opposite orientation of *catP* was used for further experiments as this matched the design of the first generation of dual reporter constructs (pTREF1, pTREF2 and pTREF3) (see Chapter 4).

We needed to construct a negative control plasmid for the Erm^R reporter constructs. Plasmid pREF0 was constructed by digesting plasmid pTREF1 (see Chapter 4) with EcoRI and self-ligating the 6168bp fragment. Briefly, plasmid pTREF1 was constructed through Golden Gate assembly and contained the *ermB* resistance gene and a dual reporter cassette, the rest of the plasmid was derived from pRPF185-phiLOV2.1Opt and was therefore assumed identical to pDSW1728 (both plasmids derived from pRPF185).

3.1.2 Measurement of mCherry expression levels

3.1.2.1 Flow cytometry in Escherichia coli

To first verify that we are able to detect mCherry expression in *E. coli*, the *E. coli* DH5α strain, transformed with pSB1A3:P*lac*:RBS:mCherry, was grown overnight in C medium supplemented with 100 μg/mL of carbenicillin with and without the P*lacZYA* inducer IPTG. The cells were harvested by centrifugation and washed in PBS buffer. The fluorescence levels were measured using a Beckman Fortessa X-20 Flow Cytometer.

Analysis of the fluorescent signals from these bacterial cells revealed no signal in the negative control sample of DH5α alone (Figure 3.2), while high levels of fluorescence in strains containing pSB1A3:P*lac*:RBS:mCherry. Notably the signal was equally high in both the absence and presence of the inducer IPTG (overlay in Fig. 3.2 D). This is likely due to titration of the chromosomally-encoded LacI repressor by an excess of plasmid-encoded binding sites. The induced sample's geometric mean of fluorescence is very similar to the uninduced sample's and more replicates would be needed to establish whether the difference is statistically significant (Table 3.1).

Table 3.1 Statistics of red fluorescence signal of *E. coli* DH5α strains expressing mCherry. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Events	% of Visible	Mean	GM	Median	CV
DH5α *wt* (IPTG)	50000	100	92.908	77.591	81	889.73
P*lac*:RBS: mCherry (no IPTG)	50000	100	54563	27229	48762	69.709
P*lac*:RBS: mCherry (IPTG)	50000	100	59600	37149	50256	71.5

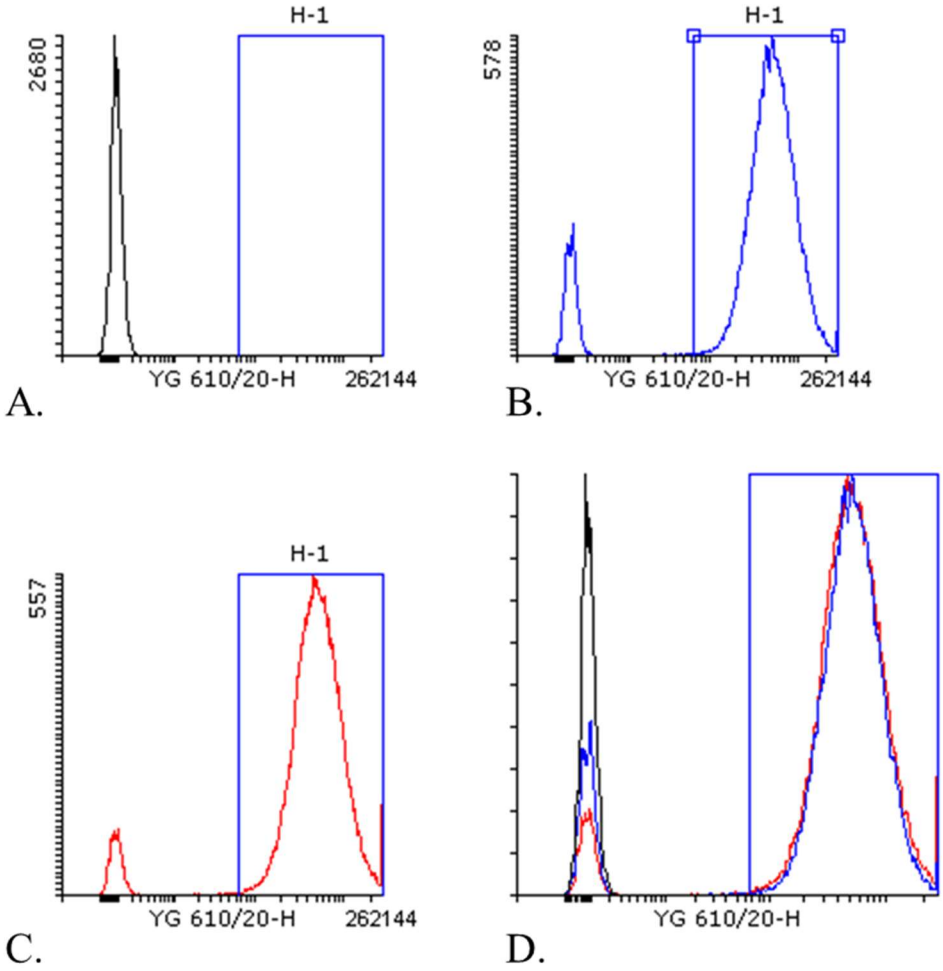


Figure 3.2 Flow cytometry histograms of red fluorescence signal of *E. coli* DH5α strains expressing mCherry.

Event count versus fluorescence intensity (λ_{ex}=561nm, λ_{em}=600-620nm) of *E. coli* DH5α strains. **A.** Red fluorescence of wild-type DH5α induced with IPTG. **B.** Red fluorescence of uninduced pSB1A3:P*lac*:RBS:mCherry **C.** Red fluorescence of IPTG-induced pSB1A3:P*lac*:RBS:mCherry **D.** Overlay histogram of **A**, **B** and **C**. Gate H1 contains the highly fluorescent cell population.

Next, we tested mCherry expression levels from the modified shuttle vector pDSW1728_SDM_erm_B in *E. coli* in order to verify its functionality after the mutagenesis and marker replacement modifications. These experiments were undertaken in *E. coli* strain CA434, which is the donor strain used for conjugation with *C. saccharoperbutylacetonicum* (see Materials and Methods).

To this end, *E. coli* CA434 strains transformed with plasmid pDSW1728_SDM_erm_B and plasmid REF0 were grown in minimal C medium (4% (w/v) glycerol, 3% (w/v) D-glucose) for 36 hours. Genotypes were inoculated in duplicate – one replicate was induced at onset of growth with 400ng/mL anhydrotetracycline (ATc) while the second culture was not. Cells were harvested and fixed according to the AFR protocol.

The results of the flow cytometry experiment indicated that mCherry was expressed under the inducing conditions (Figure 3.3C) and that its expression was tightly repressed in the absence of inducer (Figure 3.3B). Forward versus side scatter area ratios indicated consistent cell morphology between the treatments (data not shown). However, it is worth noting that the geometric mean levels of fluorescence were 29-times lower than the induced pSB1A3:P*lac*:RBS:mCherry (Table 3.2). We also tested shorter induction times of 1 and 13 hours with pre-grown cells but those yielded lower fluorescence increases (data not shown).

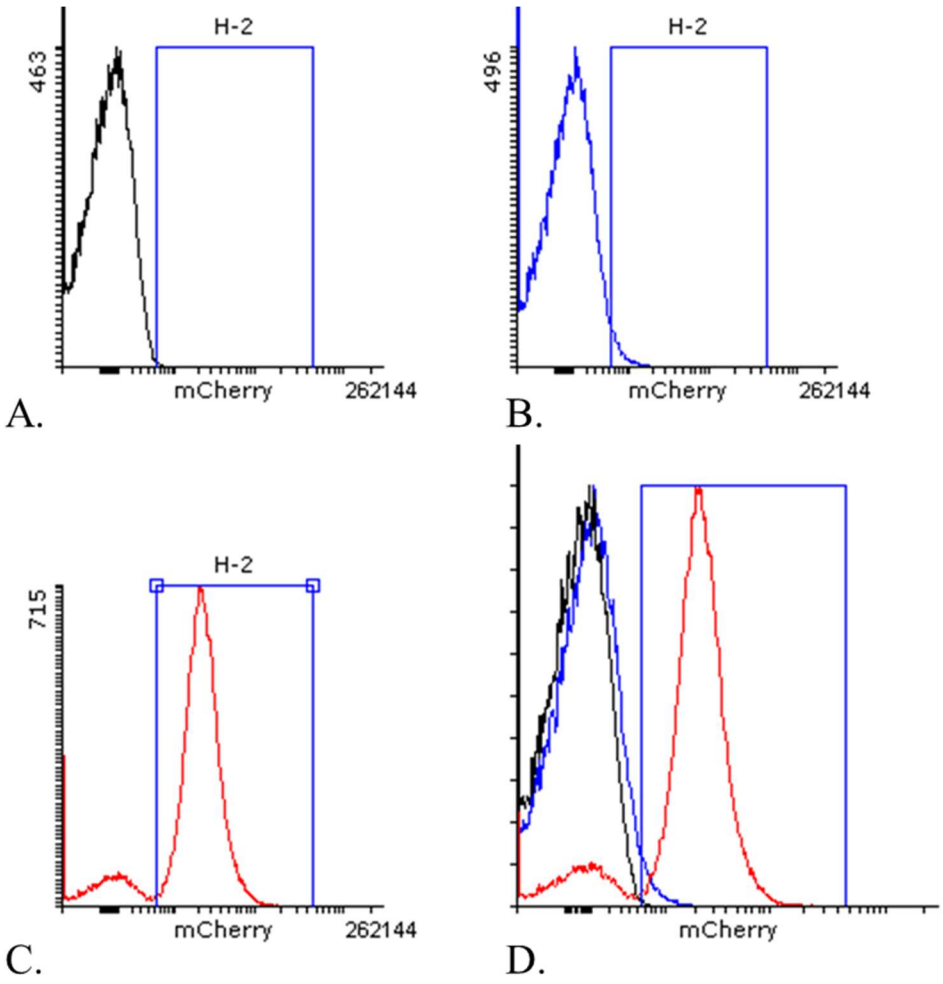


Figure 3.3 Flow cytometry histograms of red fluorescence signal of *E. coli* CA434 strains expressing mCherry.

Event count versus fluorescence intensity (λ_{ex}=561nm, λ_{em}=600-620nm)) of) of *E. coli* CA434 strains harbouring plasmids REF0 and pDSW1728_SDM_erm_B. A. Red fluorescence of ATc-induced REF0. B. Red fluorescence of uninduced pDSW1728_SDM_erm_B C. Red fluorescence of ATc-induced pDSW1728_SDM_erm_B. Overlay histrogram of A, B and C (D). Gate H2 contains the highly fluorescent cell population.

Table 3.2 Statistics of red fluorescence signal of *E. coli* CA434 strains expressing mCherry. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Inducer	Gate	Events	% of Visible	Mean	GM	Median	CV
REF0	-	All	50000	100	9.07	8.58	-1.72	1635.1
		H2	46	0.09	887.91	667.21	598.56	192.06
	+	All	50000	100	6.2	8.26	-4.3	2213.3
		H2	49	0.1	615.99	610.7	580.5	13.93
pDSW1728_SDM_erm_B	-	All	50000	100	63.88	15.8	36.12	300.94
		H2	1031	2.06	792.11	746.46	681.12	47
	+	All	50000	100	2209.6	991.98	2035.6	78.12
		H2	42186	84.37	2613.3	2289.9	2260.1	60.32

3.1.2.1 Fluorescence microscopy and flow cytometry in C. saccharoperbutylacetonicum

To determine whether we can detect mCherry expression in *C. saccharoperbutylacetonicum*, plasmid pDSW1728 was transferred into *C. saccharoperbutylacetonicum* N1-4 (HMT) via conjugation (see Materials and Methods). Liquid cultures of 2x YTG medium (1.3% (w/v) D-glucose) supplemented with appropriate antibiotic were inoculated and grown for 24-48 hours. Two identical subcultures per starter culture were inoculated (in 2x YTG 1.3% (w/v) D-glucose plus antibiotic), one of the subcultures was induced with anhydrotetracycline (400ng/mL final concentration). Culture growth was assessed by absorbance measurements. After performing aerobic fluorescence recovery (AFR) the cells were imaged using epifluorescence microscopy on an Axioskop 40 (Zeiss). Illumination was provided by a high intensity mercury lamp (Zeiss). Zeiss filter set 43 for detection of red fluorescence was used.

Red fluorescence was readily detectable in the pDSW1728 samples (Figure 3.4A) whereas wild-type cells had virtually no background red fluorescence with exposure of

0.5s (Figure 3.4C). It is also noticeable that the bulk of the population is uniformly fluorescent, while there are a few cells in the field of vision that are very intensely fluorescent.

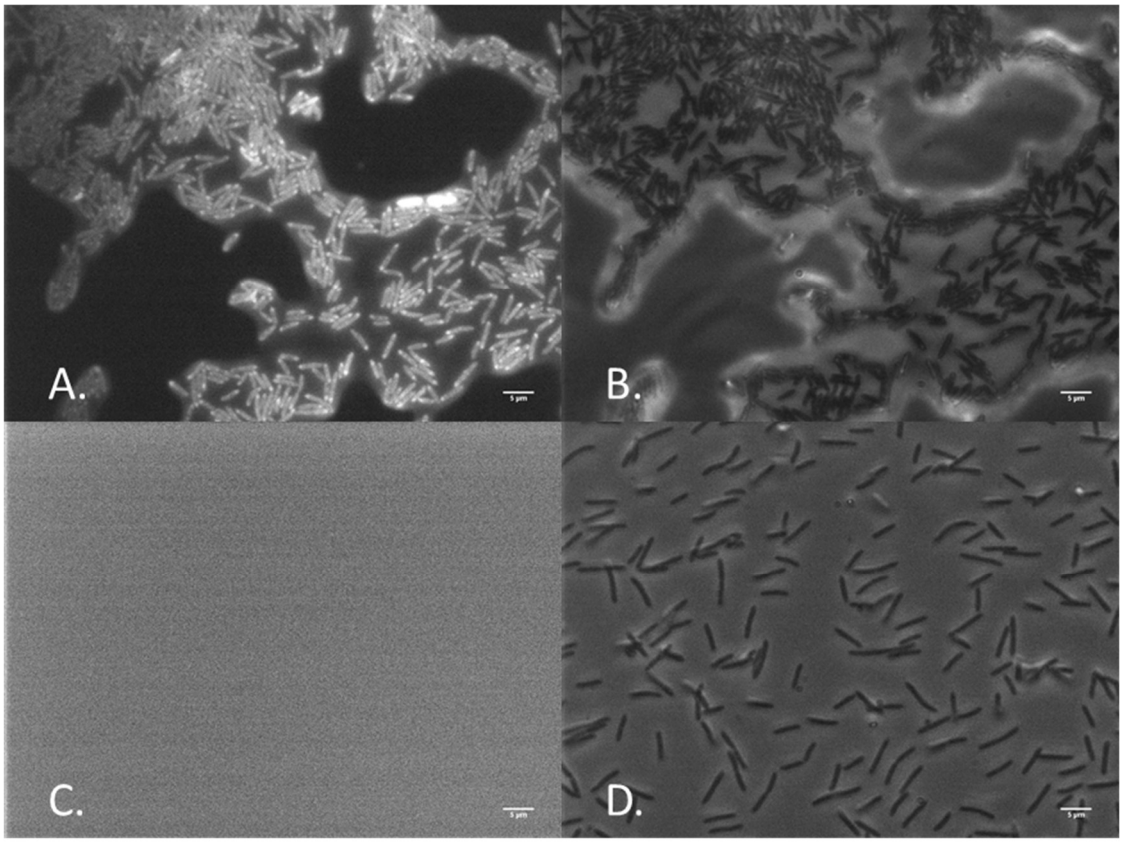


Figure 3.4 Microscopy images for mCherry expression analysis of induced wild-type *C. saccharoperbutylacetonicum* and induced pDSW1728.

A. Epifluorescence microscopy image of pDSW1728 – red channel. B. Corresponding phase-contrast microscopy image of pDSW1728. C. is fluorescence microscopy of wild-type. D. Corresponding phase-contrast image. Exposure times were 0.5s for epifluorescence and 0.01s for phase contrast. Images were processed using ImageJ for automatic adjustment of brightness. Scale bars (bottom right of each panel) are 5µm.

Next we tested our constitutive construct for expression. To that end, plasmids pDSW1728-thlA and pRPF185-thlA (containing *gusA* and used as a negative control here) were transformed into *C. saccharoperbutylacetonicum* N1-4 (HMT) via electroporation. Cultures were grown in RCMm medium, supplemented with 75µg/mL thiamphenicol. After overnight incubation, OD_{600} was measured, cells were harvested and AFR was performed. Fluorescence levels were measured using a Beckman Fortessa

X-20 Flow Cytometer. The results showed that mCherry was expressed in cells harbouring pDSW1728-thlA (Figure 3.5 C and D) and wild-type cells had very low background red fluorescence. The proportion of fluorescent cells per sample was 50-52%. The geometric mean of fluorescence was just under 4-fold higher in cells expressing mCherry (Table 3.3). The fluorescent sub-population's geometric mean was on average 14.4 times higher than the geometric mean of the control samples.

Table 3.3 Statistics of *C. saccharoperbutylacetonicum* strains expressing mCherry from pDSW1728-thlA. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean. Values are averages of two biological replicates and standard deviation is expressed following the ± sign.

Strain	Gate	Events	% of Visible	Mean	GM	Median	CV
pRPF185-thlA	All	7904± 2964.20	100±0	30.95± 13.33	12.03± 4.68	20.44± 9.23	235.10± 15.27
	H1	116.5± 30.41	1.66± 1.01	338.71± 68.78	287.37± 27.08	247.08± 14.76	85.23± 49.87
pDSW1728-thlA	All	10000±0	100±0	352.21± 20.25	73.981± 2.6198	178.79± 26.453	290.44± 131.28
	H1	4959.5± 234.05	49.595± 2.3405	662.78± 47.943	523.9±2 8.269	507.65± 28.914	132.94± 0.93116

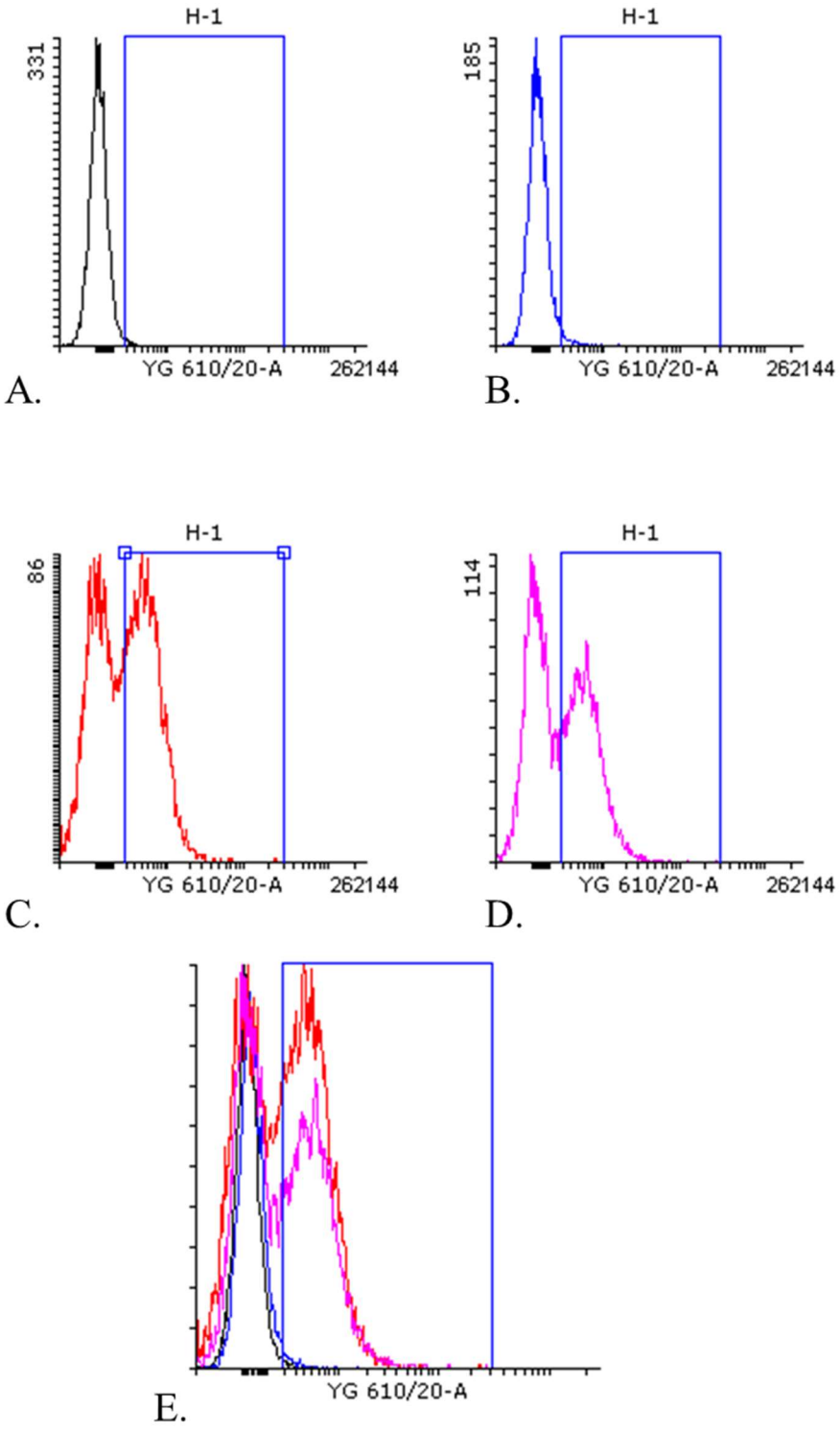


Figure 3.5 Flow cytometry histograms of *C. saccharoperbutylacetonicum* strains expressing mCherry from pDSW1728-thlA.

Event count versus fluorescence intensity (λ_{ex}=561nm, λ_{em}=600-620nm) of strains harbouring plasmids pRPF185-thlA (A, B) and pDSW1728-thlA (C, D). Overlay histogram of A, B, C and D (F). Gate H1 contains the red fluorescent cell population.

To evaluate the effect of the vector change and the site-directed mutagenesis of *mCherryOpt*, plasmids pMCH-85151 and pMCH-85151 D122E were transformed into *C. saccharoperbutylacetonicum* N1-4 (HMT) via electroporation (see Methods and Materials). Starter cultures of these two strains as well as negative control plasmids (pMTL85151 and pEMPTY) were grown in RCMm medium supplemented with 75μg/mL thiamphenicol. Sub-cultures were inoculated in triplicate at a starting OD_{600} of 0.1 and were grown at 32°C for 12 hours at which point they were harvested by centrifugation. The density of the culture at harvest was measured at OD_{600}. Harvested samples were fixed using the AFR method. Fluorescence levels were then measured using a Beckman Fortessa X-20 Flow Cytometer.

The data showed very high levels of fluorescence in the samples harbouring pMCH-85151 (Figure 5A –C). The geometric mean fluorescence of these samples was over 46.5-fold higher than the negative controls (Figure 5D-F) and over 10.8-fold higher than the plasmids expressing mCherry from pDSW1728-thlA (Figure 4B). The proportion of highly fluorescent cells amongst the samples was on average 62.29% and their average geometric mean was 461-fold higher than controls (Table 3.4).

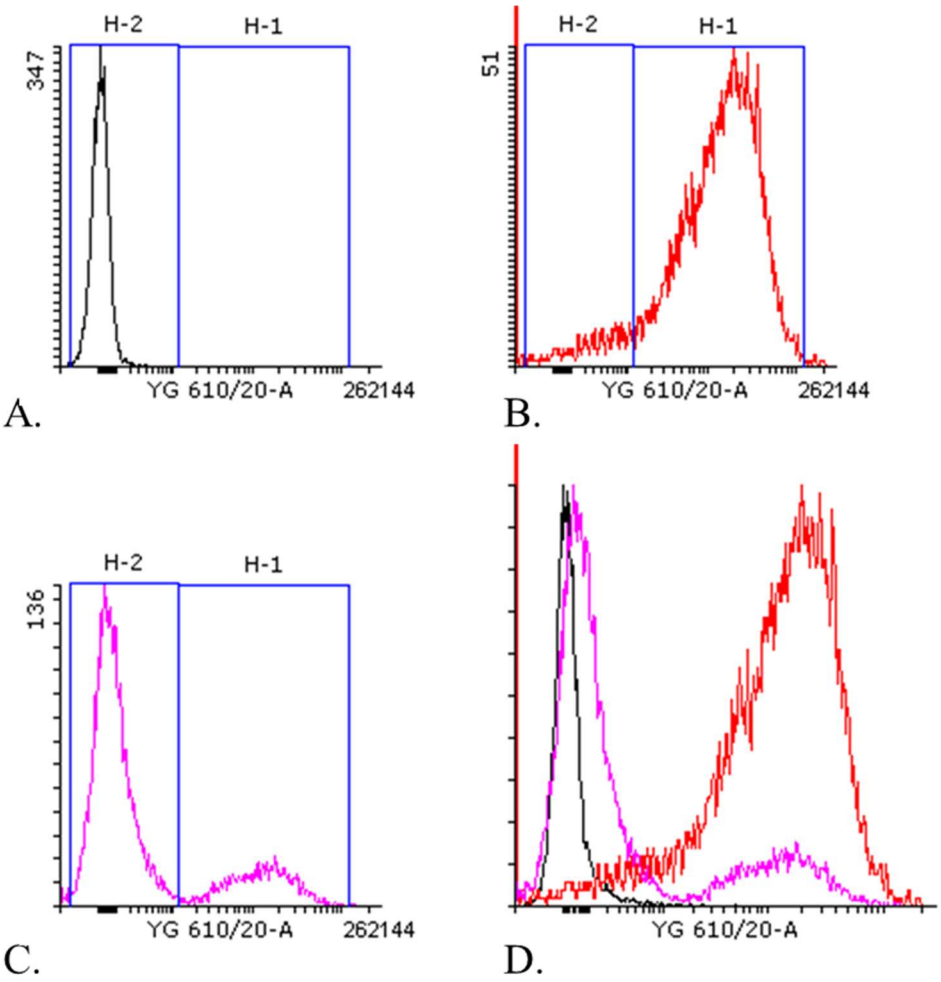


Figure 3.6 Flow cytometry histograms of *C. saccharoperbutylacetonicum* strains expressing mCherry from plasmids pMCH-85151 and pMCH-85151 D122E.

Event count versus fluorescence intensity (λ_{ex}=561nm, λ_{em}=600-620nm) of *C. saccharoperbutylacetonicum* strains harbouring plasmids pMTL85151 (**A**), pMCH-85151 (**B**) and pMCH-85151 D122E (**C**). Overlay histogram of **A**, **B** and **C** (**D**). Gate H1 contains the highly fluorescent cell population and H2 contains the non-fluorescent cell population.

Table 3.4 Statistics of red fluorescence signal of *C. saccharoperbutylacetonicum* strains expressing mCherry from plasmids pMCH-85151 and pMCH-85151 D122E. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation.Values are averages of three biological replicates and standard deviation is expressed following the ± sign.

Strain	Gate	Events	% of Visible	Mean	Geometric Mean	Median	CV
pMTL85151	All	10000±0	100±0	23.45±5.10	8.40±0.94	13.35±1.54	342.35±28.41
	H1	5.67±1.16	0.06±0.01	2208±687.65	2016.90±476.97	1703.5±232.56	40.32±23.77
	H2	9977.7±8.14	99.78±0.08	22.37±4.81	8.40±0.93	13.35±1.54	251.44±18.39
pMCH-85151 E122D	All	10000±0	100±0	14207±987.13	541.85±166.92	7414.3±1264.19	138.72±13.76
	H1	6210±297.97	62.1±2.98	22538±1132.6	16014±1323.5	18388±1553.9	78.30±6.02
	H2	467±56.63	4.67±0.57	379.04±56.23	137.19±46.39	288.36±86.17	94.36±13.49
pMCH-85151 D122E	All	10000±0	100±0	2621.61±948.16	59.14±48.27	53.70±38.28	347.11±73.10
	H1	1564±496.51	15.64±4.97	16050±3107.8	11069±1827.8	11304±2312.8	94.56±7.25
	H2	8007.3±193.96	80.07±1.94	85.01±24.66	21.68±11.91	37.09±23.16	198.01±52.13

Cells expressing pMCH-85151 D122E (Figure 6A-C) unexpectedly exhibited a slightly lower geometric mean fluorescence than pDSW1728-thlA. The population was clearly bi-modal, with 16% (11-21%) on average per sample being strongly fluorescent with the rest of the population having no or low fluorescence. The fluorescent sub-population had an average geometric mean fluorescence that was 304 times higher than the control. While the cause of this result has not been characterized in detail thus far, there are some possible causes – (i) loss of plasmid in the source glycerol culture, (ii) reduced translation due to codon preference (due to silent mutations), (iii) incomplete maturation or lower stability of fluorescent protein after AFR (difference due to residue 122) or (iv) other unidentified causes.

3.1.3 Conclusions on the use of mCherry as a reporter for gene expression in *C. saccharoperbutylacetonicum*

The reporter mCherry is routinely used in *E. coli* and a wide range of other organisms. We were able to successfully detect mCherry fluorescence in *C. saccharoperbutylacetonicum* which has not been reported previously. We have demonstrated detection of mCherry fluorescence with both an inducible and a constitutive promoter in *C. saccharoperbutylacetonicum*. The use of the mCherry reporter allows for single cell resolution and enables the measurement of differences in expression levels that appear to be directly related to the origin of replication (higher expression level in pIM13 plasmid than the pCD6 one). Future work can use mCherry as a sensitive genetically-encoded reporter for promoter, RBS and 5' and 3' UTR characterization. The brightness and photostability of mCherry allow for applications such as Fluorescence-Activated Cell Sorting (FACS). While in our current implementation, cells are killed during AFR because of our use of paraformaldehyde

fixation, AFR can be performed without fixation and recently FACS was used with live *C. beijerinckii* FbFP-expressing cells that were then grown for plasmid re-isolation and sequencing[68], thus it might be possible to culture some aerotolerant *Clostridium* strains post AFR and FACS. Even if the cells are not viable, plasmid sequencing is readily achievable from FACS-sorted suspensions (see Discussion for further applications of mCherry in *Clostridium* species).

3.2 Flavin-binding protein phiLOV2.1 as an anaerobic fluorescent reporter

To develop a dual terminator reporter, a second fluorescent reporter was required to create a matched pair for simultaneous *in vivo* measurements with mCherry (see Chapter 4 for detailed explanation of dual reporter systems for terminator characterization). The Flavin-binding fluorescent protein (FbFP) phiLOV2.1 was assessed as a second fluorescent reporter of gene expression in *C. saccharoperbutylacetonicum*. The main advantage of this protein as a reporter in anaerobes is its oxygen-independence since it only requires FMN as a co-factor. Similar to the work described for mCherry, a number of plasmid constructs containing phiLOV2.1 were created for this work (Figure 3.7 and Table 2.3).

3.2.1 Creating the phiLOV2.1 plasmid series

The work with phiLOV2.1 started with plasmid pRPF185-phiLOV2.1Opt which was a kind donation from Dr Gillian R. Douce, University of Glasgow. This plasmid is an *E. coli-C. difficile* shuttle vector (like pDSW1728 it is also derived from pRPF185) and allows for the tetracycline-inducible expression of *phiLOV2.1Opt* which is codon-optimized for *C. difficile*.

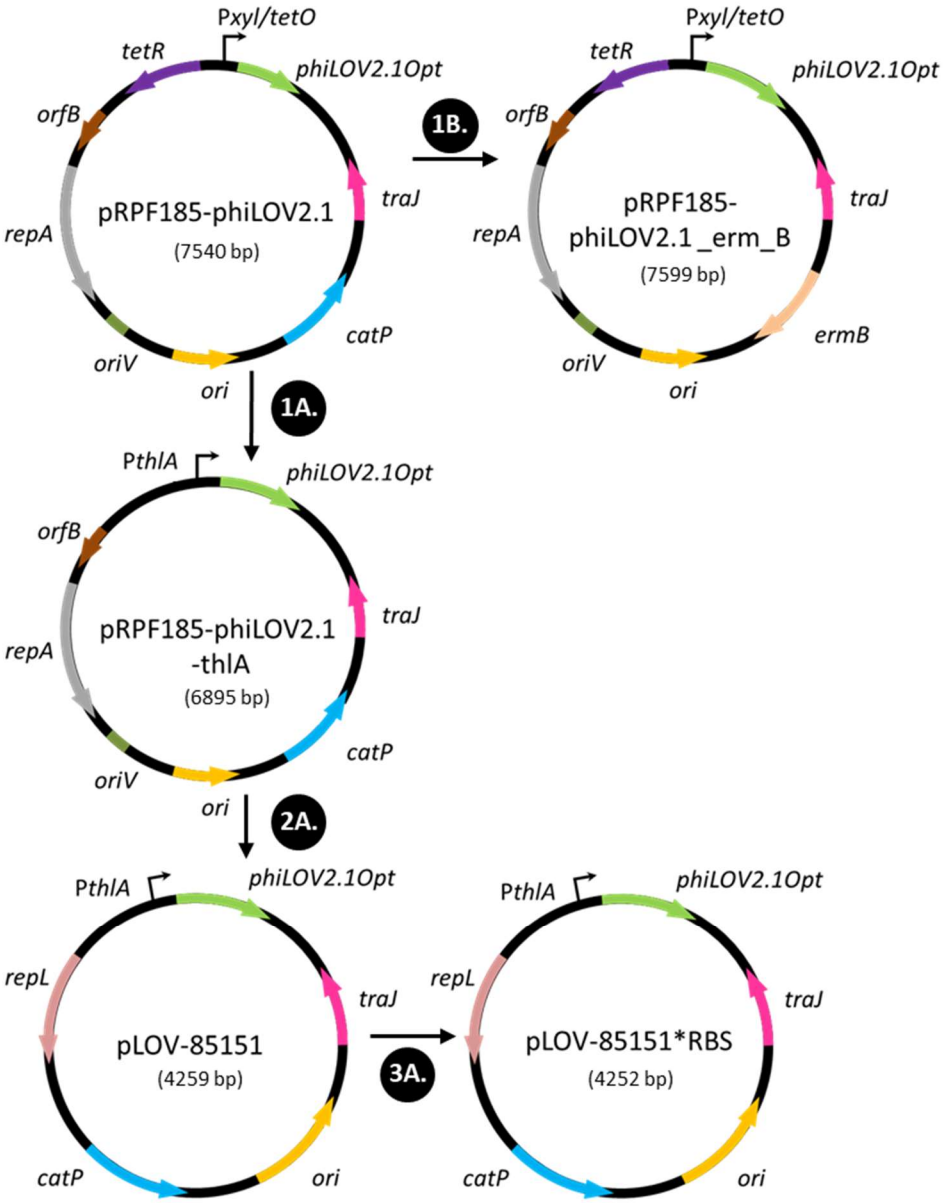


Figure 3.7 Schematic diagram of constructed plasmids of the phiLOV2.1 series.

1A. TetR-P*xyl/tetO* fragment is replaced with P*thlA*. **2A.** Backbone is exchanged for pMTL85151. **3A.** Subcloning of *phiLOV2.1Opt* fragment to replace the RBS upstream of *phiLOV2.1Opt*. **1B.** Replacement of *catP* with *ermB* resistance marker.

3.2.1.1 Promoter replacement

We made a series of changes to pRPF185-phiLOV2.1Opt to try and improve expression levels and minimize overlaps between reporters and selectable markers. The series of

modifications was purposefully replicating the modifications we made to pDSW1728 in order to obtain fully comparable reporter plasmid vectors.

As outlined in section 3.1.1, the tetracycline promoter was not optimal for our purposes and again the P*thlA_Cac* was chosen to replace it. Briefly, plasmid pDSW1728-thlA was cut with SacI-HF and BamHI-HF, dephosphorylated and the resulting 6528bp fragment was gel extracted. It was then ligated to a 367bp gel-extracted *phiLOV2.1*-containing restriction fragment (which was generated by digesting plasmid pRPF185-phiLOV2.1 with BamHI-HF and SacI-HF) to form plasmid pRPF185-phiLOV2.1-thlA . Plasmid pRPF185-phiLOV2.1-thlA was then sequence-verified by Sanger sequencing using primers IG0119 and IG0186.

3.2.1.2 Vector exchange

We then wanted to generate a matched reporter construct of plasmid pMCH-85151 which outperformed pDSW1728-thlA, and replace the backbone of pRPF185-phiLOV2.1. As previously described, pRPF185-phiLOV2.1 contains a pCD6 Gram-positive origin while pMCH-85151 contains a pIM13 origin which may result in higher expression levels in *C. saccharoperbutylacetonicum*.

Therefore, the P*thlA*::*phiLOV2.1Opt* transcription unit from pRPF185-phiLOV2.1-thlA was amplified using primers IG0164 and IG0165, the resulting 948bp PCR product was then PCR purified, digested with AscI and SbfI and gel extracted. Plasmid pMTL85151 was digested with AscI and SbfI and the 3329bp fragment was gel extracted. The two fragments were ligated to form plasmid pLOV-85151. The plasmid was sequence verified by Sanger sequencing using primer IG0119.

3.2.1.3 RBS replacement

As described in the Introduction, RBSs are very important for efficient translation initiation and therefore to translation and overall expression levels. The RBS of *phiLOV2.1Opt* in the pRPF185-phiLOV2.1 plasmid and by extension its derivatives (pRPF185-phiLOV2.1-thlA and pLOV-85151) differed from that of related plasmids pDSW1728 (*mCherryOpt* plasmid) and pRPF185 (parent plasmid) and their derivatives. In addition, the *phiLOV2.1Opt* RBS was less close to the conserved RBS sequence than the one in the latter two plasmids.

Therefore, to reduce the number of variables between the reporter plasmids and potentially improve phiLOV2.1 expression levels, we changed the RBS of pLOV-85151 to match the RBS of pDSW1728, pRPF185 and pMCH-85151. Plasmid pLOV-85151 was used as a template for PCR amplification with primers IG0176 and IG0186, the resulting 595bp PCR product was digested with SacI-HF and BamHI and the 368bp fragment was gel extracted. Plasmid pMCH-85151 was digested with SacI-HF and BamHI (to excise *mCherryOpt*) and the restriction fragment of 3892bp was gel extracted. The two fragments were ligated to form pLOV-85151*RBS. The plasmid was sequence verified by Sanger sequencing using primer IG0119.

3.2.1.4 High-copy E. coli plasmid expressing FbFP

Like with our studies on the mCherry protein, we also generated a maximally expressed *E. coli* positive control plasmid with a LOV-protein variant with *E. coli* codon usage. The protein used for this control was iLOV, which is derived from the LOV2 domain. The phiLOV2.1 variant protein was derived from iLOV through mutagenesis[108].

The plasmid was assembled from commonly used biological parts from the Registry of Standard Parts. BioBrick BBa_K660004 which consists of the iLOV CDS and is carried

on plasmid backbone pSB1C3 was digested with XbaI and PstI and the 336bp fragment was gel extracted. Plasmid backbone pSB1A3 was digested with EcoRI and PstI and the 2155bp fragment was gel extracted and dephosphorylated. Promoter P*lacZYA*-RBS combination (BBa_J04500 (promoter BBa_R0010 + RBS BBa_B0034)), carried on plasmid pSB1C3 was digested with EcoRI and SpeI, the 220bp fragment was gel extracted. The three fragments were ligated together in a single ligation reaction to form plasmid pSB1A3:P*lac*:RBS:iLOV. The plasmid was sequence verified by Sanger sequencing using primer IG0209. *E. coli* colonies growing at 37ºC on LB agar, supplemented with 100µg/mL of carbenicillin, had a noticeable green-yellow hue compared to the wild-type strain.

3.2.1.5 Resistance marker change

As is the case for pDSW1728, plasmid pRPF185-phiLOV2.1 carries *catP* (Cm^R resistance marker). In similar fashion, we decided to replace the selectable marker of this plasmid in order to eliminate the selectable marker-reporter overlap.

To that end, plasmid pRPF185-phiLOV2.1 was amplified with primers IG0011 and IG0024. This amplification encompassed most of the plasmid with the exception of the resistance gene *catP*. The PCR product was dephosphorylated using NEB Antarctic Phosphatase. The *ermB* gene was amplified from pMTL83251 with primers IG0037 and IG0038. The *ermB* PCR product was phosphorylated using T4 kinase and blunt-end ligated with the PCR products from pRPF185-phiLOV2.1 to form plasmids pRPF185-phiLOV2.1_erm_B and pRPF185-phiLOV2.1_erm_B. Plasmids A and B differed in the orientation of the *ermB* gene with respect to the rest of the plasmid. Plasmid B was used for further experiments as it matches the structure of the pTREF constructs described in Chapter 4.

3.2.2 Measurement of phiLOV2.1 expression levels

In order to first verify the emission and excitation spectra of a LOV-domain protein in *E. coli,* an *E. coli* DH5α strain, transformed with pSB1A3:P*lac*:RBS:iLOV and a DH5α wild-type were inoculated into LB and were grown overnight without induction. Live cells were harvested by centrifugation, washed and re-suspended in PBS. Absorbance at 600nm was measured against PBS. The emission and excitation spectra of the cells were then recorded using a FluoroMax 4 fluorimeter. The measurements were performed in a quartz cuvette. The intensity values were then normalized against OD_{600} and the excitation intensity. The comparison between wild-type and iLOV spectra is shown in Figure 3.8. The result showed that the iLOV protein had the expected fluorescence excitation and emission spectra[278].

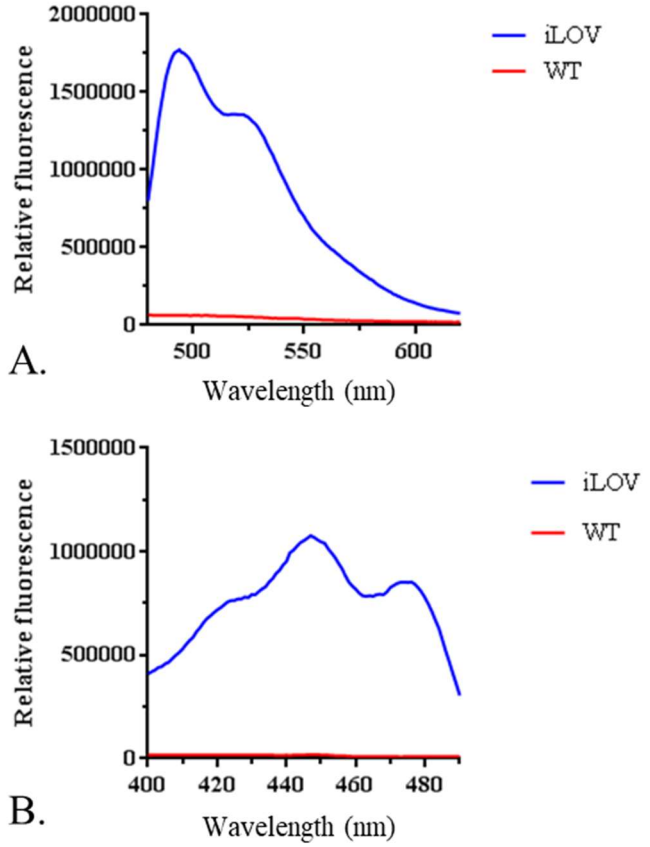


Figure 3.8 Fluorescence emission and excitation spectra of whole DH5α WT cells or cells expressing iLOV protein.

(A) Emission spectrum at λ_{ex}=405nm, recorded between 480nm and 620nm (B) Excitation spectrum at λ_{em}=530nm, recorded between 400nm and 490nm.

3.2.2.1 Flow cytometry of Escherichia coli

The *E. coli* DH5α strain transformed with pSB1A3:P*lac*:RBS:iLOV was grown in LB medium supplemented with 100µg/mL of carbenicillin overnight without the P*lacZYA* inducer IPTG. The plasmid incorporates the very high copy number ColE1-type pUC-derived origin of replication. The expression pattern from this plasmid resembled constitutive expression as we detected a high level of fluorescence in the uninduced sample compared to the *wild-type* and mCherry-expressing control strains (Figure 3.9) using flow cytometry with washed, live cells. A large proportion of the assayed iLOV-expressing cells were fluorescent (76.7%) (Table 3.5) in LB medium.

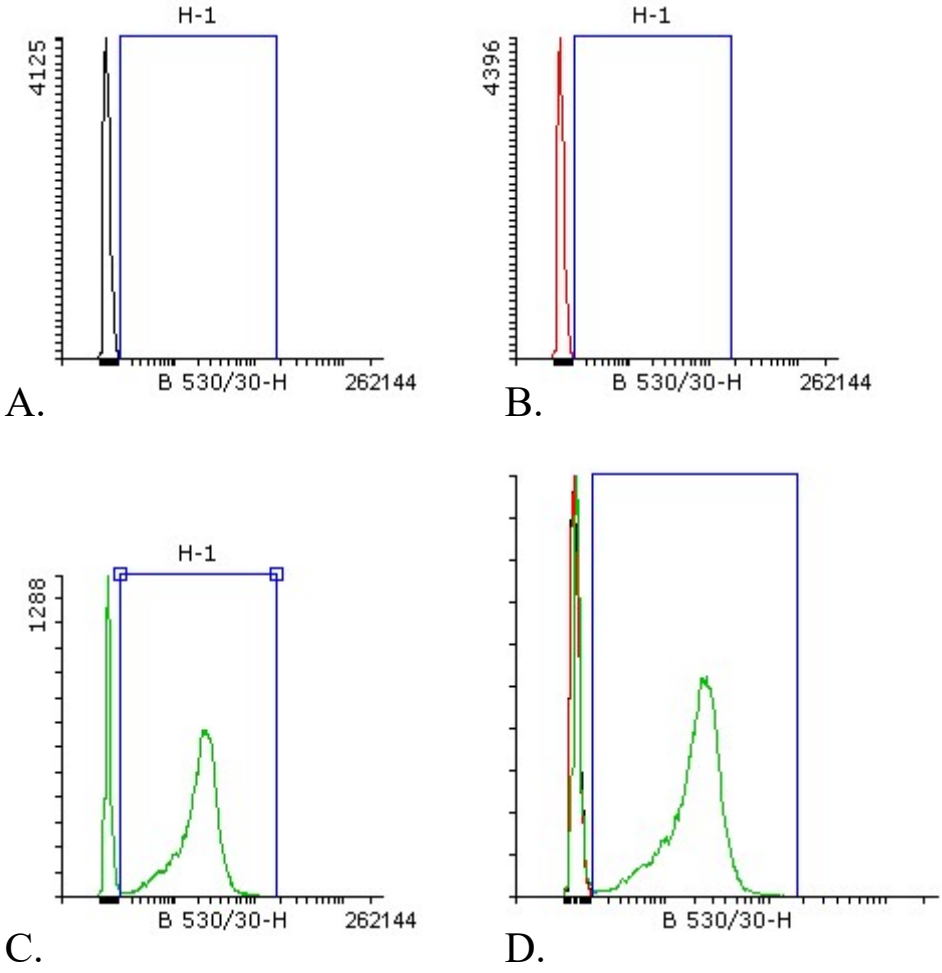


Figure 3.9 Histograms of fluorescence intensity of DH5α strains expressing mCherry and iLOV.

A. DH5α wild-type B. Plac:RBS:mCherry C. Plac:RBS:iLOV D. Overlay histrogram. λ_{ex}=488nm, λ_{em}=515-545nm

Table 3.5 Statistics of fluorescence intensity of DH5α strains expressing mCherry and iLOV. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Gate	Events	% of Visible	Mean	GM	Median	CV
DH5α *wt* (no IPTG)	All	50000	100	35.08	31.96	33	65.69
	H-1	49	0.1	198.71	128.47	117	273.51
P*lac*:RBS: mCherry (no IPTG)	All	50000	100	34.15	31.01	32	64.1
	H-1	39	0.08	249.18	145.29	119	220.98

P*lac*:RBS: iLOV (no IPTG)	All	50000	100	1781.5	771.94	1810.0	85.03
	H-1	38384	76.77	2301.8	1924.6	2245.0	56.58

We then assayed the *E. coli* DH5α strain with pSB1A3:P*lac*:RBS:iLOV in minimal C medium with and without inducer IPTG. The medium was supplemented with 100µg/mL of carbenicillin and cells were grown for 36 hours before harvesting by centrifugation. The results indicated detectable levels of iLOV fluorescence in the unindcued sample compared to the control *wt* strain (Figure 3.10). Interestingly, the proportion of fluorescent cells dropped to 45% (Table 3.6). Futhermore, the sample which was induced from the start of cultivation had an even lower percentage of fluorescent cells (11%) (Table 3.6) with a slightly higher geometric mean of fluorescence levels for the fluorescent sub-population (shown in Gate H1, Figure 3.9 and Table 3.6).

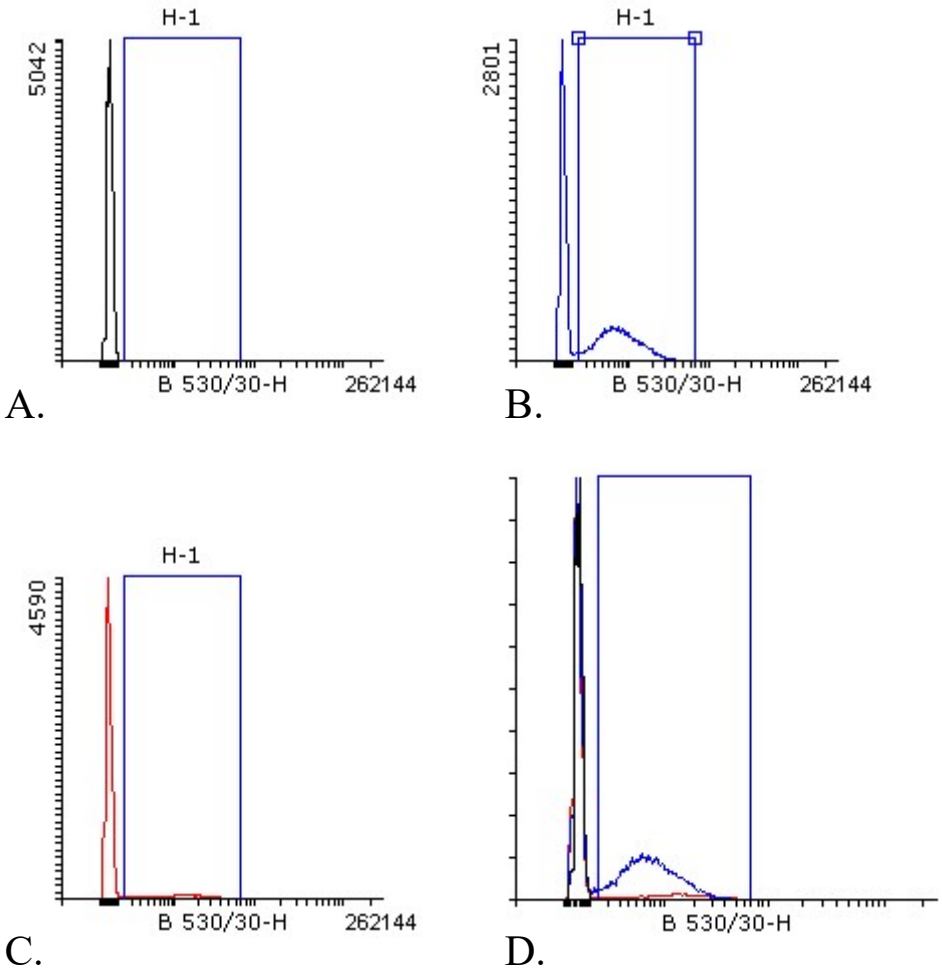


Figure 3.10 Fluorescence intensity of DH5α strains expressing iLOV in minimal medium.

A. DH5α wild-type + IPTG B. Plac:RBS:iLOV w/o IPTG C. Plac:RBS:iLOV + IPTG D. Overlay histrogram. λ_{ex}=488nm, λ_{em}=515-545nm.

Table 3.6 Statistics of fluorescence intensity from *E. coli* strains expressing iLOV in minimal medium. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Gate	Events	% of Visible	Mean	GM	Median	CV
DH5α *wt* (IPTG)	All	50000	100	47.78	45.98	47.00	29.21
	H-1	14	0.03	309.57	251.93	173.00	71.13
P*lac*:RBS: iLOV (no IPTG)	All	50000	100	429.13	150.19	61.00	149.15
	H-1	22545	45.09	890.42	701.17	693.00	75.8
P*lac*:RBS:	All	50000	100	216.68	59.05	44.00	312.29

iLOV (IPTG)	H-1	5590	11.18	1514.2	1091.5	1229.0	79.85

Next, we transformed the *E. coli* CA434 with plasmid pRPF185-phiLOV2.1_erm_B. The pRPF185-phiLOV2.1_erm_B and plasmid REF0-carrying strains were grown in minimal C medium (4% (w/v) glycerol) for 36 hours. Genotypes were inoculated in duplicate – one replicate was induced at onset of growth with 400ng/mL anhydrotetracycline (ATc) while the second culture was not. Cells were harvested and fixed according to the AFR protocol.

The results of the flow cytometry experiment indicated that phiLOV2.1 was expressed under the inducing conditions (Figure 3.11D) and that its expression was tightly repressed in the absence of inducer (Figure 3.11C). The negative control did not have significant fluorescence levels in the uninduced (Figure 3.11A) and induced (Figure 3.11B) states.

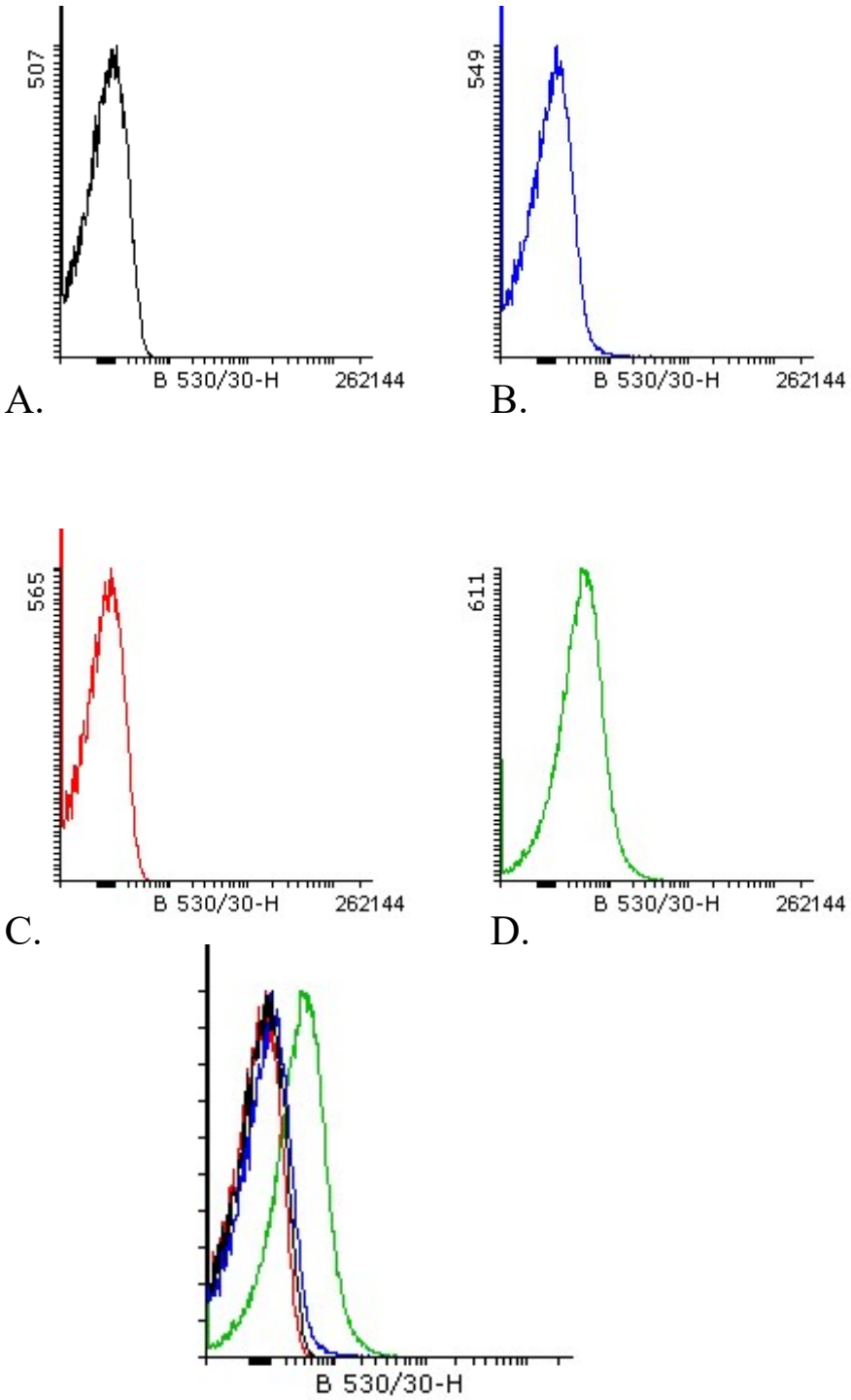


Figure 3.11 Histograms of fluorescence intensity from *E. coli* strains expressing phiLOV2.1.

A. REF0 w/o ATc B. REF0 + ATc C. pRPF185-phiLOV2.1_erm_B w/o ATc. D. pRPF185-phiLOV2.1_erm_B + ATc E.Overlay histogram. λ_{ex}=488nm, λ_{em}=515-545nm

While the fluorescence intensity of the population was lower than with the pSB1A3:P*lac*:RBS:iLOV construct, the fold-increase over the negative control was similar, possibly owing to a lower background fluorescence of the negative control REF0 CA434 strain (Table 3.7).

Table 3.7 Statistics of fluorescence intensity from CA434 strains expressing phiLOV2.1. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Events	% of Visible	Mean	GM	Median	CV
REF0 (no ATc)	50000	100	8.61	9.30	6.40	1562.3
REF0 (ATc)	50000	100	54.23	15.41	36.80	428.63
pRPF185-phiLOV2.1_erm_B (no ATc)	50000	100	9.28	8.76	6.40	1230.8
pRPF185-phiLOV2.1_erm_B (ATc)	50000	100	415.51	206.23	344.00	97.09

In conclusion, both iLOV and its further engineered *C. difficile* codon-optimized derivative of interest, phiLOV2.1, could be actively detected in *E. coli*. A reduction in fluorescence levels in *E. coli* was noticed when using C medium over LB medium. With C medium, the aim was to minimize autofluorescence in the green channel which is known to be influenced by Flavin-rich yeast extracts (common ingredient in complex media)[279].

3.2.2.2 Fluorescence microscopy and flow cytometry in Clostridium saccharoperbutylacetonicum

Next, we conjugated plasmid pRPF185-phiLOV2.1 into *C. saccharoperbutylacetonicum* in order to determine whether we can detect phiLOV2.1 expression in this organism. Liquid cultures of 2x YTG medium (1.3% (w/v) D-glucose) supplemented with appropriate antibiotic were inoculated and grown for 24-48 hours. Two identical subcultures per starter culture were inoculated (in 2x YTG 1.3% (w/v) D-glucose plus antibiotic), one of the subcultures was induced with anhydrotetracycline (400ng/mL final concentration). Culture growth was assessed by absorbance measurements. After performing aerobic fluorescence recovery (AFR) the cells were imaged using epifluorescence microscopy on an Axioskop 40 (Zeiss), with filter set 10 for detection of green fluorescence. We detected a high level of autofluorescence in the wild-type cells (Figure 3.12A). A low level of fluorescence was detectable in the plasmid-containing sample (Figure 3.12C). Exposure times for fluorescence detection varied between 0.01 s and 2s for phase-contrast and epifluorescence, however, it must be noted that the fluorescence images shown in Figure 3.12 are automatically adjusted for brightness and contrast using ImageJ. This result was somewhat promising but proved difficult to reproduce as shown later in this chapter.

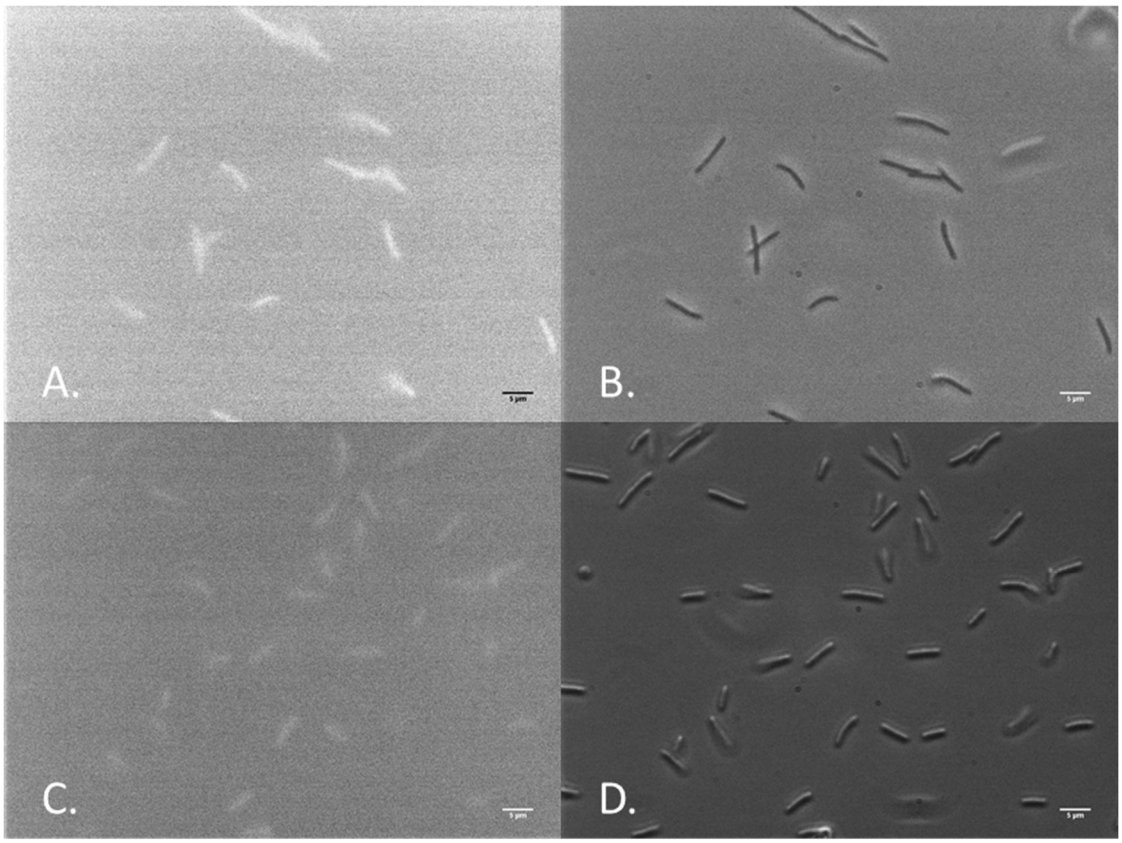


Figure 3.12 Epifluorescence microscopy of *C. saccharoperbutylacetonicum* strains expressing phiLOV2.1.

A. Green fluorescence of pRPF185-phiLOV2.1Opt with inducer – 2s exposure. B. pRPF185-phiLOV2.1Opt with inducer phase contrast image (0.01s exposure). C. Green fluorescence of WT with inducer – 2s exposure. D WT with inducer – phase contrast (0.01s exposure). Images were processed using ImageJ for automatic adjustment of brightness. Scale bars (bottom right of each panel) are 5µm.

The pRPF185-phiLOV2.1 *C. saccharoperbutylacetonicum* strain was grown in 2x YT (1.3% (w/v) D-glucose) supplemented with appropriate antibiotic until an OD_{600} near 1.0 was reached when the cells were induced with anhydrotetracycline (400ng/mL final concentration). One hour after induction the cells were harvested and treated according to the AFR protocol. Wild-type control strain was induced with an appropriate amount of antibiotic solvent. The green fluorescence intensity was quantified using flow cytometry on BD Fortessa X-20. The results indicated a high level of autofluorescence in the negative control that overlapped with that observed for the pRPF185-phiLOV2.1

strain (Figure 3.13). The negative control exhibited a very similar geometric mean of fluorescence to the plasmid-carrying strain (Table 3.8).

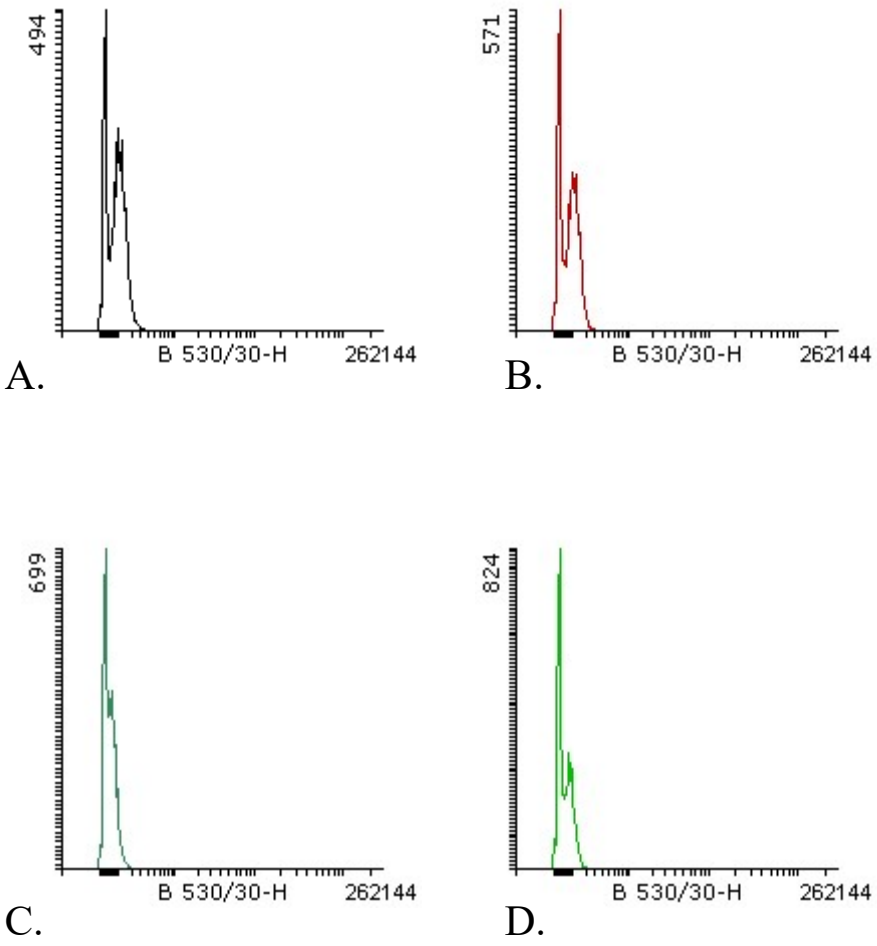


Figure 3.13 Histograms of fluorescence intensity from *C. saccharoperbutylacetonicum* strains expressing phiLOV2.1.

A. WT without inducer B. WT with inducer (ATc at 400ng/) C. pRPF185-phiLOV2.1Opt - no inducer D. pRPF185-phiLOV2.1Opt– with inducer. λ_{ex}=488nm, λ_{em}=515-545nm

Table 3.8 Statistics of fluorescence levels from induced and uninduced WT *C. saccharoperbutylacetonicum* and pRPF185-phiLOV2.1Opt. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	In-ducer	Events	% of Visible	Mean	GM	Median	CV
WT	-	10000	100	82.03	62.20	80.00	65.58
	+	10000	100	76.41	56.67	73.00	69.83
pRPF185-phiLOV2.1	-	10000	100	47.89	39.25	41.00	114.17
	+	10000	100	51.14	39.46	36.00	69.61

We also transformed *C. saccharoperbutylacetonicum* with plasmid pRPF185-phiLOV2.1_erm_B via electroporation in order to test whether we can detect phiLOV2.1 from that plasmid. After induction and expression as for pRPF185-phiLOV2.1Opt, flow cytometry failed to detect increases in fluorescence in the induced sample (Figure 3.14 and Table 3.9).

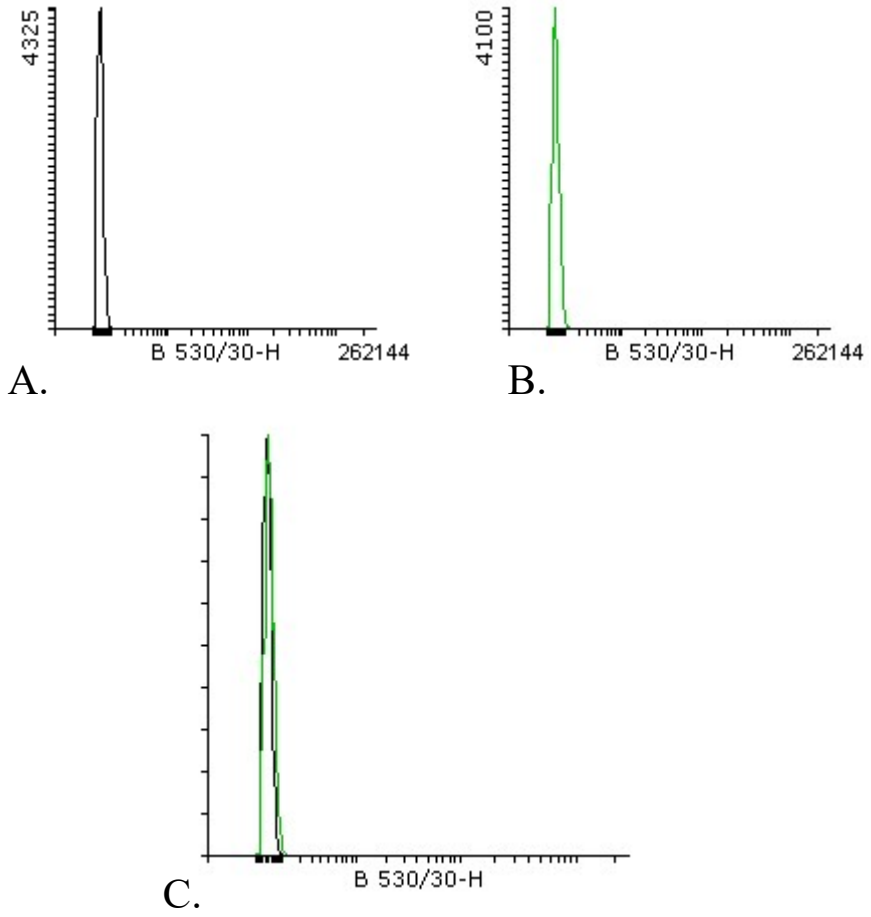


Figure 3.14 Histograms of fluorescence intensity from *C. saccharoperbutylacetonicum* strains expressing phiLOV2.1 from pRPF185-phiLOV2.1_erm_B.

A. pRPF185-phiLOV2.1_erm_B without inducer (ATc). B. pRPF185-phiLOV2.1_erm_B with inducer. C. Overlay of A.and B. λ_{ex}=488nm, λ_{em}=515-545nm.

Table 3.9 Statistics of fluorescence levels from induced and uninduced pRPF185-phiLOV2.1_erm_B. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean.

Strain	Events	% of Visible	Mean	GM	Median	CV
pRPF185-phiLOV2.1_erm_B (no ATc)	50000	100	34.906	32.304	33	86.32
pRPF185-phiLOV2.1_erm_B (ATc)	50000	100	42.811	39.311	41	90.27

In a last effort to detect phiLOV2.1 in *C. saccharoperbutylacetonicum*, we transformed the strain with plasmid pLOV-85151*RBS via electroporation. The backbone of pLOV-85151*RBS had previously resulted in high mCherry expression in *C. saccharoperbutylacetonicum*. The strain was grown in triplicate in RCMm medium, supplemented with 75µg/mL thiamphenicol. After 12 hour incubation from a starting OD_{600} of 0.1, OD_{600} was measured, cells were harvested and AFR was performed. Flow cytometry experiments indicated the lack of measurable expression from the pLOV-85151*RBS strain in *C. saccharoperbutylacetonicum* (Figure 3.15 and Table 3.10).

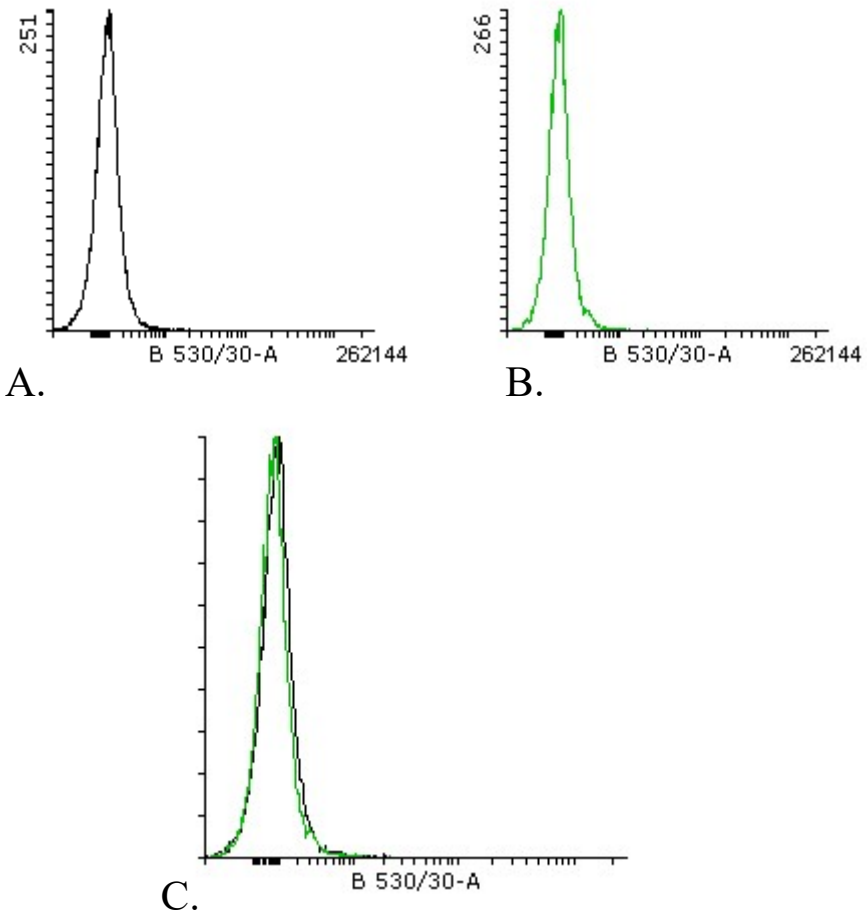


Figure 3.15 Histograms of fluorescence intensity from *C. saccharoperbutylacetonicum* expressing phiLOV2.1 from pLOV-85151*RBS.

A. pMTL85151 B. pLOV-85151*RBS C. Overlay of A.and B. λ_{ex}=488nm, λ_{em}=515-545nm.

Table 3.10 Statistics of fluorescence levels from *C. saccharoperbutylacetonicum* strains carrying pMTL85151 and pLOV-85151*RBS. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean, % VIS - % of Visible.

Strain	Events	% VIS	Mean	GM	Median	CV
pMTL85151	10000	100	76.29±23.89	39.07± 14.92	62.72± 19.33	176.16± 41.56
pLOV-85151*RBS	10000	100	98.01±12.80	47.47± 7.48	71.40± 7.47	219.82± 66.88

3.2.3 Conclusions on the use of phiLOV2.1 as a reporter for gene expression in *C. saccharoperbutylacetonicum*

We were unable to consistently detect phiLOV2.1 expressed from a variety of constructs in *C. saccharoperbutylacetonicum* even though its expression was detectable in *E. coli*. The limiting factors are as yet unknown and could be due to poor transcription elongation, poor stability of the transcript or the protein, poor translation elongation, protein misfolding or lack of co-factor availability. For a discussion of possible reasons for the lack of expression see the Discussion Chapter. Due to the observed minimal expression levels, we moved onto exploring enzymatic reporters of gene expression as a reporter alternative that we regarded as needing less development to meet our requirement for a matched set of reporters for terminator characterization.

3.3 Enzymatic reporter GusA

Enzymatic reporters are an established alternative to fluorescent reporters that offers the advantage of enzymatic amplification while presenting the drawback of bulk population measurements (see Introduction for a detailed list of enzymatic reporters tested in *Clostridium* species). The β-glucuronidase from *E. coli* (GusA) is one of the most

commonly used such reporters across a variety of species and taxa. Importantly, its use was reported in the solventogen *Clostridium acetobutylicum* by multiple independent laboratories[117,125]. We investigated the potential of this enzyme to serve as a reporter for gene expression in *C. saccharoperbutylacetonicum*.

3.3.1 Construction of *gusA* plasmid series

Our work with GusA started with plasmid pRPF185, which was a kind donation by Dr Robert P. Fagan, University of Sheffield. This plasmid is an *E. coli-C. difficile* shuttle vector (aforementioned plasmids pDSW1728 and pRPF185-phiLOV2.1Opt are based on the backbone of this plasmid and are otherwise identical except for the reporter gene and its RBS, the latter differs only in pRPF185-phiLOV2.1Opt), which allows for the tetracycline-inducible expression of *gusA* which is identical to the genomic copy of *E. coli*. We engineered two changes to this plasmid summarized in Figure 3.16.

3.3.1.1 Promoter replacement

In order to replace the tetracycline inducible promoter and repressor with the constitutive strong promoter P*thlA_Cac*, plasmid pRPF185 was digested with SacI-HF and KpnI and the 8243bp restriction fragment was dephosphorylated and gel extracted. It was then ligated to the P*thlA* restriction fragment (as per Section 3.1.1.1) to form plasmid pRPF185-thlA.

3.3.1.2 Vector exchange

We then replaced the backbone of pRPF185 with a modular vector with an alternative Gram-positive origin of replication which we had shown to contribute to higher expression levels in *C. saccharoperbutylacetonicum* using the mCherry reporter. To that end the P*thlA*::*gusA* transcription unit from pRPF185-thlA was amplified using primers IG0164 and IG0165, the resulting 2438bp PCR product was then PCR purified,

digested with AscI and SbfI, and gel extracted. Plasmid pMTL85151 was digested with AscI and SbfI, dephosphorylated and the 3329bp fragment was gel extracted. The two fragments were ligated to form plasmid pGUS-85151.

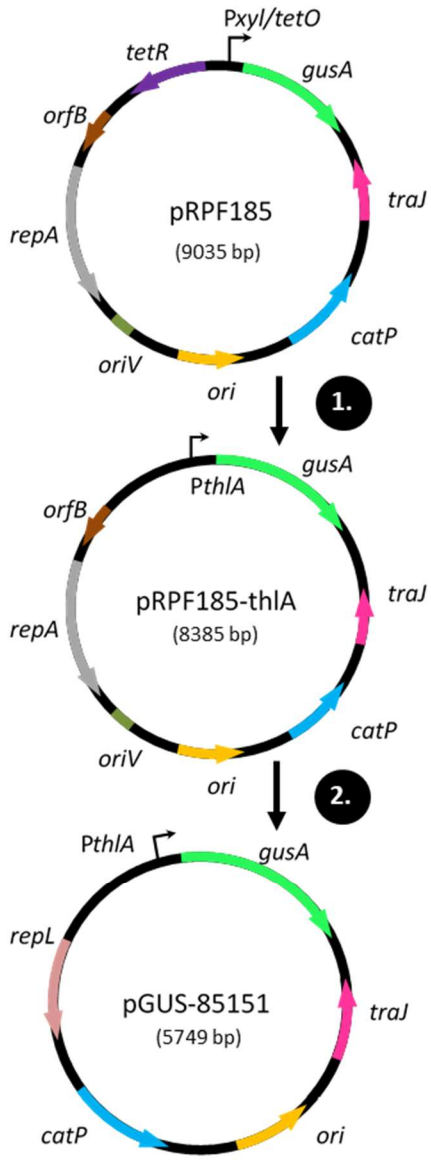


Figure 3.16 Schematic of the two derivatives of pRPF185 constructed for the assessment of GusA function in *C. saccharoperbutylacetonicum*.

1. TetR-P*xyl/tetO* fragment is replaced with P*thlA*. **2.** Backbone is exchanged for pMTL85151.

3.3.2 Measurement of GusA_Ec activity

Detection of GusA activity from our expression plasmid in *Escherichia coli*, *B. subtilis* and *C. saccharoperbutylacetonicum* was assessed, by measuring the hydrolysis of the synthetic substrate p-nitrophenyl-β-D-glucuronide and following absorbance at 405nm. This assay was adapted to a 96-well-plate format with automated absorbance measurement, which afforded improved reliability in terms of quantifying the specific activity of a sample because the main drawback of the test tube-cuvette format is the arbitrary timing of assay stoppage which can contribute to variability between technical replicates. By measuring absorbance at regular intervals and using only the linear phase of the reaction for activity calculations we circumvented that issue.

3.3.2.1 Activity assays in Escherichia coli

Cultures of *E. coli* carrying plasmid pGUS-85151 were grown in LB overnight at 37ºC and harvested. The cells were permeabilized using the toluene permeabilization method and assayed in a 96 well format (see Methods and Materials). The assays produced linear increases in absorbance while a strain containing a control plasmid (pLAC-85151_Syn described later in this chapter) did not produce any noticeable increases in absorbance (Figure3.17A and B).

The intention was to use the GusA reporter in conjunction with a thermostable LacZ enzyme, so we undertook a series of experiments in order to determine the extent to which we can combine sample pre-processing for both reporters.

First the toluene permeabilization method was compared to making cell extracts with the proprietary detergent BugBuster protein extraction protocol. The results indicated

that both methods resulted in measurable activity with toluene-permeabilized cells retaining significantly more of their activity than BugBuster lysates (Figure 3.18A).

Next, samples were heat treated at 60ºC and the enzymatic assay performed at 60ºC temperature, as per the protocol for thermostable LacZ (discussed later in this chapter). The toluene permeabilized cells were possibly removed during treatment (possibly during centrifugation after the heating step) which resulted in a significant loss of activity (Figure 3.18B). BugBuster lysates on the other hand retained their activity, displaying higher activity values (GusA is more active at 60ºC and has moderate thermostability) (Figure 3.18B). This result indicated that we could use either BugBuster or toluene for measurement of GusA activity but it would be preferable to use BugBuster for the LacZ reporter. In order to keep the measurements comparable with the second reporter and streamline sample processing we began using BugBuster without heat treatment with incubation at 37ºC for subsequent GusA measurements.

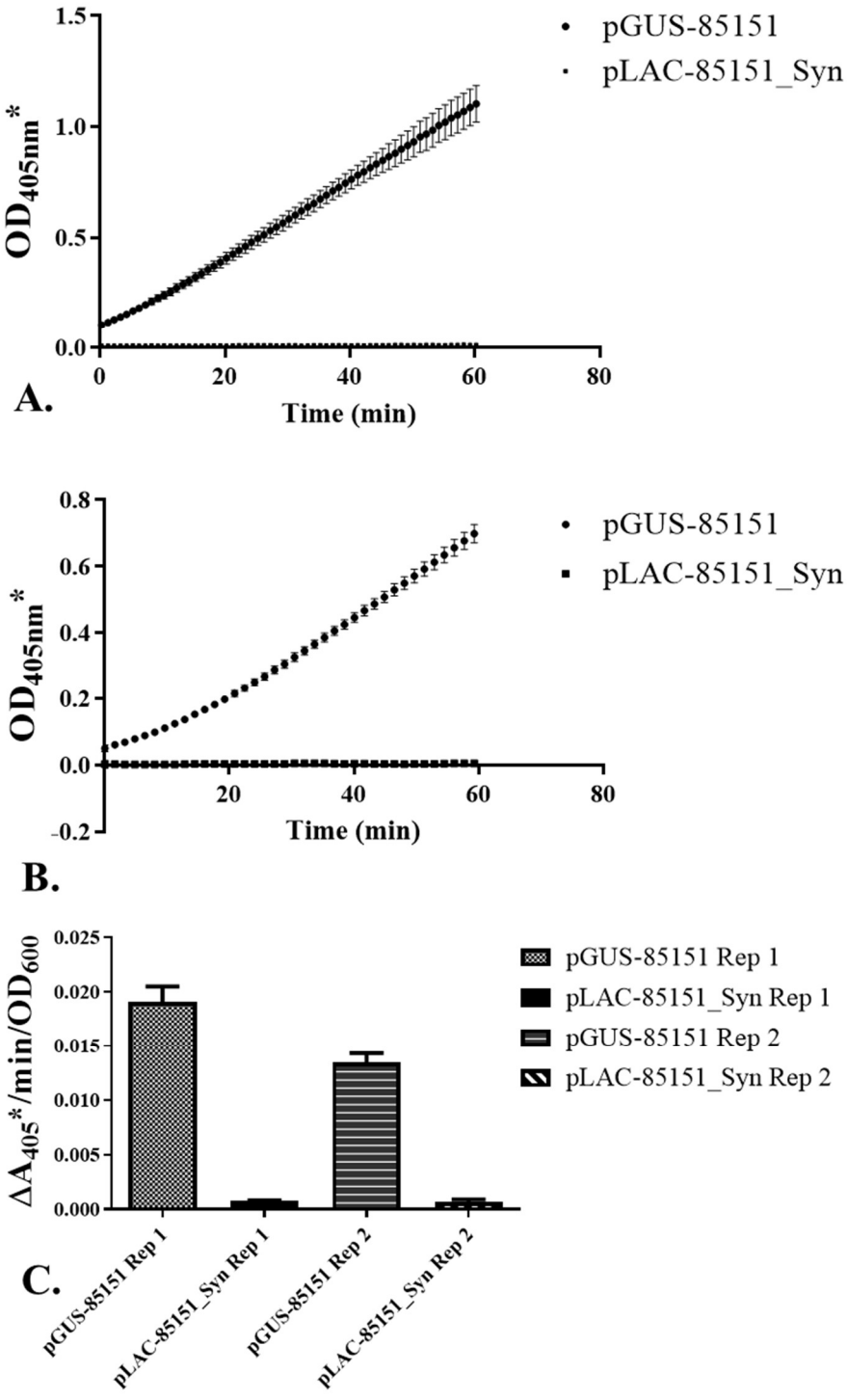


Figure 3.17 Results of assays with pGUS-85151.

A. Linear increase in absorbance with time during assay of β-glucuronidase. Error bars – standard deviation, N=3 (technical replicates) **B.** Biological replicate of A: Error bars – standard deviation, N=3 (technical replicates). **C.** Increase in absorbance over time per unit OD_{600}. * - Absorbance at 405nm is not normalized to 1cm pathlength. Error bars – standard deviation, N=3 (technical replicates).

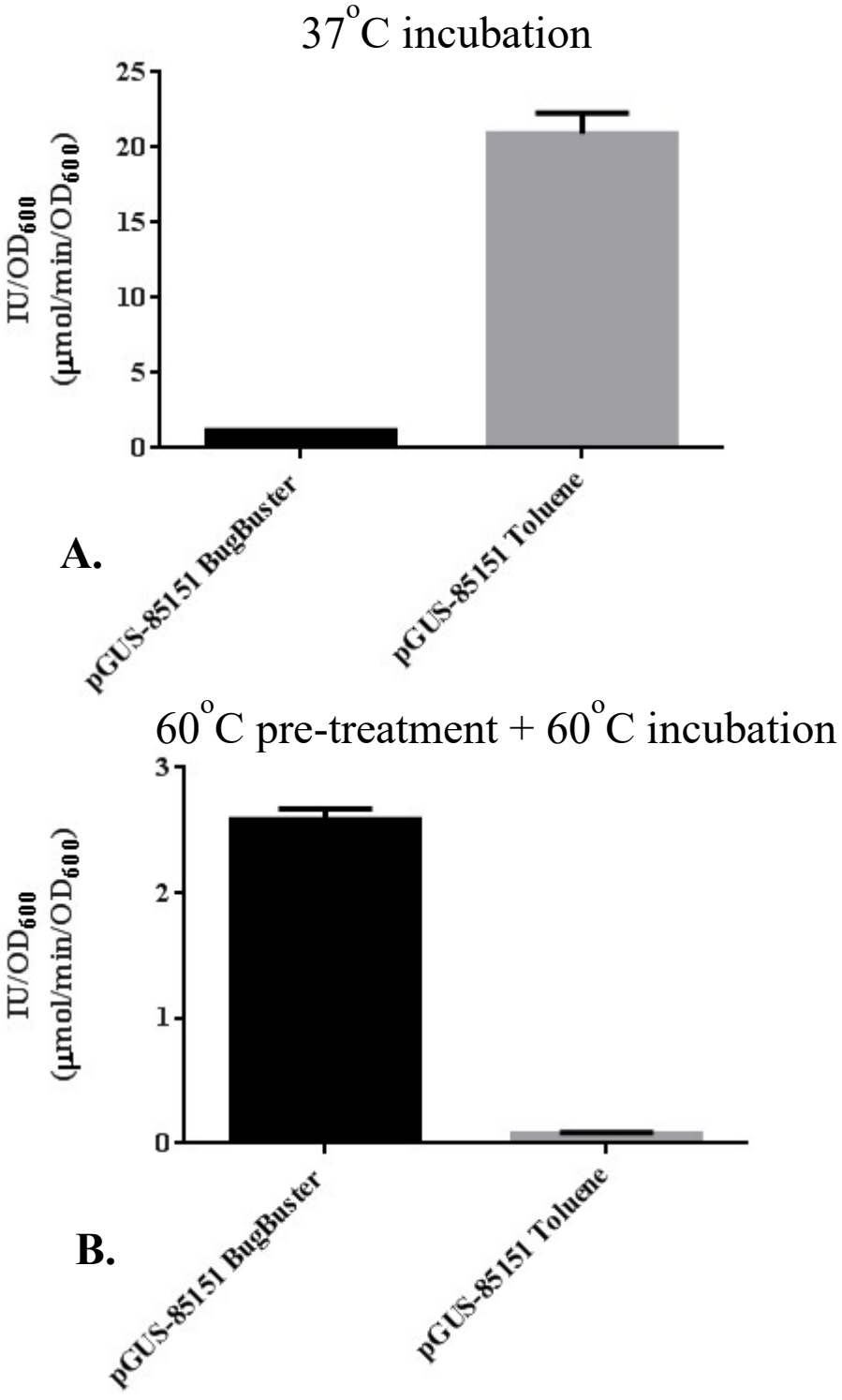


Figure 3.18 Comparison of β-glucuronidase activity at 37ºC versus 60ºC.

β-glucuronidase activity at 37ºC (A) versus 60ºC (B) as well as a comparison of Bug Buster extraction and toluene permeabilization. Note the significantly different y-axis scales on the two plots. Error bars – standard deviation of technical replicates (N=2).

To verify the assay's linear range when using BugBuster we performed a standard curve with dilutions of the p-nitrophenol product in the assay buffer with appropriate amount of BugBuster reagent added (Figure 3.19). We found the assay to be linear up to absorbance values of around 1.0.

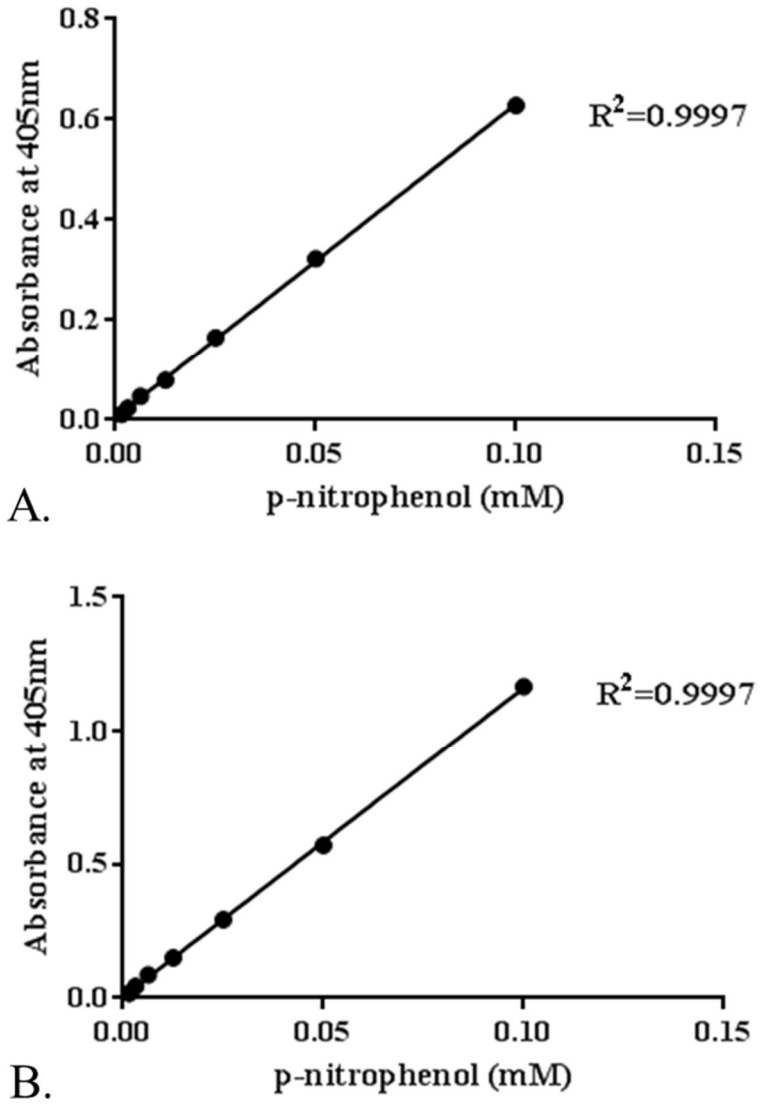


Figure 3.19 Standard plots of p-nitrophenol concentration against A_{405}.

A. Absorbance versus concentration of p-nitrophenol in Buffer Z and Bug Buster from 96 well plate. **B.** Pathlength-normalized absorbance versus concentration of p-nitrophenol in Buffer Z and Bug Buster. Absorbance displayed a linear relationship to

concentration up to A_{405}=4.48 (R^2 = 0.9967, simple linear regression analysis). Data displayed is from a representative experiment.

3.3.2.2 Activity assays in Bacillus subtilis

To test the wider subsequent use of this reporter, expression levels were also assessed in *Bacillus subtilis* because of its conserved gene expression regulatory features and the aforementioned pIM13 origin (originally derived from a strain of *B. subtilis*). The pGUS-85151 plasmid was transformed into *B. subtilis* str. 168 using induced natural competence (see Materials and Methods). A strain containing the empty vector had negligible background activity and in the presence of *gusA* gave significant activity, demonstrating that the regulatory elements used for a *Clostridium* species design are active in *B. subtilis* (Figure 3.20).

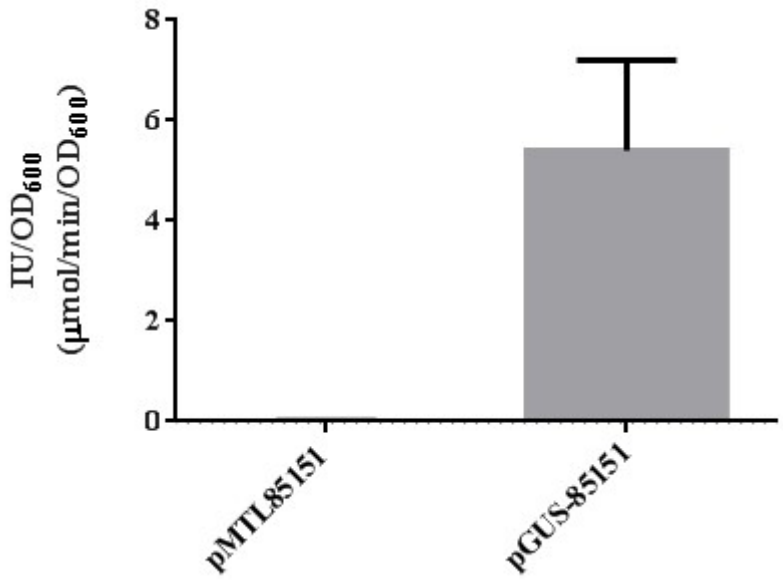


Figure 3.20 The β-galactosidase activity of *B. subtilis* strains with plasmids pMTL85151 (negative control) and pGUS-85151. Error bars – standard deviation of biological replicates (N=2) and technical replicates (N=2).

3.3.2.3 Activity assays in Clostridium saccharoperbutylacetonicum

To assess the function of the GusA reporter in *C. saccharoperbutylacetonicum*, the pGUS-85151 plasmid was introduced by electroporation along with negative control plasmids pMTL85151 and pEMPTY. The β-glucuronidase activity measurements indicated that GusA is expressed and its activity is detectable above the low but detectable background activity level (Figure 3.21).

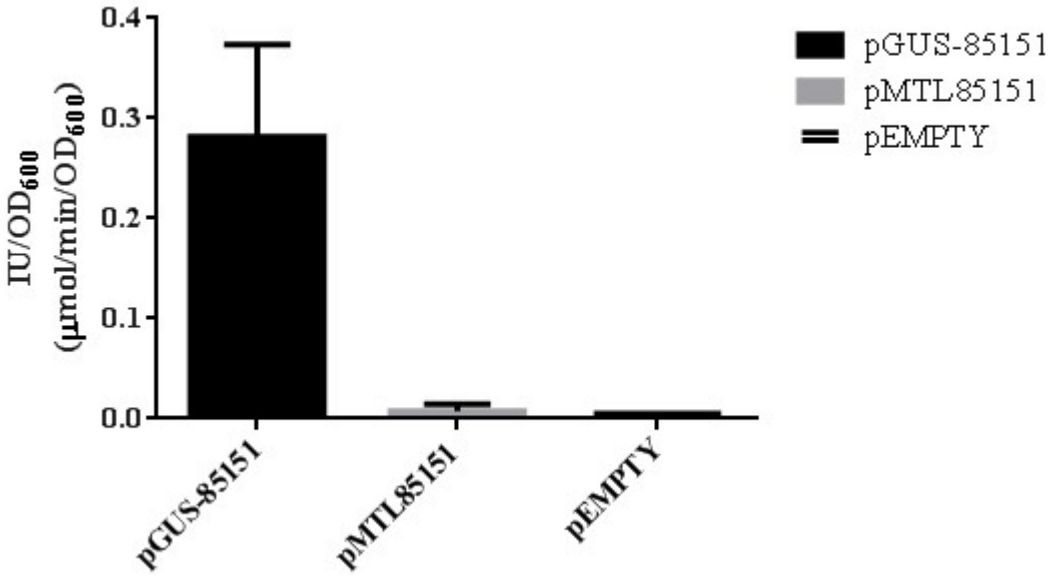


Figure 3.21 β-glucuronidase activity of *C. saccharoperbutylacetonicum* strains with plasmids pMTL85151, pEMPTY (negative controls) and pGUS-85151.

Error bars - standard deviation of biological replicates (N=3) and technical replicates (N=2).

Based on these results, we concluded that GusA is a suitable reporter of gene expression in *C. saccharoperbutylacetonicum*. Although the β-glucuronidase background activity was highest in *C. saccharoperbutylacetonicum* (when compared to *E. coli* and *B. subtilis*) it was low enough to allow for terminator characterization.

3.3.2.4 Conclusion on the use of GusA_Ec

The β-glucuronidase reporter was active in the three species tested when expressed from the pGUS-85151 construct, hinting at the potential of this reporter system to be used for comparative studies on regulatory elements – namely terminators in this study – for the identification of species/taxon-specific features with functional significance.

3.4 Enzymatic reporter thermostable LacZ_Tts

We then moved onto working on developing the thermostable β-galactosidase reporter from *Thermoanaerobacterium thermosulfurigenes* for work in *C. saccharoperbutylacetonicum*. This reporter had been used in *C. acetobutylicum* by several laboratories most of which obtained the physical DNA construct from one of the original laboratories that developed it with one laboratory reportedly cloning it directly from source organism DNA themselves[280]. A synthetic copy of the gene was purchased based on the sequence in the published sequences from 1991[281], revised in 2004[134]. However, we found the encoded protein to be largely inactive (data shown in Section 3.4.3.1). Modifications to the RBS did not produce a big increase in activity and assuming that sequence reported in 2004 is not correct, the gene was cloned from the source organism *Thermoanaerobacterium thermosulfurigenes* EM1 (=DSM3896) genomic DNA. The published sequence of the gene was found to diverge from the genomic copy by several non-synonymous mutations including frameshift mutations at the C-terminus that were not reported or corrected in the literature since 2004. The genomic copy of the gene had activity that matched that reported in previous studies.

The series of constructs we made to investigate the potential of LacZ as a reporter for *C. saccharoperbutylacetonicum* is outlined in Figure 3.22.

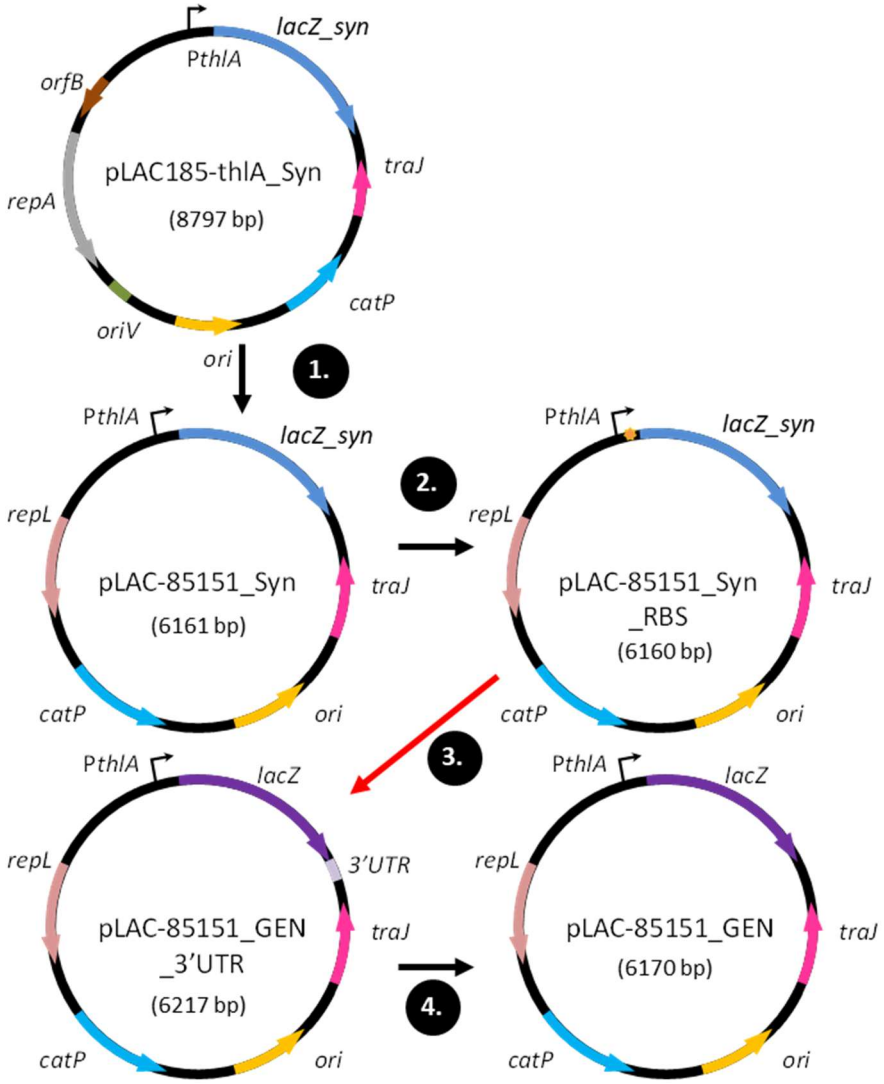


Figure 3.22 The LacZ_Tts series of plasmids constructed to reach pLAC-85151_GEN .

1. The backbone of pLAC185-thlA_Syn is exchanged for pMTL85151. **2.** Site-directed mutagenesis to modify the RBS spacer upstream of *lacZ_syn*. **3.** Cloning of genomic *lacZ* (including its 3' UTR) copy from *T. thermosulfurigenes* and replacement of *lacZ_syn* with it (red arrow indicates the significance for activity – see main text). **4.** The 3' UTR of *lacZ* is removed to increase construct similarity to pGUS-85151.

3.4.1 Cloning of synthesized, partially codon-optimized *lacZ_syn* from *Thermoanaerobacterium thermosulfurigenes*

3.4.1.1 Design and cloning of gBlocks

The gene sequence of the thermostable β-galactosidase gene from *Thermoanaerobacterium thermosulfurigenes* EM1 (=DSM 3896) was obtained from GenBank (M57579.1). The sequence of the gene (coordinates 252…2403 of accession M57579.1) was divided into three fragments for DNA synthesis by Integrated DNA Technologies, Inc. The synthesis option chosen was double-stranded linear DNA fragments (gBlocks). Due to restrictions of the sequence that IDT were able to synthesize (as of January, 2018) two regions of the gene were re-designed to increase their GC-content. To this end, the IDT codon optimization tool was used with settings for *C. acetobutylicum*. The codon optimization algorithm increased the 252...377 region's (per M57579.1) GC-content from 24.6% to 40.48% and the 828...956 region's (per M57579.1) GC-content from 20.16% to 28.68%. The three gBlocks were designed with 2 overlapping regions of 42bp and 33bp predicted melting temperatures of with 65.8ºC and 65.2ºC (SantaLucia, 1999), respectively. All three gBlocks were 751bp long. The gBlock DNA was resuspended in ddH_2O, following the manufacturer's instructions, to a concentration of 50ng/µL (see Materials and Methods). The three gBlocks were assembled into a single fragment using overlap-extension PCR of 2178bp. This fragment was then PCR amplified using primers IG0087 and IG0088 and gel extracted. The purified PCR fragment was used in a BpiI-mediated Golden Gate Assembly reaction with plasmid DVA_CD to generate plasmid Clo_LacZ_syn_CD.

3.4.1.2 Addition of C-terminus-coding sequence

After construction of Clo_LacZ_syn_CD was completed, it was realized that the sequence of M57579.1 contains known sequence errors that result in non-sense

mutations and a truncation of the predicted C-terminus of the LacZ protein by 31 amino acids that had been demonstrated to be critical for activity. Therefore, correction of that error was necessary. Plasmid Clo_LacZ_syn_CD was used as a template for PCR with primers IG0087 and IG0089. Overlapping oligonucleotides IG0077 and IG0078 were annealed and phosphorylated. The two fragments were used together with DVA_CD in a BpiI-mediated Golden Gate Assembly reaction with plasmid DVA_CD to generate plasmid Clo_LacZ_syn_CD_full.

3.4.1.3 Correction of SNP by site-directed mutagenesis

Plasmid Clo_LacZ_syn_CD_full was Sanger sequenced (GATC Biotech) using primers IG0209, IG0210, IG0083 and IG0084. An unintended mutation was found at position 2114 (per M57579.1). To correct the mutation, plasmid Clo_LacZ_syn_CD_full was amplified with primers IG0153 and IG0154, the reaction was PCR purified, digested with DpnI and transformed into chemically competent *E. coli* cells. Plasmid was purified from single colonies and the corrected sequence of Clo_LacZ_syn_CD_full was verified by Sanger sequencing using primers IG0209, IG0210, IG0083 and IG0084.

3.4.1.4 Sub-cloning of Clo_LacZ_syn

The *lacZ* gene was then sub-cloned into an expression vector.

Briefly. plasmid Clo_LacZ_syn_CD_full was used as a template for PCR with primers IG0093 and IG0090. The 2285bp product was PCR purified and digested with SacI-HF and BamHI. Plasmid pDSW1728-thlA was digested with SacI-HF and BamHI, dephosphorylated and the 6528bp fragment was gel extracted. The two gel extracted fragments were ligated to form pLAC185-thlA. The resulting plasmid was verified by Sanger sequencing using primers IG0119, IG0186, IG0083 and IG0084.

3.4.1.5 Vector exchange

As results with mCherry in pMTL85151 indicated that this vector results in higher expression levels in *C. saccharoperbutylacetonicum* and allowed expression in *B. subtilis* we sub-cloned the *lacZ* expression cassette into that background.

Briefly, the P*thlA*::*lacZ_syn* transcription unit from pLAC185-thlA was amplified using primers IG0164 and IG0165, the resulting 2850bp PCR product was then PCR purified, digested with AscI and SbfI and gel extracted. Plasmid pMTL85151 was digested with AscI and SbfI, dephosphorylated and the 3329bp fragment was gel extracted. The two fragments were ligated to form plasmid pLAC-85151_Syn. The plasmid was sequence verified by Sanger sequencing using primers IG0119, IG0186, IG0083 and IG0084.

3.4.1.6 RBS modification

As expression levels were very low from pLAC-85151_Syn, we sought to identify the causes of this – one possible reason was an extra A in the RBS spacer when compared to pGUS-85151, pMCH-85151 and pLOV-85151*RBS.

In an attempt to improve expression, we modified the RBS spacer length in pLAC-85151_Syn. Still the activity of pLAC-85151_Syn_RBS remained very low.

Briefly, plasmid Clo_LacZ_syn_CD_full was used as a template for PCR amplification with primers IG0082 and IG0090, the resulting 2284bp PCR product was digested with SacI-HF and BamHI and was gel extracted. Plasmid pMCH-85151 (section 3.5.3) was digested with SacI-HF and BamHI and the restriction fragment of 3892bp was gel extracted. The two fragments were ligated to form pLAC-85151_Syn_RBS. The plasmid was sequence verified by Sanger sequencing using primers IG0119, IG0186, IG0083 and IG0084.

3.4.1.7 BLAST search and alignment with related proteins

To troubleshoot the expression issues along with protocol modifications, a literature and database search was undertaken.

A BLAST search of the full length amino acid sequence of the LacZ_syn protein returned a hit from a related organism with a full genome sequence – *Thermoanaerobacterium xylanolyticum* LX-11. The predicted protein (accession WP_013788775.1) had query coverage of 99% and identity of 88.41%. Meanwhile, the translation (accession P26257.1) of the uncorrected sequence that LacZ_syn was based on - had a coverage score of 95% and identity of 100% (reflective of the 31 amino acid truncation of the C-terminus). Upon closer inspection, the identity of the C-terminus of WP_013788775.1 was found to be different from the query and also 2 amino acids longer. This was surprising, as the aforementioned experimental evidence indicated that the C-terminus is critical for activity and was thus expected to be conserved.

A nucleotide sequence alignment revealed possible deletions leading to a frameshift -the *T. xylanolyticum*-like C-terminus amino-acid sequence was found in an alternative open-reading frame in the *T. thermosulfurigenes* nucleotide sequence (Figure 3.23 and Figure 3.24).

Figure 3.23 Alignment of *lacZ* nucleotide sequences. An alignment of the published *lacZ_Tts* sequence (based on M57579.1[281] with correction[134]) and a homologue - *lacZ_xyl* (AEF18046.1) - identified through BLAST search of full length protein highlights possible frameshifts in the 3' end of *lacZ_Tts*.

Next, the divergent C-terminal amino acids of LacZ_syn and LacZ_xyl were used as queries in independent BLAST searches. The results of the LacZ_syn-derived C-terminus did not return any predicted β-galactosidases. However, the LacZ_xyl C-terminus (last 10 amino acids) produced hits to several predicted β-galactosidases from Clostridia (Figure 3.24).

Figure 3.24 **Alignment of C-termini of putative glycoside hydrolases from Clostridia.** The LacZ_Tts (P26257.1) C- terminus is aligned to four putative glycoside hydrolases from Clostridia, three of which were obtained by BLASTP search of the translated *lacZ_xyl* (AEF18046.1) 3' end sequence. WP_096145356.1 - glycoside hydrolase family 2 protein from *Clostridium chauvoei* ;WP_011591873.1 - glycoside hydrolase family 2 protein from *Clostridium perfringens*; ORX22838.1 - hypothetical protein BVF91_09310 from *Thermoanaerobacterium* sp. PSU-2 (fragment).

These results indicated that it is likely that the C-terminus of the LacZ_syn protein is inaccurate. It was reasoned that this inaccuracy likely stems from sequencing errors in the original publication. This was not particularly surprising given that researchers had already demonstrated sequencing errors in the original sequence. Other sequencing errors resulting in missense mutations along the length of the protein could not be excluded. Additionally, next-generation sequencing data often contains misassemblies. Therefore, to validate the prediction of the true nucleotide sequence of *lacZ*'s 3' end (corresponding to the LacZ C-terminus) and to ensure a functional copy of the *lacZ* gene was available it was decided to re-isolate it from the source organism. An alternative explanation for the low activity of the *lacZ_syn* constructs would be the codon optimization. As mentioned in 3.4.1.1, codon-optimization was kept to a minimum to avoid changing the properties of the coding sequence, the reasoning being

that this coding sequence had already been demonstrated to be functional in both *E. coli* and *C. acetobutylicum*.

3.4.2 Cloning of lacZ_gen from *Thermoanaerobacterium thermosulfurigenes* genomic DNA

3.4.2.1 Cloning of lacZ_gen

Genomic DNA of *T. thermosulfurigenes* EM1 (=DSM 3896) was obtained from the German Collection of Microorganisms and Cell Cultures GmbH (Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ)). The gene coding for the LacZ protein was amplified with primers IG0095 and IG0094. The PCR product (expected size of 2341bp) was then PCR purified, digested with SacI-HF and BamHI and gel extracted. The fragment was ligated into the 3892bp SacI-HF and BamHI-digested fragment of the pMCH-85151 plasmid to form plasmid pLAC-85151_GEN_3'UTR. The plasmid was verifed using Sanger sequencing with primers IG0119, IG0186, IG0083 and IG0084.

3.4.2.2 Sequencing of lacZ cloned from T. thermosulfurigenes EM1 (=DSM 3896)

Within the 2281bp coding region of the *lacZ* gene we found a total of 7 sequencing errors (if we assume that the strain we received is isogenic with the one studied) in the original gene sequence (6 in addition to the 1 described in the literature) (Figure 3.25). That results in a calculated 99.65% accuracy of the original sequencing. Three of these errors impair the C-terminal amino-acid sequence.

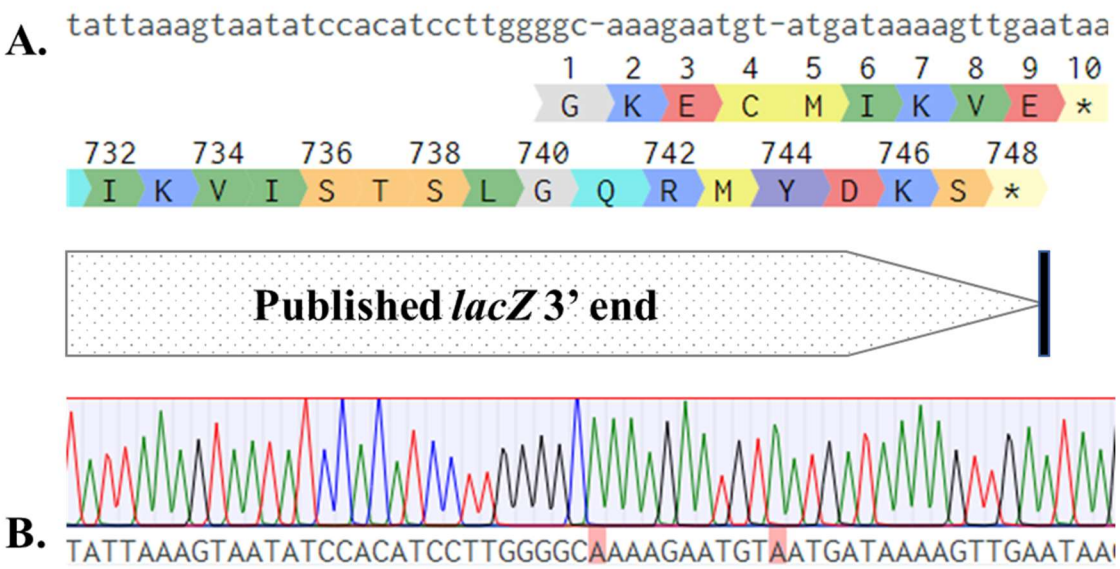


Figure 3.25 Sanger sequencing trace files (from pLAC-85151_GEN_UTR) aligned to published *lacZ_Tts* sequence showing the positions of two unreported deletions resulting in frameshift mutations. A. Top- nucleotide sequence of published *lacZ_Tts* 3' end. Top Middle – Amino acid sequence predicted by us based on homology to related proteins. Top Bottom – amino acid sequence based on published nucleotide sequence. Bottom – illustration of the *lacZ* 3' end according to the published data. Black vertical line marks the end of the CDS. **B.** Sanger sequencing obtained from recloning *lacZ_Tts* shows locations of insertions relative to the published sequence.

3.4.2.3 Removal of 3'UTR from pLAC-85151_GEN

Since the *lacZ* genomic region was amplified to capture the entire C-terminal region it also included the UTR region bordering the downstream gene and the start of the downstream gene's coding sequence. In order to avoid unexpected effects, this region (referred to as 3'UTR) was removed.

Briefly, the genomic copy of *lacZ* gene was PCR amplified from plasmid pLAC-85151_GEN_3'UTR with primers IG0095 and IG0086. The PCR product (expected size of 2341bp) was then PCR purified, digested with SacI-HF and BamHI and gel extracted. The fragment was ligated into the 3892bp SacI-HF and BamHI-digested

fragment of the pMCH-85151 plasmid to form plasmid pLAC-85151_GEN which was verified by Sanger sequencing using primers IG0119, IG0186, IG0096 and IG0084.

3.4.2.4 Generation of a negative control plasmid without residual β-galactosidase activity

Plasmid pMTL85151 which serves as the vector backbone for the constitutive single reporter constructs contains a multiple cloning site (MCS) that consists of a promoterless *lacZα* coding region. This plasmid was still able to complement *lacZΔM15* strains and result in hydrolysis of XgaL on LB-agar plates, indicating leaky *lacZα* expression. Indeed, testing of the CD0164 terminator upstream of the *lacZα* (see Chapter 5) revealed it to have weak promoter-like activity in *E.coli* and was very close to the no terminator control in *B. subtilis*. This is unexpected as the structure of this terminator is canonical and it is derived from a convergent gene region. Therefore, to remove the *lacZα* and thus background activity in *E. coli*, pMTL85151 was digested with XbaI and NheI, the 3515bp fragment was gel extracted and self-ligated to form plasmid pEMPTY-85151 (referred to as pEMPTY for brevity).

3.4.3 Measurement of LacZ_Tts activity

3.4.3.1 Activity assays with LacZ_Tts in E. coli

We first began testing for thermostable β-galactosidase activity using plasmids pLAC-85151_Syn and pLAC-85151_Syn_RBS. However, as alluded to earlier in the chapter the plasmids resulted in non-detectable to very low substrate hydrolysis.

After cloning the genomic copy encoding LacZ_Tts (referred to as *lacZ_Gen*), we observed rapid development of deep blue color on Xgal agar plates at 37ºC and after transfer to 4ºC. We then compared the activity of these constructs using a quantitative activity assay with o-nitrophenyl-β-d-galactopyranoside (ONP-Gal) as the substrate.

Two methods of cell preparation were compared – Bug Buster protein extraction and toluene permeabilization (see Materials and Methods) as well as two lysate treatment incubation temperature combinations. A standard curve of o-nitrophenol was produced for one of the Bug Buster reaction conditions (Figure 3.26). Interestingly, the two methods of cell preparation produced almost inverse results when used with the two different lysate treatment-incubation temperature combinations. The Bug Buster extraction resulted in similar activity levels for the positive control pADAPT (expressing *E.coli lacZ α*-peptide, complementing a *lacZΔM15* deletion in DH5α) and the pLAC-85151_Gen plasmid at 37ºC without heat pre-treatment (Figure 3.27A and Figure 3.29B). Whereas with regard to the toluene permeabilized samples, the pADAPT plasmid had a significantly higher activity (Figure 3.28A and Figure 3.29A). On the contrary, when incubated at 60ºC with heat pre-treatment the Bug Buster-extracted sample of pLAC-85151_Gen (Figure 3.27B) had similar activity to the toluene treated sample (Figure3.28B). Heat treated pADAPT retained more of its activity in Bug Buster than with toluene – similar to GusA (previous chapter) while they both had higher activity with toluene without the heat pre-treatment. While toluene permeabilization gave satisfactory results and higher activity levels, there were issues associated with its use such as a higher degree of variability between technical repeats due to issues such as cell clumping and loss of activity after heat treatment (as mentioned in ***Section 3.3.2.1***). Briefly, the toluene permeabilization step results in a cell suspension which when heated (as per the modified protocol with heat inactivation, see ***Section 2.9***) tends to clump together (data not shown) and were removed by centrifugation. It appears that during heat inactivation at 60ºC the protein content of the cells was not extracted in the supernatant. A modification of the protocol may be able solve this issue. However, Bug Buster was chosen for subsequent experiments with both β-galactosidase and β-

glucuronidase assays because it allowed for a more streamlined process and produced better activity with heat-treated samples.

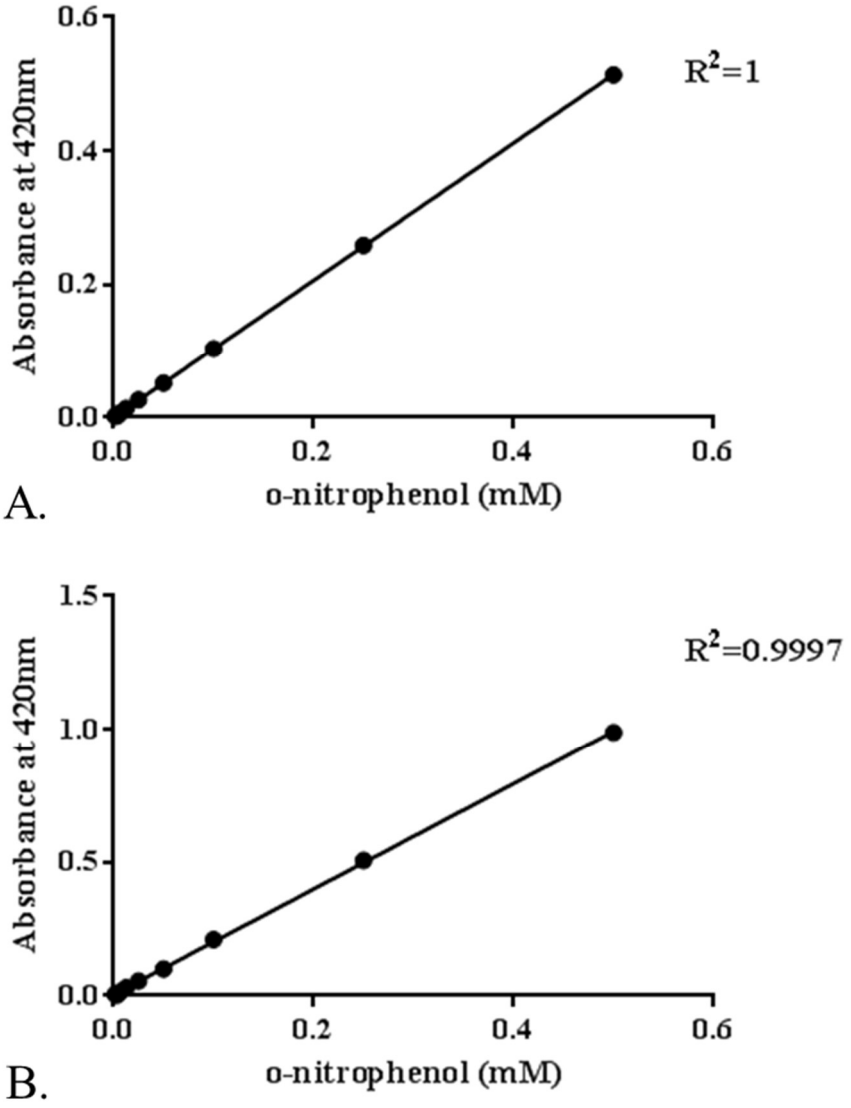


Figure 3.26 Standard curves of o-nitrophenol concentration versus A_{420}.

A. Absorbance versus concentration of p-nitrophenol in Buffer Z and Bug Buster from 96 well plate. B. Pathlength-normalized absorbance versus concentration of o-nitrophenol in Buffer Z and Bug Buster. Data displayed is from a representative experiment. R^2 was calculated as described in Materials and Methods.

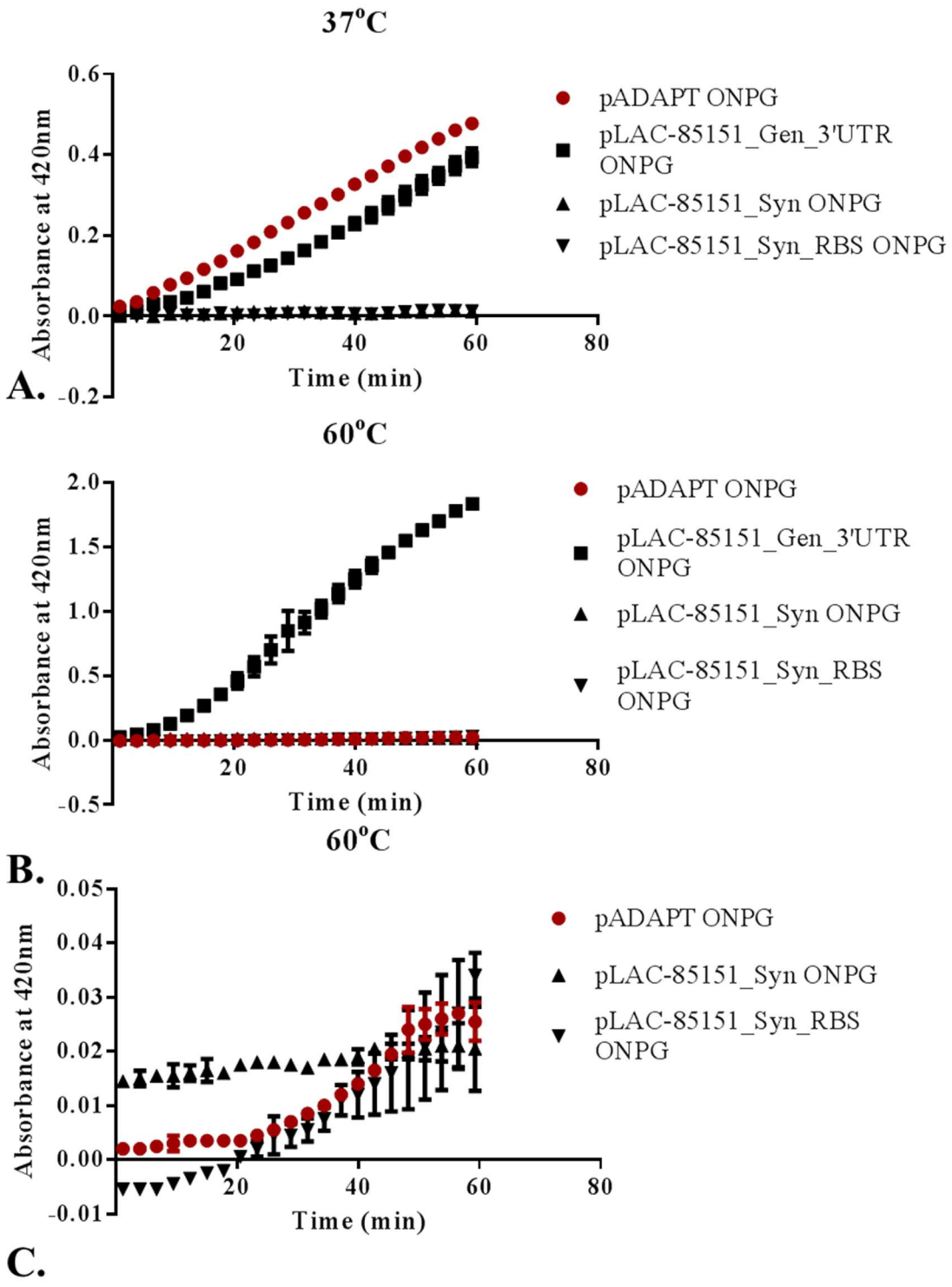


Figure 3.27 Absorbance increase during assays with Bug Buster-extracted LacZ-construct variants.

Absorbance increase during assays with Bug Buster-extracted LacZ-construct variants incubated at A. 37°C and at B. 60°C (with pre-treatment at 60°C). Panel C. is a zoom-in of low activity samples from panel B. * - Absorbance at 420nm is not normalized to

1cm pathlength. Error bars – standard deviation to the mean (N=2, technical replicates). Representative experiment shown.

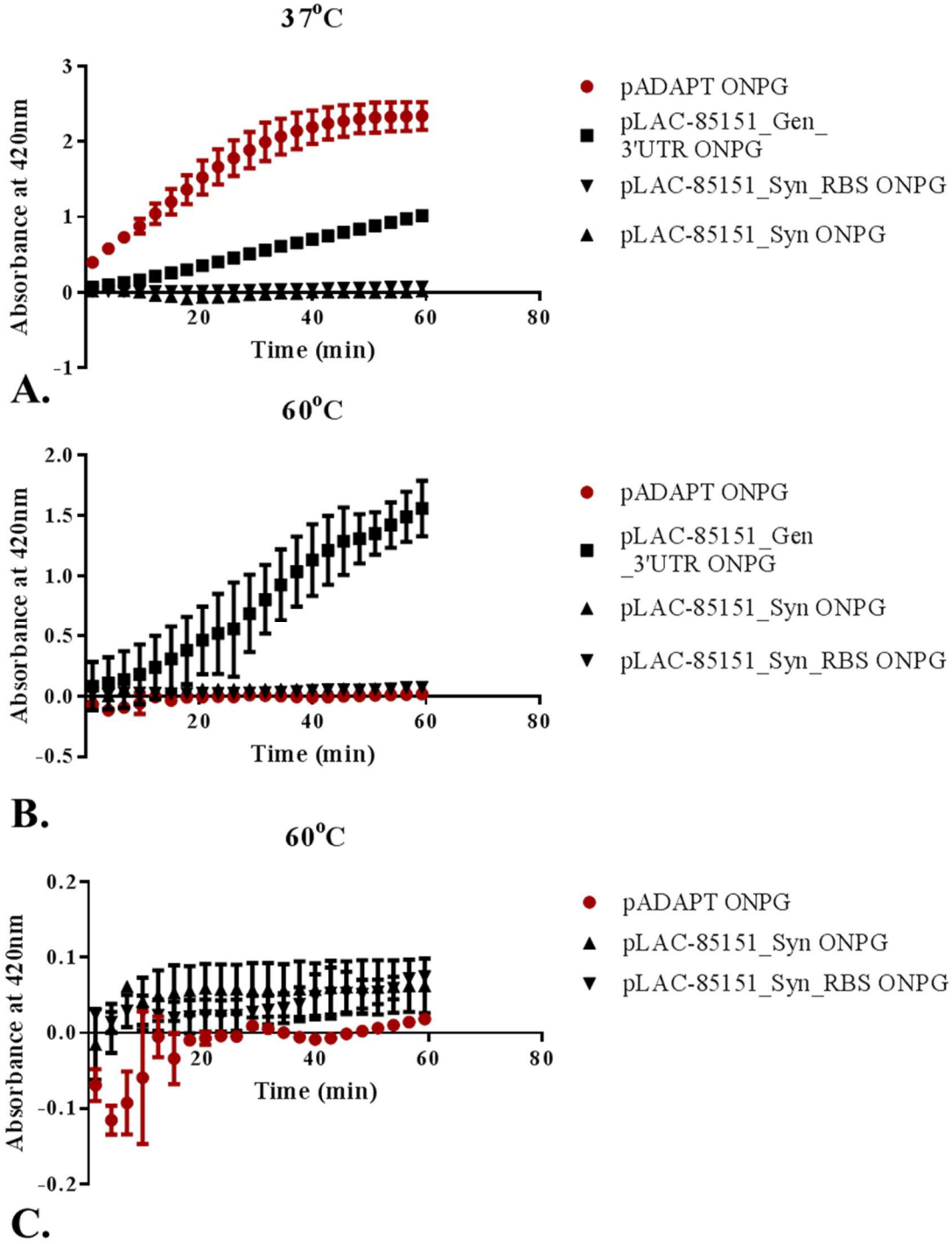


Figure 3.28 Absorbance increase over time during assays with toluene-permeabilized LacZ-construct variants

Incubation at A. 37ºC and at B. 60ºC (with pre-treatment at 60ºC). Panel C. is a zoom-in of low activity samples from panel B. * - Absorbance at 420nm is not normalized to

1cm pathlength. Error bars – standard deviation (N=2, technical replicates). Representative experiment shown.

We noticed that blanks near samples of high β-galactosidase activity in a 96-well plate were increasing in absorbance over the course of the assay (not shown). We reasoned that high volatility of o-nitrophenol was the cause, over several days on the bench o-nitrophenol would evaporate from 96 well plates while p-nitrophenol (β-glucuronidase assay product) would not. An alternative substrate for β-galactosidase – p-nitrophenyl-β-d-galactopyranoside (PNP-Gal) was chosen. According to the literature LacZ_Tts does not distinguish between PNP-Gal and ONP-Gal whereas *E. coli* LacZ has a clear preference for ONP-Gal (added benefit of reduced background). Since p-nitrophenol is released from PNP-Gal as well as from PNP-Gluc, both reactions can be monitored at the same wavelength of 405nm, increasing the throughput of measurements in the plate reader and thus decreasing time intervals. We tested activity on PNP-Gal at 0.6mM final concentration and compared it to activity on 0.6mM ONP-Gal and 1.3mM ONP-Gal. The assays were performed with aliquots of the same lysate (pLAC-85151_GEN). The results indicated that LacZ_Tts is just as active on PNP-Gal as on ONP-Gal (Figure 3.30) which was in agreement with the literature[143]. Indeed, the raw absorbance values produced by p-nitrophenol were higher than those by o-nitrophenol. For activity calculation see Methods and Materials. The assay was performed at 37ºC without heat pre-treatment.

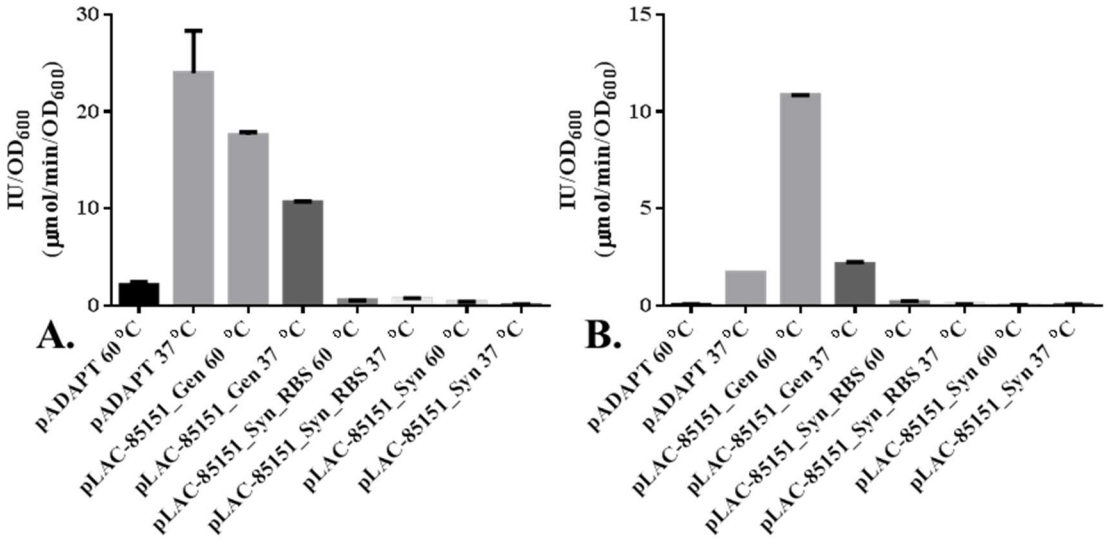


Figure 3.29 Activity of LacZ_Tts-containing cells lysates prepared by toluene permeabilization or BugBuster extraction – assayed at 60°C or at 37°C.

A. Toluene permeabilized cells. Error bars – standard deviation (N=2). **B.** BugBuster extracted lysates. Error bars – standard deviation (N=2, technical replicates).

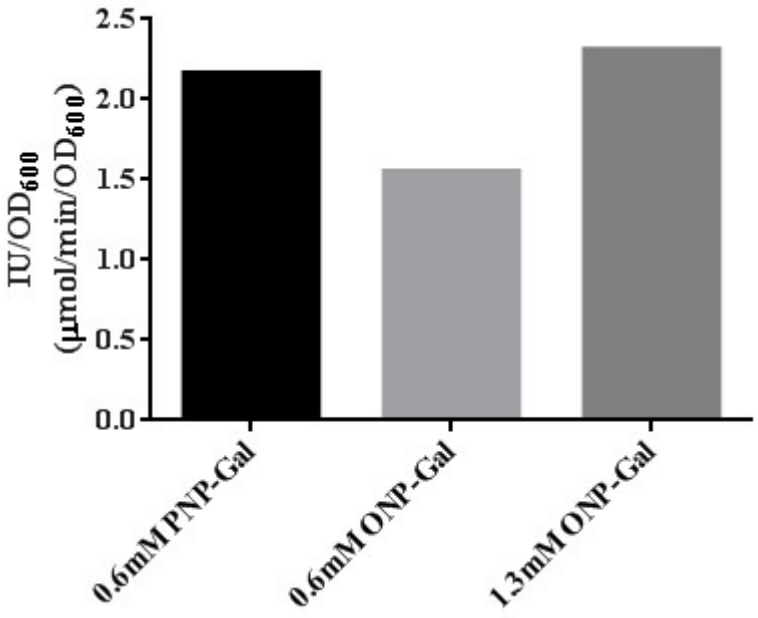


Figure 3.30 Comparison of substrates of LacZ substrates. Single experimental replicate result of enzyme activity per absorbance unit of bacterial culture with 0.6mM p-nitrophenyl-β-d-galactopyranoside and 0.6mM or 1.3mM o-nitrophenyl-β-d-galactopyranoside.

3.4.3.2 Activity assays with LacZ_Tts in B. subtilis

Having established activity in *E. coli,* the pLAC-85151_GEN plasmid was transformed into *B. subtilis* str. 168 cells to test for expression. Overnight cultures were grown and harvested by centrifugation and extracted with BugBuster. The comparison of the results of negative control pMTL85151 and plasmid pLAC-85151_GEN revealed low background activity and readily detectable β-D-galactosidase activity (Figure 3.31). The assay was performed at 37°C without heat pre-treatment.

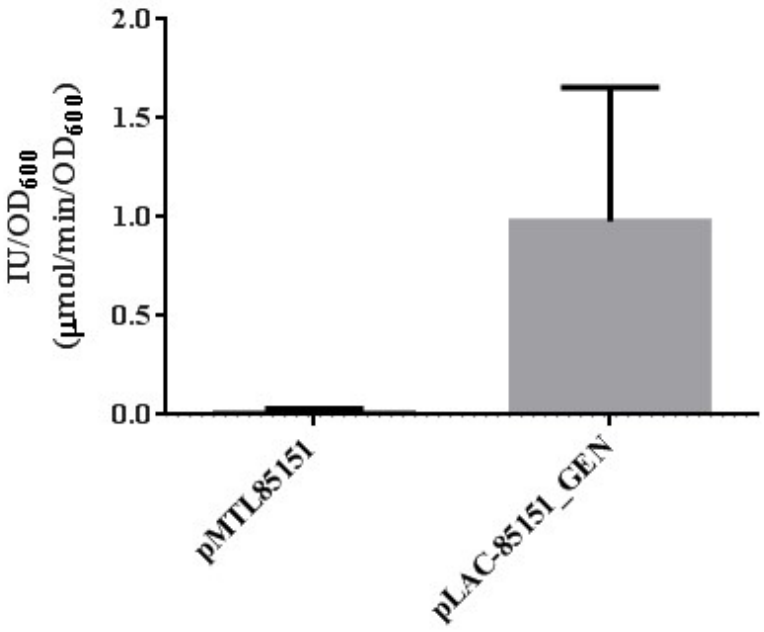


Figure 3.31 β-galactosidase activity of *B. subtilis* strains with plasmids pMTL85151 (negative control) and pLAC-85151_GEN. Error bars - standard deviation of biological replicates (N=2) and technical replicates (N=2).

3.4.3.3 Activity assays with LacZ_Tts in C. saccharoperbutylacetonicum

Next, we transformed *C. saccharoperbutylacetonicum* with plasmid pLAC-85151_GEN and negative controls pMTL85151 and pEMPTY. Triplicate cultures were grown for 12

hours from a starting OD_{600} of 0.1 and harvested. Cells were treated with Bug Buster and assayed at 60ºC. Preliminary tests activity tests at 37ºC without heat pre-treatment revealed the presence of background activity. This activity was eliminated with heat pre-treatment and subsequent incubation at 60ºC owing to the thermostability of the LacZ enzyme used (*C. saccharoperbutylacetonicum* is a mesophilic organism and its enzymes were expected to be broadly mesophilic). We were able to readily measure activity from the pLAC-85151_GEN construct and significantly decrease the background activity found in *C. saccharoperbutylacetonicum* by heat pre-treating lysates and incubating the reactions at 60ºC (Figure 3.32).

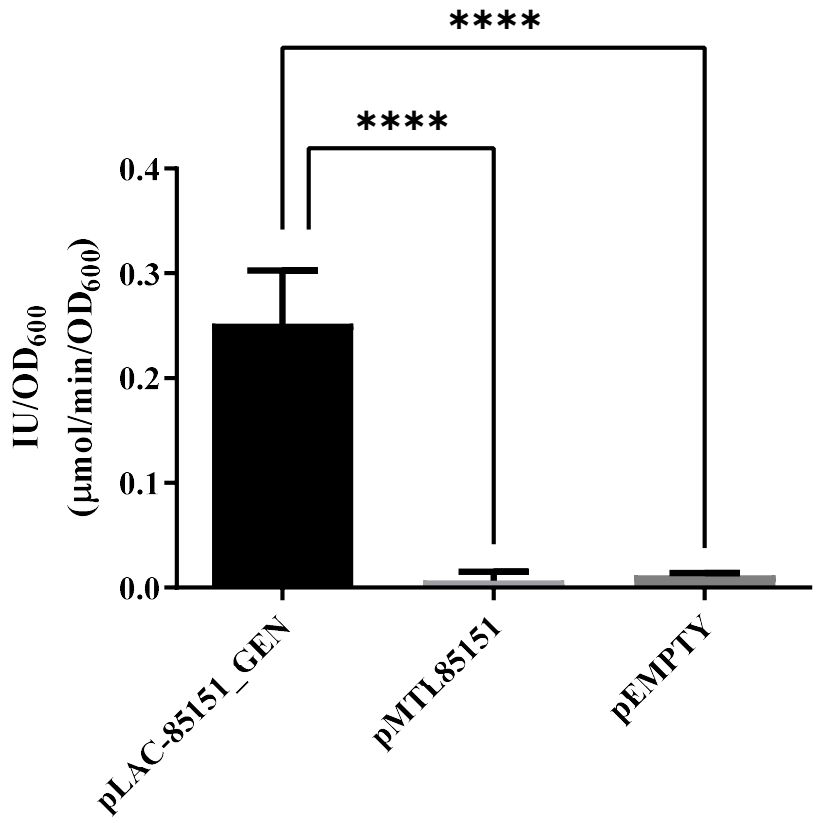


Figure 3.32 **β-galactosidase activity of *C. saccharoperbutylacetonicum* strains with plasmids pMTL85151, pEMPTY (negative controls) and pLAC-85151_GEN.** Error bars - standard deviation of the mean of 6 measurements - biological replicates (N=3) and technical replicates (N=2); **** - significant *p*-value < 0.0001 (Unpaired t-test).

3.4.4 Conclusion on the use of LacZ_Tts

After some troubleshooting of poor activity that turned out to be due to incorrect published DNA sequence, LacZ_Tts proved to be a very useful reporter of gene expression. Moreover, the construct pLAC-85151_GEN had activity in three distantly related bacterial species – model organisms *E. coli* and *B. subtilis* and the industrial solventogen *C. saccharoperbutylacetonicum*. Together with GusA, which was also active in the aforementioned species we would be able to construct a dual reporter vector series to allow for the measurement of termination efficiency *in vivo* in three species. We reasoned that comparing termination efficiency in the same sequence context across these species would begin to inform us on possible differences in terminator recognition and other species-specific effects that would improve future synthetic biology designs.

3.5 Conclusion on the choice of reporters

Overall, out of the four reporters that were extensively tested in this chapter, three proved suitable for future work. While the FbFP phiLOV2.1 anaerobic fluorescent protein was initially considered promising its low activity in both *E. coli* and *C. saccharoperbutylacetonicum* proved a major flaw. Its parent variant however (LOV2.1) was fluorescent in *E. coli* and further work with it is reported in Chapter 4. ***Figure 3.11*** summarizes the molecular designs of the reporter constructs chosen to be used as the basis for dual reporter constructs in Chapter 4.

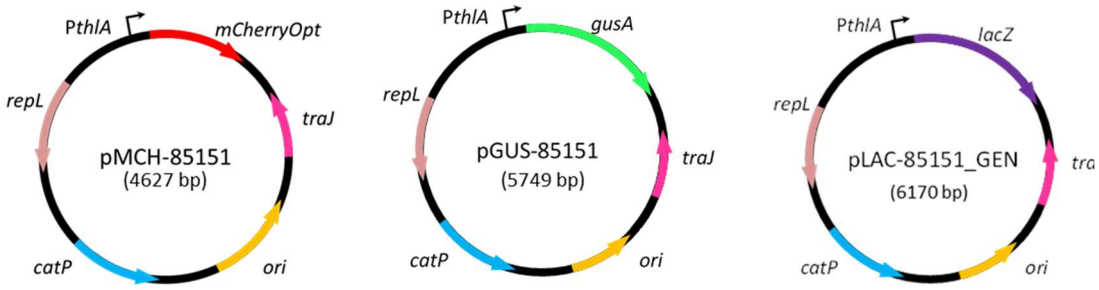


Figure 3.33 Summary of reporter designs chosen for further work.

Left to right – mCherry-containing plasmid, GusA_Ec-encoding plasmid, LacZ_Tts-encoding plasmid.

Chapter IV

Development of a novel broad-host-range dual reporter system to measure and compare termination efficiency in *Escherichia coli*, *Bacillus subtilis* and *Clostridium saccharoperbutylacetonicum*

4. Development of a novel broad-host-range dual reporter system to measure and compare termination efficiency in *E. coli*, *B. subtilis* and *C. saccharoperbutylacetonicum*

4.1 Types of reporter systems and the *in vivo* measurement of termination efficiency

The minimal reporter system for the measurement of biological activity consists of a fusion of a single reporter gene to a transcription and translation initiation element and is an established method for the *in vivo* quantification of transcription and translation initiation activity of promoters and RBSs[282]. The activity of promoters is quantified as promoter strength[283,284] (an approximation of the rate of RNA (usually mRNA for reporter systems) synthesis from a DNA or RNA (for RNA-dependent RNAPs) template driven by a promoter). Factors involved in contributing to promoter strength include but are not limited to rate of RNAP association with the promoter, rate of initiation from promoter and elongation rate[284]. Multiple sub-factors contribute to the aforementioned factors – for example, rate of initiation is influenced by promoter melting and promoter escape amongst others[285]. The activity of translation initiation elements (such as RBSs) is also expressed in terms of strength[286] (an approximation of the rate of protein synthesis from a mRNA driven by a RBS/translation initiation element). Contributing factors are known to be rate of initiation. Terminator activity is quantified and expressed as termination efficiency representing the percentage of transcripts terminated[287,288].

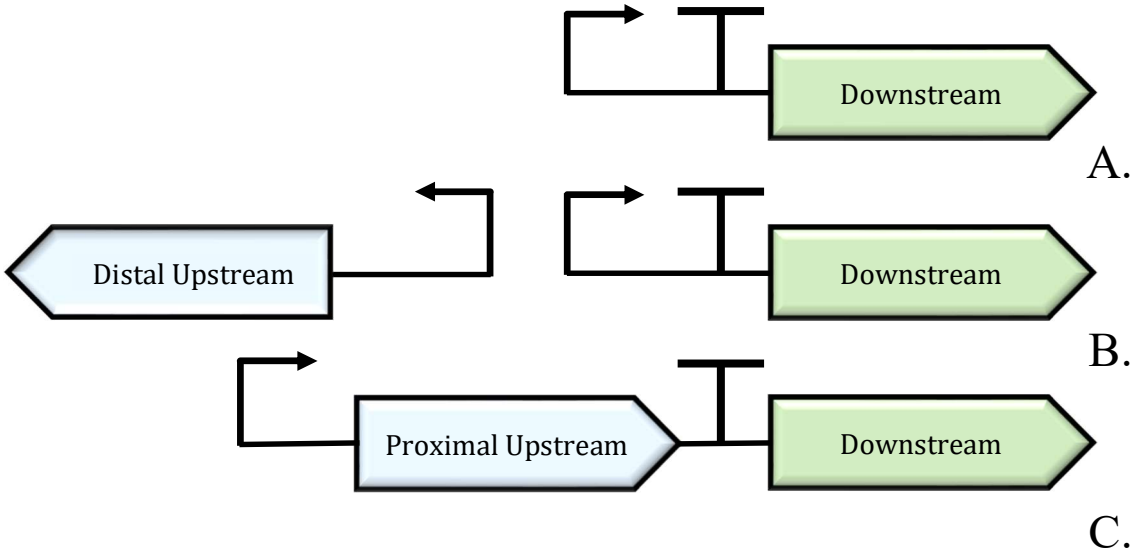
While for most applications a single reporter might be sufficient there are several instances when two or more orthogonal reporters are desirable. For instance, in the characterization of promoters a ratiometric approach can be undertaken where an

additional reporter, expressed from an independent, constitutively expressed promoter and located on the same replicon as the promoter-of-interest-reporter fusion, can serve as an internal control for plasmid copy number. This system was used to characterize the expression profile of growth rate-dependent promoters[289]. More recently, an implementation of the ratiometric promoter characteristic in conjunction with computational analysis was used to drastically reduce the variance attributable to external factors (factors other than the promoter itself such as growth medium, colony choice and replicate in addition to unidentified sources of variance (variance not attributable to any of the above)) across a variety of growth conditions in *E. coli* for a set of constitutive promoters[290].

In an additional application of dual reporter systems, researchers placed two reporters under the control of the same promoter in different but equidistant regions of the bacterial chromosome and combined with single cell measurements were able to quantify the contributions of extrinsic and intrinsic noise to cell-to-cell gene expression variation[291].

The third application of dual reporter systems and focus of this study is in the *in vivo* measurement of termination efficiency (TE)[148,292]. Termination efficiency can be expressed as a fraction of 1 or more commonly as a percent (0-100%). At present, we were able to identify three types of terminator reporter systems (**Figure 4.1**). The simplest construct (referred to here as Type I, **(Figure 4.1 A**) consists of a promoter, terminator and single reporter fusion[282,293]. This system relies solely on the decrease of reporter levels (referred to here as the downstream reporter) in comparison to a no-terminator control construct (efficiency calculation expressed in **Figure 4.1 D - Equation 4.1**). Such a system has also been used for riboswitch and anti-termination studies[294]. The advantage of this system is its simplicity, evidently it is also the only

possible system when only one reporter is available or if alternative available reporters are incompatible due to signal overlap or other reasons. The disadvantages include a lack of an internal promoter control – variability due to promoter activity can contribute to variation in the measured efficiencies. Additionally, cases where the sequence introduced influences the rate of transcriptional initiation from the promoter will not be detected. Its limitations can be overcome by RNA quantification techniques targeting the RNA immediately upstream of the terminator, however, this would increase processing time, costs and scalability of the reporter use and is sub-optimal for most applications.



$$TE = 1 - \frac{\langle DW \rangle_{smp}}{\langle DW \rangle_{ref}} \qquad \textit{Equation 4.1}$$

$$TE = 1 - \left(\frac{\langle DW \rangle_{smp}}{\langle UP_{dist} \rangle_{smp}} \right) \left(\frac{\langle DW \rangle_{ref}}{\langle UP_{dist} \rangle_{ref}} \right)^{-1} \qquad \textit{Equation 4.2}$$

$$TE = 1 - \left(\frac{\langle DW \rangle_{smp}}{\langle UP_{prox} \rangle_{smp}} \right) \left(\frac{\langle DW \rangle_{ref}}{\langle UP_{prox} \rangle_{ref}} \right)^{-1} \qquad \textit{Equation 4.3}$$

D.

Figure 4.1 Types of terminator reporter constructs and corresponding termination efficiency calculations.

A. Promoter-terminator-reporter configuration used in single reporter systems (Type I) **B**. The structure of the distal upstream reporter gene in a dual monocistronic reporter system (Type II). **C**. Synthetic bicistronic operon – dual reporter system with proximal upstream reporter (Type III). **D.** Equations for deriving termination efficiency (TE) for different reporter systems. **Eq 4.1**. TE formula for Type I system. **Eq 4.2**. TE formula for Type II system. **Eq. 4.3**. TE formula for Type III system.

Equation Legend: DW – downstream reporter levels; UP – upstream reporter levels; smp – sample (terminator-of-interest), ref –reference plasmid (no-terminator control); prox – proximal; dist – distal. Chevrons used for visual clarity. $\langle UP_{dist} \rangle_{smp}$ - distal upstream reporter level of sample; $\langle UP_{dist} \rangle_{ref}$ - distal upstream reporter level of reference; $\langle UP_{prox} \rangle_{smp}$ - proximal upstream reporter of sample; $\langle UP_{prox} \rangle_{ref}$ - proximal upstream reporter level of reference plasmid; $\langle DW \rangle_{ref}$ – downstream reporter level of reference plasmid; $\langle DW \rangle_{smp}$ - downstream reporter level of sample.

The second type (type II, **Figure 4.1 B**) of terminator reporter system uses a second reporter (referred to here as distal upstream) located on the same replicon (e.g. plasmid) that is expressed from a separate promoter[295] much in the same way of the ratiometric promoter characteristic system. The second reporter serves as an internal plasmid copy number control. The decrease in downstream reporter levels is normalized to the levels of the upstream distal reporter (**Figure 4.1 D - Equation 4.2**).

Finally, type III terminator reporter systems (**Figure 4.1 C**), which in recent years have become the most common type in synthetic biology research in the model bacterium *E. coli*, consist of a synthetic dual reporter operon that in the test samples is interrupted by a terminator. In this system, the proximal upstream reporter serves as an internal promoter control. Similar to type II, the decrease in the downstream reporter levels are normalized to the levels of the upstream proximal reporter (**Figure 4.1 D - Equation 4.3**). The application of such a system is not exclusive to terminators – the quantification of the effects of engineered intergenic elements (dubbed TIGRs for tunable intergenic regions) has also been reported[296].

In addition to termination efficiency (TE), a metric referred to as terminator strength (Ts) has also been used by researchers. Terminator strength (Ts) is simply the reciprocal of the remainder of 1 minus termination efficiency (a fraction of 1). Ts can be calculated from TE by applying ***Equation 4.4*** (**Figure 4.4**). T_S is also equivalent to the fold-change in the downstream reporter.

As RNA polymerase transcribes along the synthetic operon, it first transcribes the cistron, encoding the first reporter, and then begins transcribing the transcriptional terminator. RNAP polymerase can either continue transcribing past the transcriptional

terminator – transcriptional read-through- or terminate transcription at the terminator site – termination (**Figure 4.2**). Bacterial canonical intrinsic transcriptional terminators consist of a stable hairpin-loop structure followed immediately by a U-tract. The hairpin component of bacterial transcription terminators may be similar to Class I Pause signals[297] and termination's first step may be pausing[298]. There is evidence that the U-tract facilitates pausing by itself[242] but the hairpin may also contribute[298]. The bacterial RNAP ternary elongation complex is processive and can produce transcripts upwards of 10kb[299]. While a relationship between transcript length and the probability of full transcript synthesis hasn't been established experimentally to our knowledge, it is reasonable to assume that the probability of encountering elongational roadblocks (such as DNA-bound proteins, lesions in the DNA, RNA secondary structure, pause sites and others[300]) increases as the transcript length increases.

An additional characteristic of terminators that has been previously reported is δTE (hereby referred to as upstream effect). The upstream effect aims to quantify the influence of a terminator on the expression levels of the upstream reporter and is calculated according to Equation 4.5 (***Figure 4.3***). A known mechanism through which an intrinsic terminator could influence the upstream gene's expression levels is the effect of RNA secondary structure on RNA stability[249]. The aforementioned TIGRs are an example of a synthetic biology application of the concept[296].

As mentioned in the Introduction, canonical terminators appear to be a minority in *E. coli*[256]. Recently, it was found that the *in vivo* 3' ends of transcripts terminated by rho-dependent terminators in *E. coli* were also associated with hairpin stem-loops very similar to those of intrinsic terminators, those however had imperfect U-tracts and were followed by an enrichment for cytosine over guanosine in the downstream DNA sequence[301]. It has been demonstrated that such elements (which prior to

characterization may be considered putative intrinsic terminators) contribute little to termination but function by providing a boundary against degradation by RNAses post-transcription and that at such genomic sites termination is mediated primarily through the Rho-dependent mechanism[302]. Additionally, as mentioned in the Introduction, hairpins in the 5'UTR of transcripts have been explored in *Clostridium acetobutylicum* for improvements in gene expression[192].

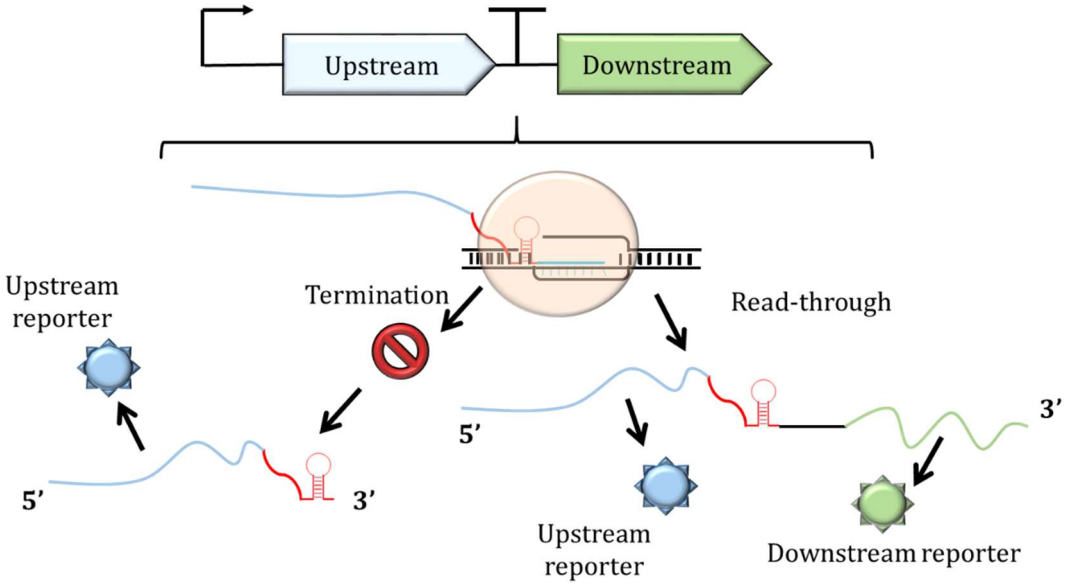


Figure 4.2 Principle of *in vivo* reporter-based dual reporter termination efficiency measurement. RNAP transcribing a synthetic operon. At a termination site, RNAP can either terminate the transcript, leading to detection of resulting of only the upstream reporter, or transcribe through the terminator (read-through) resulting in detection of both upstream and downstream reporters.

Figure 4.3 Equations for terminator strength and upstream effect calculations.

Equation 4.4 Terminator strength (T_S) **Equation 4.5** Terminator Upstream effect (δTE). Terms in equations: $\langle UP \rangle_{smp}$ - upstream reporter level of sample; $\langle UP \rangle_{ref}$ - upstream reporter level of reference plasmid; $\langle DW \rangle_{ref}$ – downstream reporter level of reference plasmid; $\langle DW \rangle_{smp}$ - downstream reporter level of sample; TE – Termination efficiency.

$$T_s = \frac{1}{1-TE} = \left(\frac{\langle UP \rangle_{smp}}{\langle DW \rangle_{smp}}\right)\left(\frac{\langle UP \rangle_{ref}}{\langle DW \rangle_{ref}}\right)^{-1} \qquad \textit{Equation 4.4}$$

$$\delta_{TE} = \frac{\langle UP \rangle_{smp}}{\langle UP \rangle_{ref}} - 1 \qquad \textit{Equation 4.5}$$

4.2 Design of broad-host-range dual reporter systems for termination efficiency measurement

When considering the design of a dual terminator reporter construct there are several important points of consideration, many of which were considered in Chapter 3 as they were also relevant when testing the single reporter test vectors for use in *C. saccharoperbutylacetonicum*. These include choice of vector backbone, promoter system, reporter, assembly-cloning strategy and "spacer" design. Hence, the design of the dual reporter constructs was based on the single reporter studies described in Chapter 3.

Two series of the dual reporter constructs were designed and assembled. The first one was constructed in an attempt to create a functional dual fluorescent reporter and was based on the pRPF185-phiLOV2.1_erm plasmid and the P*xylO/tetO* promoter system and included variants with two fluorescent reporters (phiLOV2.1 and mCherry). The second series of reporter used a more widely tested set of two enzymatic reporters (GusA_Ec and LacZ_Tts) and was based on the pGUS-85151 and pLAC-85151_GEN plasmids and the P*thlA_ac* promoter.

Both series of plasmids were designed to be compatible with the Golden Gate assembly method and in particular the CIDAR MoClo assembly standard[267]. The terminators of interest and terminator insertion site (TIS) in the reporter vector were designed to have compatible overhangs that correspond to the D and E overhangs of the MoClo standard which are commonly designated for terminators. This allowed simple sub-cloning of terminators of interest into the reporter vectors using Golden Gate assembly. The TIS consists of outwardly facing Type IIS restriction sites that when cut are excised from the plasmid and leave compatible overhangs for an incoming nucleic acid to be ligated

to. Two variants of TIS were designed – an R-variant (R for replacement) with BsaI restriction sites and B-variant (B for BsmBI) with BsmBI (isoschizomer Esp3I was used instead of BsmBI in routine Golden Gate manipulations) restriction sites. After assembly with a terminator, the end products constructed with either variant are intended to be identical.

The design of the intergenic, untranslated region in the synthetic operon can likely influence the measured termination efficiency of a terminator in several ways. Firstly, as mentioned in the Introduction, it has been reported that a terminator's distance to the stop codon of a preceding coding sequence can influence termination efficiency. Reportedly, if a terminating ribosome's footprint overlaps with the terminator hairpin, there can be a decrease in termination efficiency at intrinsic termination sites. The proposed mechanism is by the ribosome preventing folding of the hairpin structure[153]. To address this, a spacer sequence of 27bp was designed for the first series of reporter constructs (**Figure 4.4**). In the second series, the length of the spacer was either kept the same or increased to 32bp for vectors based on the pLAC-85151_GEN plasmid. In the first series the TIS and 5'UTR of the downstream reporter follow the spacer. In the second series of constructs an insulator sequence was incorporated downstream of the spacer. This sequence was based on the literature[153] with some modifications and was intended to be as devoid of regulatory elements or secondary RNA structure as possible. The intended role of the insulator was to prevent a terminator's strong secondary structure from influencing the translation initiation region of the downstream reporter[303], which is relevant in the case of transcriptional read-through through the terminator. Thus, the insulator's purpose is to minimize the occurrence of false positives (low levels of downstream reporter) due to translational inhibition. The insulator is followed by the downstream reporter's 5'UTR which included the RBS.

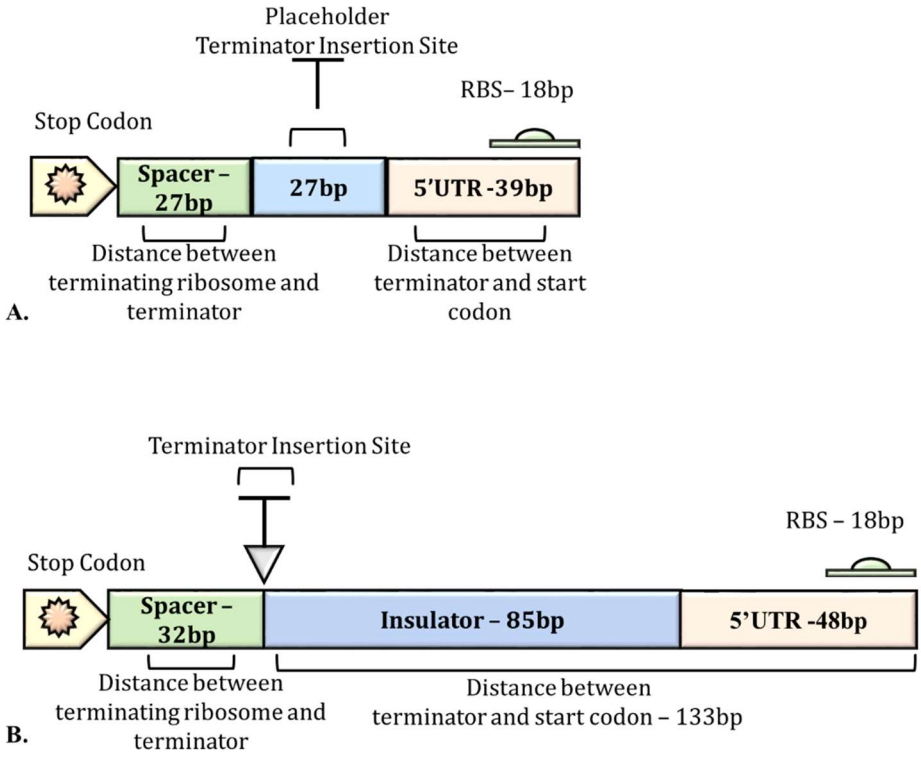


Figure 4.4 The structure surrounding the terminator insertion site (TIS) of the dual reporter vector.

A. First iteration of the dual reporter vector (TIS v1). B. Second iteration of the dual reporter vector (TIS v2). B-variant TIS v2 contain BsmBI restriction sites, R-variant contain BsaI restriction sites, X-variant do not contain restriction sites.

4.3 Construction of dual reporter systems for termination efficiency measurement

Initially, in this study two parallel sets of reporter vectors were constructed, the pTREF series of dual fluorescent vectors and the dual enzymatic (GusA and LacZ_Tts) series, the construction of which is briefly described below.

4.3.1 Construction of dual fluorescent and hybrid reporter systems (pTREF series)

As mentioned above the first series of terminator reporter vectors were based on the phiLOV2.1-pRPF185_erm with some modifications. The cloning steps taken to construct the first series of dual reporter constructs are illustrated in **Figure 4.5**. Briefly, plasmid phiLOV2.1-pRPF185 was amplified using primers IG0011 and IG0012. The downstream reporter genes *mCherryOpt_SDM*, *mCherry_BBa* and *catP* were amplified using primer pairs IG0162 and IG0163, IG0019 and IG0020, and IG0017 and IG0018 from plasmids pDSW1728_SDM, pSB1C3_ J06504 and pMTL85151, respectively. The *traJ* gene was amplified with primer pair IG0023 and IG0024 from plasmid pDSW1728. The *ermB* gene was amplified using primers IG0037 and IG0038 from plasmidpMTL832351. All PCR products were gel extracted and used in a Golden Gate assembly reaction with BpiI restriction enzyme to generate plasmids pTREF1, pTREF2 and pTRE3 with *catP*, *mCherry_BBa* and *mCherryOpt_SDM* as downstream reporter genes, respectively.

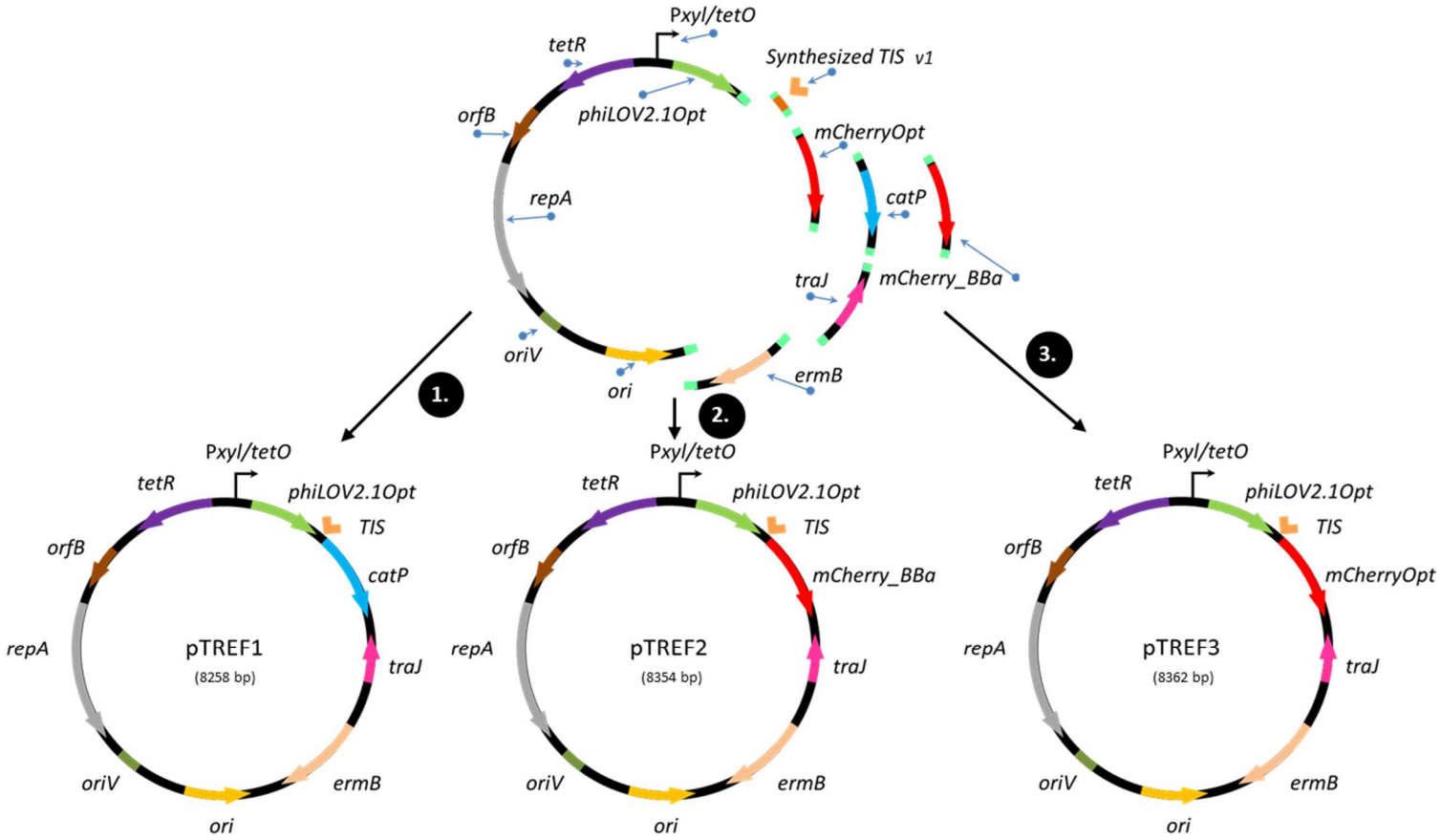


Figure 4.5 Schematic diagram of the dual fluorescent reporter (and one fluorescent-enzymatic hybrid) constructs and their assembly. The following PCR-amplified fragments were combined using Golden Gate Assembly (GGA): 1. Backbone (with *phiLOV2.1Opt*), synthesized TIS v1, *catP*, *traJ*, *ermB* form pTREF1. 2. Backbone (with *phiLOV2.1Opt*), synthesized TIS v1, *mCherry_BBa*, *traJ*, *ermB* to form pTREF2. 3. Backbone (with *phiLOV2.1Opt*), synthesized TIS v1, *mCherryOpt*, *traJ*, *ermB* to form pTREF3.

The plasmids were sequence verified using primers IG0119 and IG0186.

The pTREF series were designed with the following purposes in mind – pTREF2 was an *E. coli* control (because of the higher GC% gene of the *mCherry_BBa* and its *E. coli* RBS BBa_B0034). The pTREF3 plasmid was designed to be optimal for *Clostridium* species with its two codon-optimized genes. Finally, pTREF1 was a hybrid reporter vector (*phiLOV2.1Opt* and *catP*) designed as an alternative in case the fluorescent reporters failed.

4.3.2 Construction of dual enzymatic reporter systems (GusA and LacZ_Tts)

The second series of reporter constructs consisted of two combinations of reporter order – GusA upstream and LacZ_Tts downstream and *vice versa*, referred to as GUS-LAC and LAC-GUS for brevity.

Briefly, the GUS-LAC varants were constructed based on the aforementioned single reporter vector pGUS-85151. Plasmid pGUS-85151 was amplified with primers IG0126 and IG0208 and the 5727 bp fragment was gel extracted. The *lacZ_syn* gene was supplied on plasmid CLOLACZ_SYN_CD. The intergenic region, containing a B-variant TIS v2 (stands for BsmBI restriction sites), was synthesized as complementary oligonucleotides IG0121, IG0123, IG0185 and IG0187which were used in an overlap extension PCR together (see Materials and Methods). A “R” variant TIS v2 (R stands for replacement, contains BsaI restrictions sites instead of BsmBI) was also synthesized using oligonucleotides IG0120, IG0122, IG0185 and IG0187. The vector, second reporter and TIS fragments were used in a Golden Gate assembly reaction with the BpiI restriction enzyme to construct plasmids pGUS-B-LAC_Syn (**Figure 4.6**) and pGUS-R-LAC_Syn (not shown). The plasmids were verified by Sanger sequencing.

For LAC-GUS variants, plasmid pLAC-85151_Syn_RBS was amplified using primers IG0152 and IG0208 and gel extracted. The *gusA* gene was amplified using primer pair IG0124 and IG0125 and gel extracted. The vector, reporter and TIS fragments were assembled by Golden Gate to construct plasmids pLAC-B-GUS_Syn (**Figure 4.6**) and pLAC-R-GUS_Syn (not shown). The plasmids were verified by Sanger sequencing.

Next, the pGUS-B-LAC_Syn and pLAC-B-GUS_Syn plasmids were digested with BsmBI, treated with T4 PNK and self-ligated to yield plasmids pGUS-X-LAC_Syn and pLAC-X-GUS_Syn (**Figure 4.6**). The plasmids were verified by Sanger sequencing. The "X" variants were deemed the most suitable negative controls as they lacked the TIS region which contains the BsmBI restriction sites which form an interrupted inverted repeat, in other words an interrupted palindrome (the restriction sites of BsmBI as is typical of Type IIS restriction enzymes are not themselves palindromic).

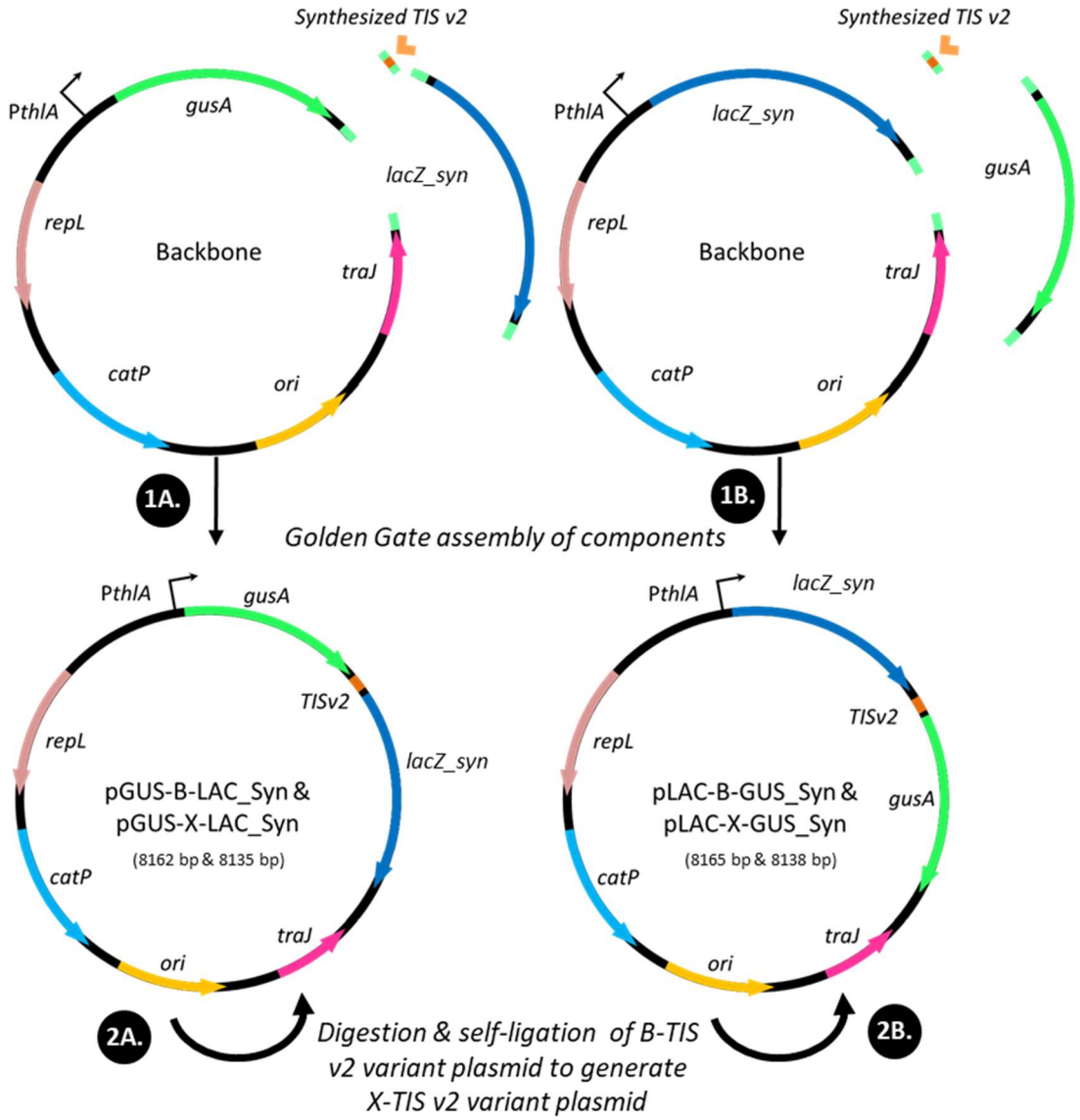


Figure 4.6 Schematic diagram of the first iteration of the dual enzymatic reporter constructs and their assembly.

1A. Backbone fragment (derived from pGUS-85151), B-TIS v2 fragment, *lacZ_syn* fragment were PCR amplified and combined in GGA to generate pGUS-B-LAC_Syn. **1B.** Backbone fragment (derived from pLAC-85151_Syn), B-TIS v2 fragment, *gusA* fragment were PCR amplified and combined in GGA to generate pLAC-B-GUS_Syn. **2A and 2B.** Plasmids were restricted pGUS-B-LAC_Syn and pLAC-B-GUS_Syn with Esp3I, blunting and self-ligation to generate pGUS-X-LAC_Syn and pLAC-X-GUS_Syn, respectively.

After extensive testing (reported in Chapter 3) on the *lacZ_syn* variant, we concluded that it is largely inactive and therefore it had to be replaced in the dual enzymatic

reporter constructs with the genome-derived copy *lacZ_gen.* Briefly, plasmid pMCH-85151 (chosen to serve as backbone because uncut plasmid background produces colonies of red color after transformation in *E. coli*, the backbone is identical to pGUS-85151) was digested with restriction enzymes SacI and BamHI and the 3892bp fragment was gel extracted. The *gusA* gene together with the downstream intergenic region was PCR amplified from pGUS-B-LAC_Syn and pGUS-X-LAC_Syn using primer pair IG0119 and IG0187. The fragments were digested with SacI and BpiI and gel extracted. The *lacZ_gen* gene was amplified from pLAC-85151_GEN_3'UTR with primers IG0086 and IG0097, digested with BamHI and BpiI and gel extracted. The gel extracted fragments were ligated to form plasmids pGUS-B-LAC_GEN and pGUS-X-LAC_GEN. The plasmids were then sequence verified.

For the LAC-GUS series, plasmid pLAC-85151_GEN was digested with BamHI, dephosphorylated and gel extracted. The *gusA* gene together with the upstream intergenic region was PCR amplified using primers IG0081 and IG0186 from plasmids pLAC-B-GUS_Syn and pLAC-X-GUS_Syn. The fragments were digested with BamHI, gel extracted and separately ligated with the cut pLAC-85151_GEN to form plasmids pLAC-B-GUS_GEN and pLAC-X-GUS_GEN, respectively. The plasmids were then sequence verified.

Figure 4.7 Schematic diagram of the second iteration of the dual enzymatic reporter constructs and their assembly.

1A. Backbone fragment (derived from pMCH-85151), fragment derived from pGUS-B-LAC_Syn (containing *gusA* and B-TIS v2 fragment), PCR fragment from pLAC-85151_GEN_3'UTR (containing *lacZ*) were combined in GGA to generate pGUS-B-LAC_GEN. **1B.** Backbone fragment (derived from pMCH-85151), fragment derived from pGUS-X-LAC_Syn (containing *gusA* and X-TIS v2 fragment), PCR fragment from pLAC-85151_GEN_3'UTR (containing *lacZ*) were combined in GGA to generate pGUS-X-LAC_GEN. **1C.** Backbone fragment (derived from pLAC-85151_GEN) and a PCR fragment (containing *gusA* and B-TIS v2) derived from pLAC-B-GUS_Syn were combined in GGA to generate pLAC-B-GUS_GEN. **1D.** Backbone fragment (derived from pLAC-85151_GEN) and a PCR fragment (containing *gusA* and X-TIS v2) derived from pLAC-X-GUS_Syn were combined in GGA to generate pLAC-X-GUS_GEN.

4.4 Testing of the dual fluorescent reporter system

4.4.1 Activity measurement of dual fluorescent reporter system in *E. coli*

The *E. coli* CA434 conjugation donor strain was transformed with plasmids pREF0, pTREF1, pTREF2 and pTREF3. The above strains, including wild-type CA434 were inoculated into defined C minimal medium (4% (w/v) glycerol, 3% (w/v) D-glucose) (see Methods and Materials). The cultures were grown at 37ºC overnight (22-23 hours), OD_{600} was measured, aliquots were taken and fixed, and doxycycline (to 400ng/mL) was added to the culture. Induction proceeded for 13.5 hours when induced samples were harvested and fixed. All fixed samples were washed three times in PBS buffer and then resuspended in PBS buffer. The samples were then kept in the dark until flow cytometry was performed. The results of the flow cytometry experiment showed proportionally higher increases in green fluorescence emission (upstream reporter) (λ_{ex}=405nm, λ_{em}=500-550nm) in the test samples (pTREF1 and pTREF2) over the empty vector control (pREF0) (**Figure 4.8**). Sample pTREF3 had a very modest increase in green fluorescence post-induction that was not different from the induced control pREF0 (*p* value = 0.3511, Unpaired t-test) (**Figure 4.8**).

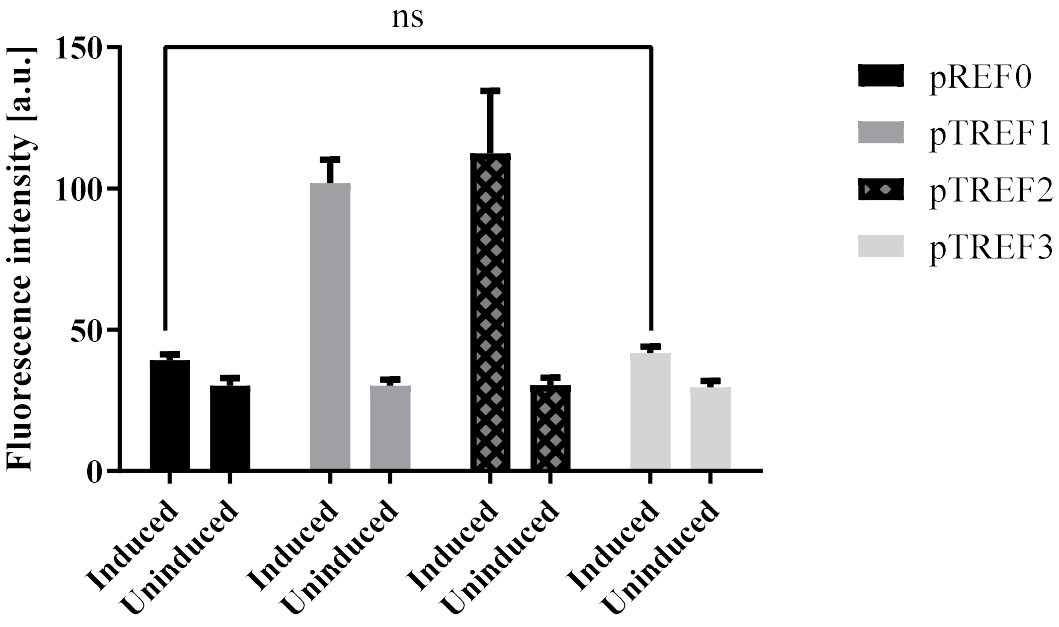


Figure 4.8 Fluorescence intensity of upstream reporter in dual fluorescent reporter constructs. Error bars – standard deviation (SD) of the geometric mean (geometric mean of fluorescence) from two biological replicates (N=2), ns – no significant difference found – *p*-value = 0.3511 (Unpaired t-test). Left column – induced; Right column – uninduced. . λ_{ex}=405nm, λ_{em}=500-550nm.

The red fluorescence emission profile of the uninduced pTREF2 sample showed that there was significant "leaky" expression of the mCherry protein (**Figure 4.9 A**). This appeared to be due to spurious promoter(s) within the *phiLOV2.1Opt* gene which as mentioned previously in Chapter 3 has a very low GC-content as the single reporter mCherry construct was tightly repressed (for these results see Chapter 3) in the absence of inducer. The spurious transcription initiation from *phiLOV2.1Opt* appears to be constitutive. When induced, however, pTREF2 does display an increase in red fluorescence – the geometric mean of fluorescence increases an average of 2.38-fold (**Table 1**). However, there is still an identifiable sub-population that is fluorescent to the level of the uninduced constitutively expressed sample (**Figure 4.9 B** and **E, Gate H2**). The proportion of cells in H1 (highly fluorescent) and H2 (moderately fluorescent) is inverse in the induced versus the uninduced sample (**Table 4.1**) while the proportion of

non-fluorescent cells remains stable (Number of cells from Gates “All” minus H1 and H2, **Table 4.1**).

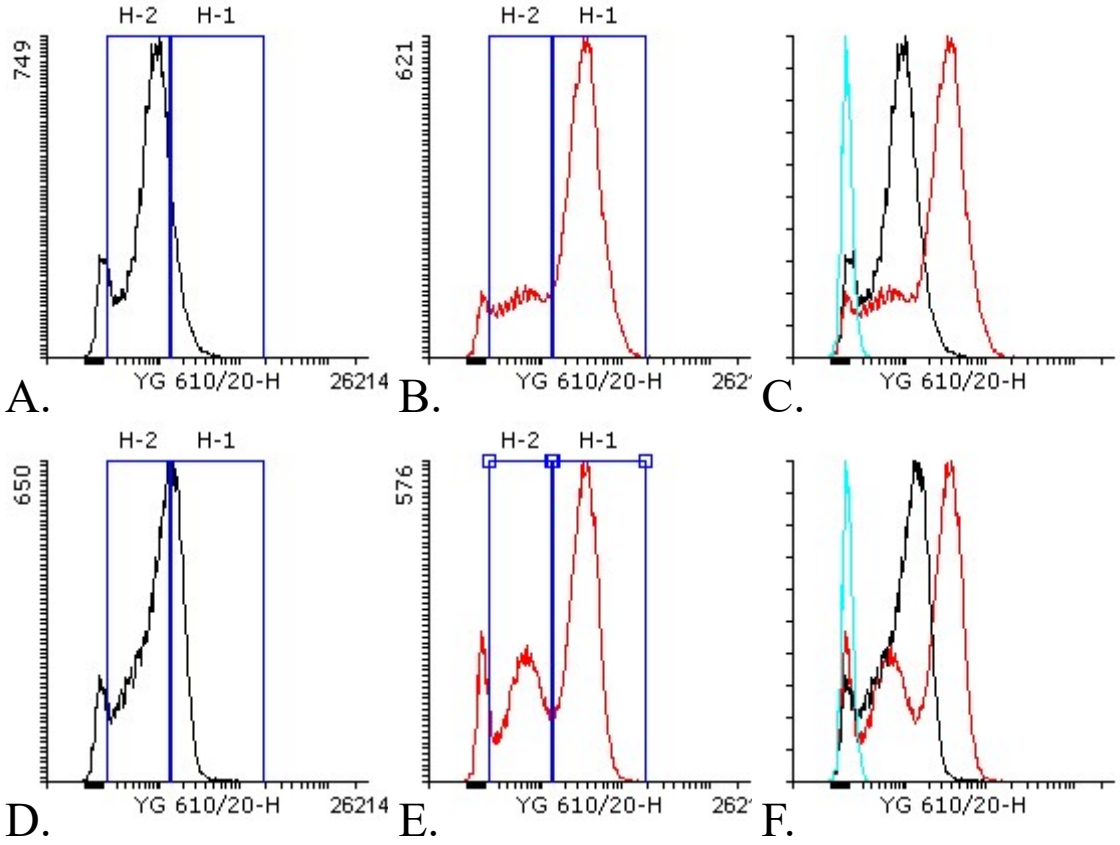


Figure 4.9 Measurement of red fluorescence from pTREF2 after 13 hour induction.

Black line – uninduced pTREF2, red line – induced pTREF2, cyan in overlay – pREF0. Gate H1 encompasses the highly fluorescent population in the induced sample. Gate H2 encompasses the population with lower fluorescence. Panels A,B and C are biological replicates of D, E and F.

The results with sample pTREF3 showed that red fluorescence was marginally increased over the empty vector pREF0 (**Figure 4.10** and **Table 4.1**) but remained largely unaffected by induction and overall was significantly lower than that observed in pTREF2. This indicated problems with the expresssion of *mCherryOpt_SDM* from pTREF3 and also confirmed the spurious promoter activity of the *phiLOV2.1Opt* gene.

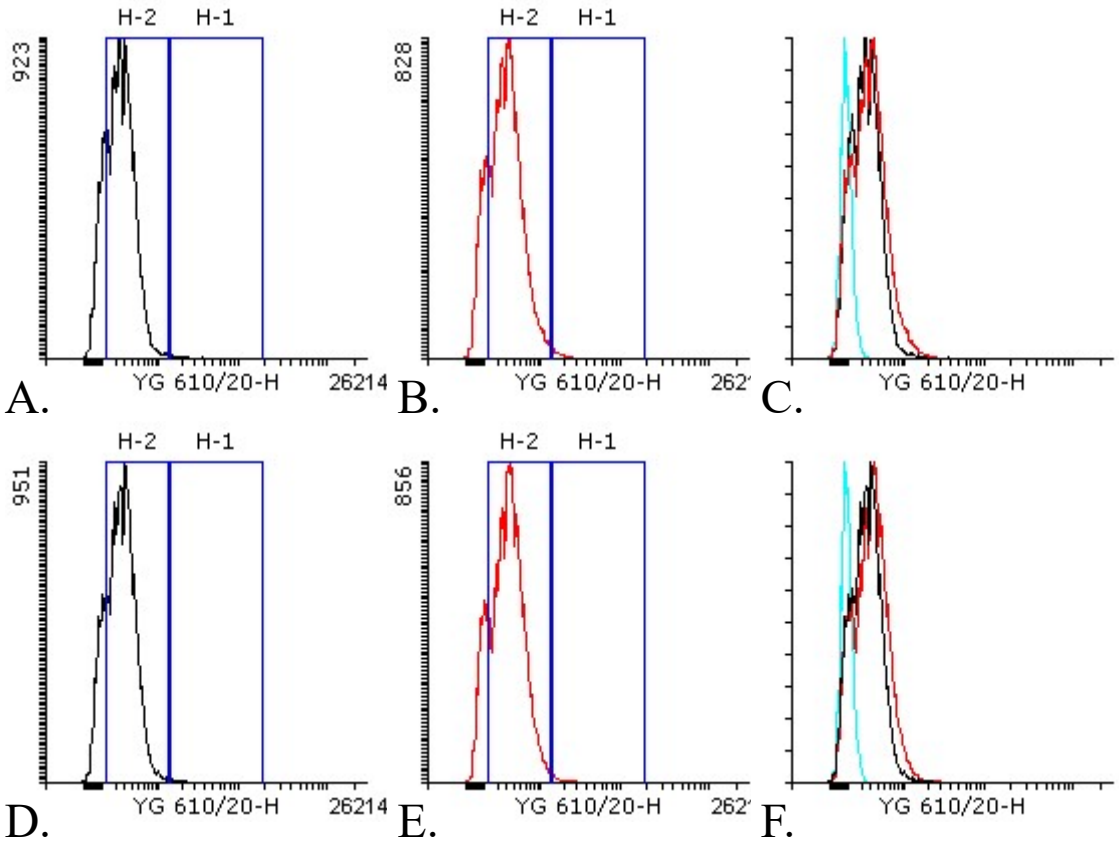


Figure 4.10 Increases in red fluorescence from pTREF3 after 13 hour induction.

Black line – uninduced pTREF3, red line – induced pTREF3, cyan in overlay – pREF0. Gate H1 encompasses the highly fluorescent population in the induced sample and corresponds to Gate H1 in **Figure 4.9**. Gate H2 encompasses the population with lower fluorescence and corresponds to Gate H2 in **Figure 4.9**. Panels A,B and C are biological replicates of D, E and F.

To troubleshoot this, *E. coli* CA434 with pTREF3 was grown in the aforementioned defined C minimal medium at 37ºC for 36 hours and strains were inoculated in duplicate – one replicate was induced at onset of growth with 400ng/mL anhydrotetracycline (ATc) while the second culture was not. The result showed that the uninduced sample had a higher geometric mean of fluorescence after 36 hours than overnight (**Figure 4.11 A** and **B**). Surprisingly, the sample which was induced from the onset of growth had a small fraction of highly fluorescent cells, while the majority of the cells were non-fluorescent, indicating loss of fluorescent gene expression or loss of

the plasmid (the conditions were selective). Plasmid pTREF2 showed similar red fluorescence profile to the previous experiments (**Figure 4.11 C** and **D**).

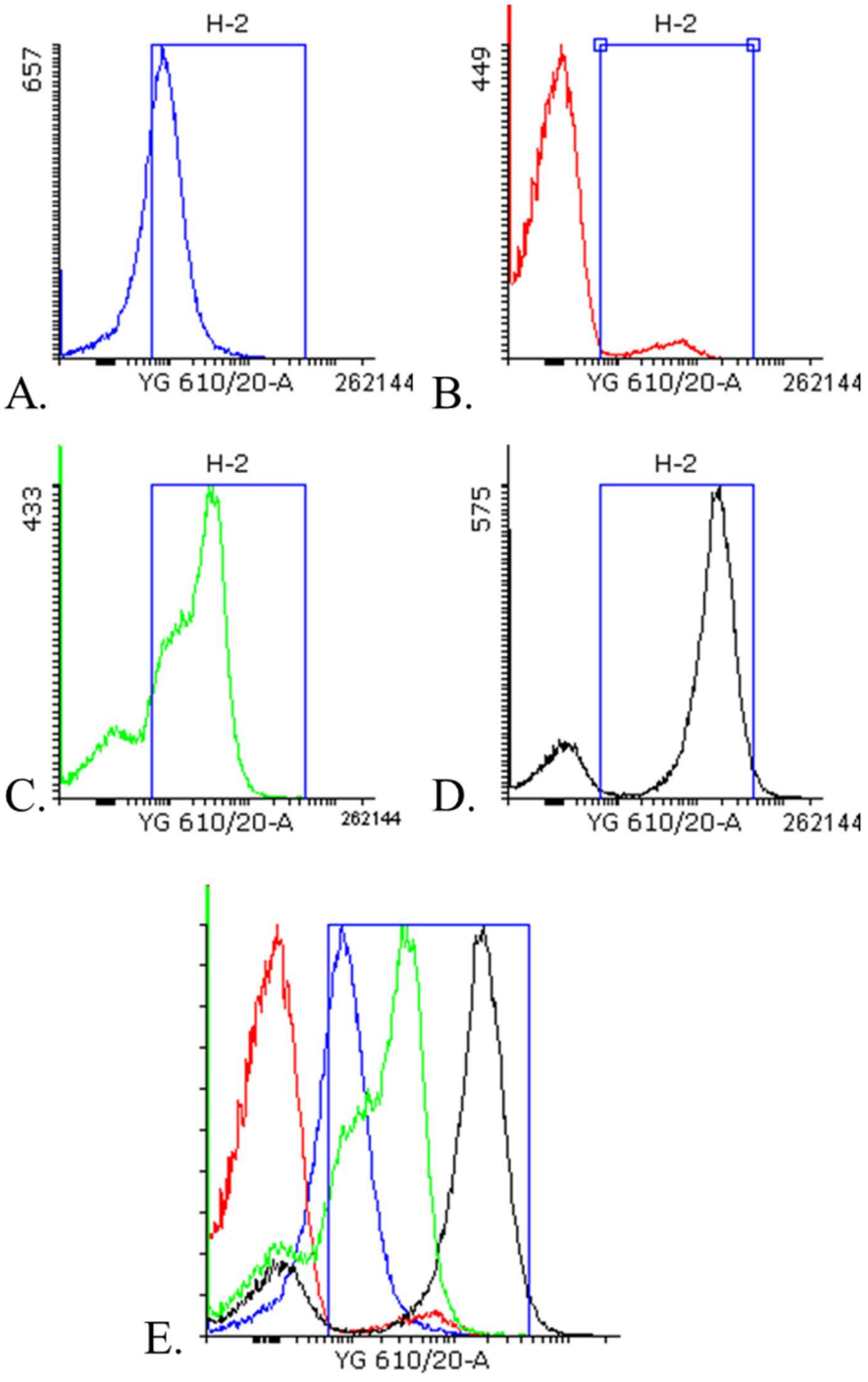


Figure 4.11 Increases in red fluorescence from pTREF2 and pTREF3 with induction at the onset of growth.

Event count (y-axis) versus red fluorescence on x-axis (YG 610/20 – A). A. Uninduced pTREF3. B. Induced pTREF3. C. Uninduced pTREF2. D. Induced pTREF2.

Additionally, an initial attempt at transformation of strain CA434 with the pTREF3 construct, when selected on Tet 10µg/ml produced only 5 colonies, whereas pTREF1 and pTREF2 produced hundreds up to thousands (with similar starting concentrations of plasmid DNA). Out of the 5 colonies, 3 were tested by PCR with primers IG0119 and IG0186 (encompassing the dual reporter cassette) and showed increased size. When two of these plasmids were sequenced, they both were found to contain insertion sequence (IS) elements. Each colony had a unique IS element insertion and insertion site. The insertions either interrupted the *mCherryOpt* gene or were just upstream of it – in the intergenic region **(Figure 4.12**).

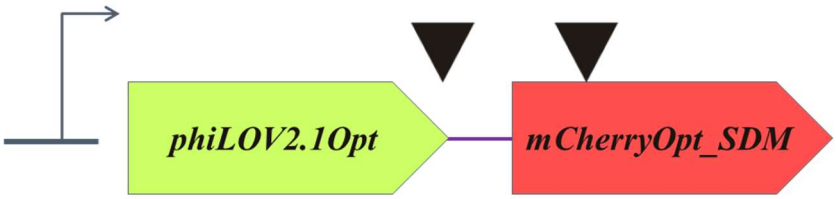


Figure 4.12 Schematic representation of the insertion sites in pTREF3 found during transfer into CA434.

The upstream insertion was determined to be IS2, with a duplicated recognition sequence of 5' TAAGG 3'. The downstream insertion was determined to be IS10 with a duplicated recognition sequence of 5' TCTAAAGCA3 '.

E. coli CA434 is Tet^R due to the conjugative plasmid pR702 and tetracycline is routinely used to maintain that plasmid but is also an inducer of the P*xylO*/*tetO* promoter. We were able to successfully isolate *E. coli* CA434 transformed with non-mutagenized pTREF3 by selecting on kanamycin instead of tetracycline (CA434 is also Kan^R).

Based on these results, we hypothesized that the simultaneous induction of the synthetic *phiLOV2.1Opt*::*mCherry_Opt_SDM* operon was detrimental in an unidentified way to the *E. coli* host cell.

Table 4.1 Statistics of red fluorescence measurements by flow cytometry. Data displayed is rounded to 5 significant figures. CV – Coefficient of variation, GM – Geometric mean, % VIS - % of Visible, IND – inducer (dox). Values are averages of two biological replicates and standard deviation is expressed following the ± sign.

Strain	IND	Gate	Events	% VIS	Mean	GM	Median	CV
pREF0	-	All	50000	100	84.37±1.07	75.60±3.84	78.5±3.54	453.23±580.04
	+	All	50000	100	85.12±2.91	78.22±2.18	81±2.12	40.14±137.93
pTREF 2	-	All	50000	100	983.34±152.37	702.18±84.83	911.50±166.17	73.61±1.15
		H1	11043±5919.9	22.09±11.84	2002.9±21.30	1922.4±9.94	1804.5±28.99	37.06±0.35
		H2	33605±6710.4	67.21±13.42	751.97±25.18	651.97±41.84	755.50±20.51	46.07±4.55
	+	All	50000	100	2780.6±472.82	1674.4±441.79	2656±401.64	77.38±4.67
		H1	33373±4838.0	66.75±9.68	3892.4±200.33	3572.30±112.67	3588.5±48.79	43.37±6.51
		H2	12908±3261.2	25.82±6.52	594.21±5.41	499.33±18.28	542±14.14	54.69±4.40
pTREF 3	-	All	50000	100	276.36±8.34	220.05±6.02	240±8.49	81.22±1.02
		H1	194±21.21	0.39±0.04	2344.2±57.03	2160.4±33.45	1930±29.70	48.47±2.40
		H2	39844±438.41	79.688±0.88	314.10±7.31	282.60±6.65	282±9.90	49.67±0.35
	+	All	50000	100	338.76±14.41	259.60±14.44	286.5±20.51	86.15±4.83
		H1	338±56.57	0.676±0.11	2028.17±99.88	1912.3±68.05	1713±24.04	43.02±5.96
		H2	40815±694.38	81.63±1.39	377.55±14.13	330.83±14.12	329±16.97	54.83±2.19

4.4.2 Using the pTREF2 plasmid to determine TE of a small set of terminators

To assess the potential of the pTREF series further, the pTREF2 plasmid was used to measure the TE of a small set of known transcriptional terminators that had been previously studied in *E. coli*[148]. The terminators chosen were 3 strong synthetic terminators and 1 strong double synthetic terminator. Whilst reportedly very strong TE (over 90%), the 3 synthetic terminators were still distinguishable in the original publication and as such are a suitable test for assay sensitivity.

The cloning of terminators is described in Materials and Methods in greater detail. Briefly, the four well characterized terminators (BBa_B0015, L3S2P00m, L3S2P36m and L3S2P21m) were sub-cloned into pTREF2 using Golden Gate assembly and sequence verified (see Methods and Materials). The plasmids were transformed into *E. coli* CA434 and then the transformed strains (including no-terminator pTREF2) were inoculated in C medium, grown at 37ºC overnight (22-23 hours) and induced with doxycycline (to 400ng/mL) for 13.5 hours. Samples of induced and uninduced strains were harvested and fixed as described in Materials and Methods. The level of red and green fluorescence of each sample was then analyzed using flow cytometry (as described previously).

The TE and T_S were then computed from these fluorescence levels using ***Equation 4.3*** (**Figure 4.1**) and ***Equation 4.4*** (**Figure 4.4**). The results of these measurements were compared to the published measurements from *E. coli* DH5α[148] (**Figure 4.13**). There was general agreement with regards to the order of strength with the published data. Notably, terminator L3SP21 was remarkably reproducible between our measurements and the literature.

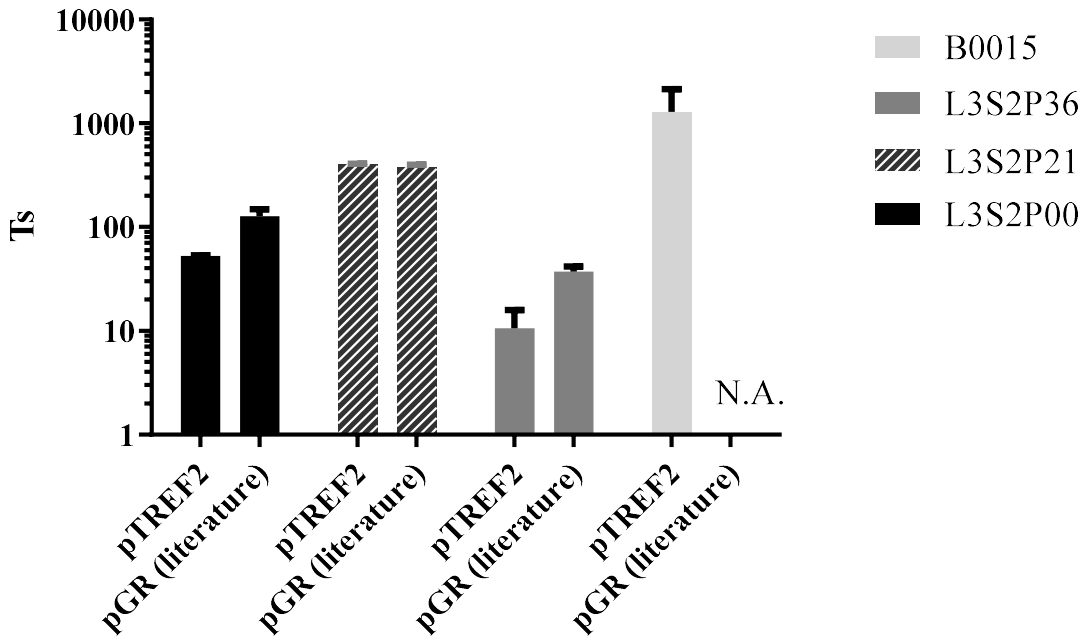


Figure 4.13 **Plot of terminator strengths (T_S) measured with pTREF2.** Error bars – standard deviation (SD) of the geometric mean (geometric mean of fluorescence) from two biological replicates (N=2) for pTREF2 and three for pGR (N=3). Left column – measured in pTREF2 (this study); Right column – data measured in plasmid pGR from Chen *et al.* 2013 study[148]. N.A – data not available.

Overall, the measurements were reproducible, with small standard deviation of the mean (**Table 4.2**). We quantified the TE according to equation 4.1 (**Figure 4.1**) and δTE according to equation 4.5 **(Figure 4.4**) and found those metrics to be more variable (**Table 4.2**).

Table 4.2 Statistics of terminator strength and termination efficiencies derived from pTREF2. Data displayed is rounded to 5 significant figures. N.A. – data not available.

Measurement	L3S2P00	L3S2P21	L3S2P36	B0015
Measured T_S (Eq. 4.4)	52.54 ±1.01	406.38 ±1.22	10.55 ±0.42	1296.61 ±521.04

Measured TE (Eq. 4.3)	0.98096 $\pm 3.7 \times 10^{-4}$	0.99754 $\pm 7.1 \times 10^{-6}$	0.90516 $\pm 3.8 \times 10^{-3}$	0.99916 $\pm 3.39 \times 10^{-4}$
Measured TE (Eq. 4.1)	0.98368 $\pm 1.54 \times 10^{-3}$	0.99756 $\pm 4.243 \times 10^{-4}$	0.94095 $\pm 3.36 \times 10^{-2}$	0.99916 $\pm 2.19 \times 10^{-4}$
δTE (Eq. 4.5)	-0.14336 $\pm 6.41 \times 10^{-2}$	-0.0079 $\pm 1.74 \times 10^{-1}$	-0.36975 $\pm 0.38 \times 10^{-1}$	0.04275 $\pm 0.16 \times 10^{-1}$
Published T_S	127.47	382.13	37.37	N.A.

In conclusion, pTREF2 performed satisfactorily in the *E. coli* strain tested and would be a suitable system for further testing in *E. coli*. When tested in *C. saccharoperbutylacetonicum,* pTREF2 did not produce fluorescence in either color (data not shown), however, this could be due to issues with induction conditions. However, pTREF3 did not function as expected in *E. coli*, combined with the difficulty in detecting fluorescence from phiLOV2.1 in *C. saccharoperbutylacetonicum*, a matched enzymatic dual reporter was preferred for further study.

4.5 Testing of the dual enzymatic reporter system

4.5.1 Assessment of compatibility of GusA and LacZ_Tts

First, as a continuation of the experiments described in Chapter 3, a cross-talk evaluation experiment was performed with *E. coli* DH5α strains harbouring plasmids pGUS-85151 (**Figure 4.14**) and pLAC-85151 (**Figure 4.15**). Briefly, we measured non-specific hydrolysis of the preferred substrate for one enzyme by the other. First, we compared GusA activity with p-nitrophenyl-β-D-glucuronide and o-nitrophenyl-β-D-galactopyranoside (**Figure 4.14**). We found that the GusA had detectable levels of

background activity toward o-nitrophenyl-β-D-galactopyranoside that increased significantly at 60°C, the increase in activity at the higher temperature towards the native substrate was more significant (**Figure 4.14**). The LacZ_Tts on the other hand had negligible levels of activity towards p-nitrophenyl-β-D-glucuronide (**Figure 4.15**) and the activity increase towards the non-native substrate at the higher temperature was not significant.

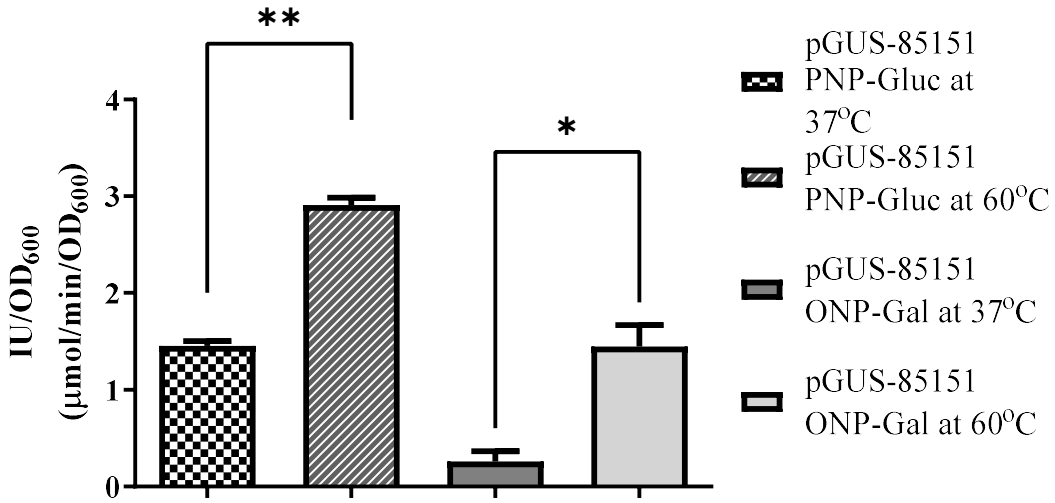


Figure 4.14 Assessment of non-specific activity of *E. coli* cells expressing GusA.

Activity of *E. coli* cells expressing GusA on the native substrate p-nitrophenyl-β-D-glucuronide (PNP) and the LacZ substrate o-nitrophenyl-β-D-galactopyranoside (ONP). Activity is expressed in IU/OD_{600} and the data is a mean of technical replicates (N = 2) of the same lysate for purposes of direct comparison. Error bars – standard deviation. ** - significant difference - *p*-value = 0.0018; * - significant difference - - *p*-value = 0.0206.

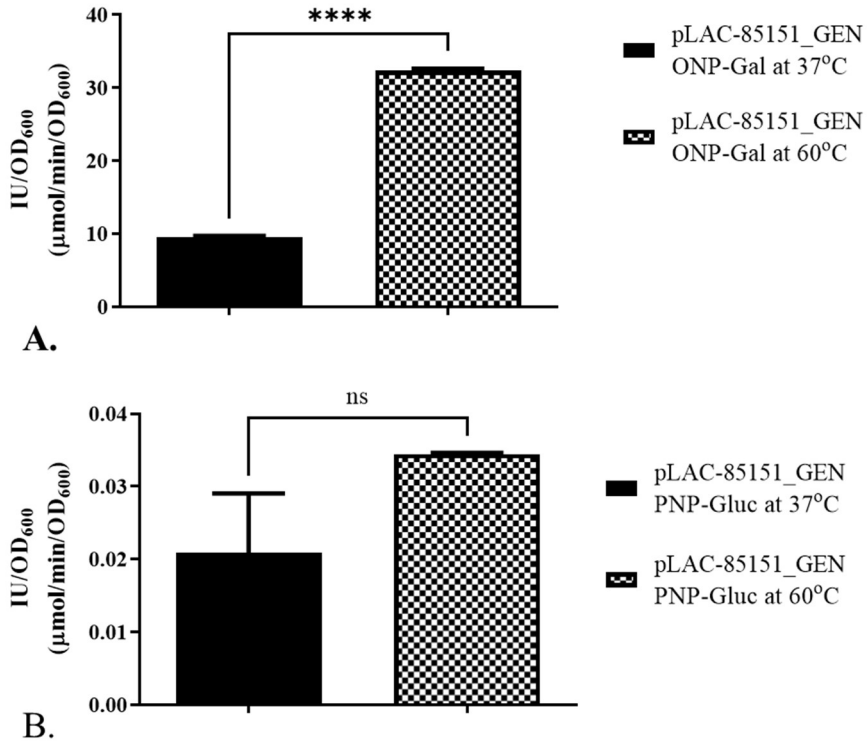


Figure 4.15 Assessment of non-specific activity of *E. coli* cells expressing LacZ_Tts.

A. Activity of *E. coli* cells expressing LacZ on the native substrate o-nitrophenyl-β-D-galactopyranoside (ONP-Gal) and (**B.**) the GusA substrate p-nitrophenyl-β-D-glucuronide (PNP-Gluc). Activity is expressed in IU/OD_{600} and the data is a mean of technical replicates (N=2) of the same lysate for purposes of direct comparison. Error bars – standard deviation. **A.** **** - significant difference - *p*-value < 0.0001 (Unpaired t-test); **B.** ns – no significant difference found – *p-value* = 0.1437 (Unpaired t-test).

We then tested the background activity of GusA toward the alternative β-galactosidase substrate p-nitrophenyl-β-D-galactopyranoside (**Figure 4.16**), adopted at the end of Chapter 3. The use of this substrate resulted in much lower non-specific activity by GusA and as shown in Chapter 3 LacZ_Tts activity toward it is comparable to o-nitrophenyl-β-D-galactopyranoside (**Figure 3.30**).

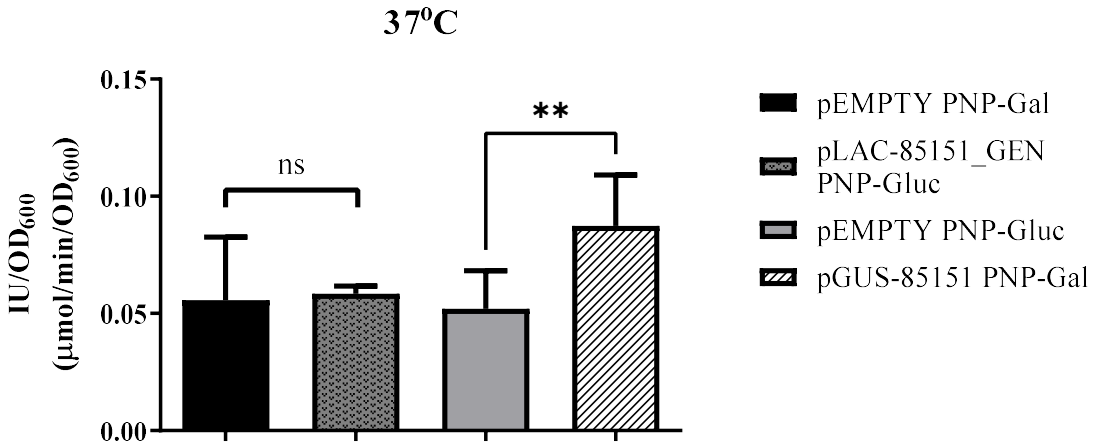


Figure 4.16 **Assessment of non-specific activity of GusA and LacZ_Tts on p-nitrophenyl-β-D-galactopyranoside and p-nitrophenyl-β-D-glucuronide, respectively.** Activity is expressed in IU/OD_{600} and the data is a mean of biological replicates (N=3) and technical replicates per biological replicate (pEMPTY plasmid - N=4, other plasmids - N=2); ns – not significant *p*-value = 0.8649 (Unpaired t-test); ** - significant *p*-value = 0.0085 (Unpaired t-test).

Therefore, p-nitrophenyl-β-D-galactopyranoside appears to be a much more suitable substrate for the purposes of a dual enzymatic reporter assay in 96-well polystyrene plates as there is a low level of cross-talk with GusA and a lower rate of diffusion out of wells and into neighbouring wells (data not shown).

4.5.2 Activity measurement of dual enzymatic reporter system in *E. coli*

To assess the function of the dual enzymatic reporters, they were first tested in *E.* coli before moving onto other organisms.

The dual reporter constructs pLAC-X-GUS_GEN, pLAC-B-GUS_GEN, pGUS-X-LAC_GEN and pGUS-X-LAC_GEN (described in **Section 4.3.2**) were transformed in *E. coli* DH5α. The resulting strains were incubated overnight, harvested and tested for

activity. These tests were performed with Bug Buster extracts which were aliquoted into matched samples to enable comparison of heat-treatment or no heat-treatment for testing β-galactosidase and β-glucuronidase.

The results indicated both reporters are being expressed in all 4 constructs (**Figure 4.17**), although some other interesting patterns were observed. In both reporter order variations the downstream reporter had a higher activity than the upstream reporter (**in Figure 4.17)** it should be taken into account that the activity of LacZ at 60^{o}C is about 5-fold higher than at 37^{o}C). The reasons behind this might be due to the strong RBS used for downstream reporter expression (derived from P*thlA_Cac*) whereas the upstream RBS is that present in pRPF185 (derived from *tcdB* of *C.difficile*). Additionally, it is possible that there is spurious intragenic transcription from the upstream reporter in each case like with *phiLOV2.1Opt*. However, that possibility cannot be fully assessed without a no-promoter control reporter construct.

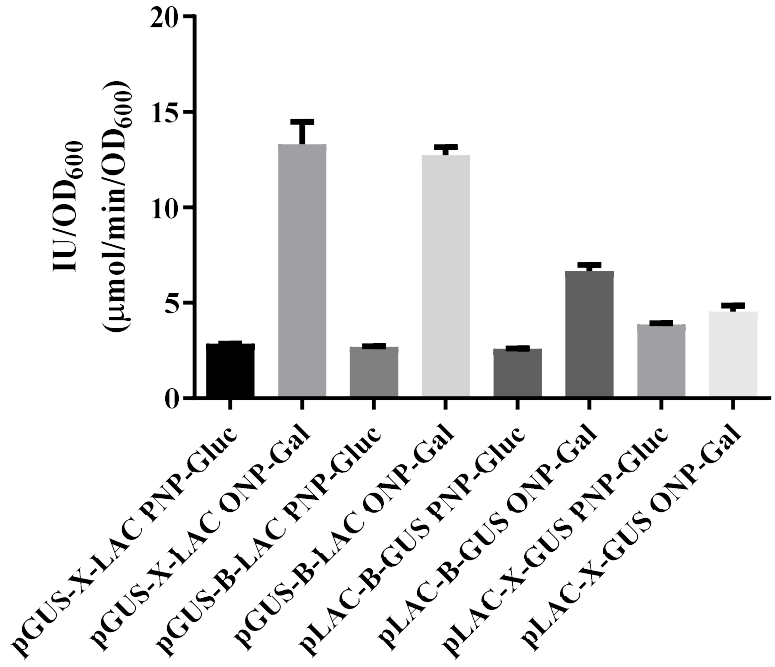


Figure 4.17 Activity measurements from the dual enzymatic constructs in *E. coli.*

Activity was measured with o-nitrophenyl-β-D-galactopyranoside at 60°C and activity was measured p-nitrophenyl-β-D-glucuronide at 37°C. Activity is expressed in IU/OD_{600} - and the data is a mean of technical replicates (N=2) of the same lysate, error bars represent standard deviation. Data shown is from a single representative experiment.

Based on the activity profiles of the LAC-GUS and GUS-LAC series, the LAC-GUS construct was chosen for further work and for sub-cloning of the terminator libraries. Part of the reasoning is that if two reporters have overlapping activity, then it would be preferable for the purposes of measuring a decrease in the downstream reporter to have that downstream reporter be the one with activity with both substrates as its expression would be more apparent even if it contributes to the signal of the upstream reporter.

We then compared the pLAC-X-GUS_GEN and pLAC-B-GUS_GEN constructs for activity using an updated protocol without heat treatment, with the assay performed at 37ºC using p-nitrophenyl-β-D-galactopyranoside and p-nitrophenyl-β-D-glucuronide as substrates (**Figure 4.18 A**). The β-galactosidase activities appear to be lower than those reported in **Figure 4.15**. However, it should be noted that the assays in **Figure 4.15** were performed at 60ºC which when using BugBuster increased activity of LacZ_Tts over 5-fold (see Chapter 3 for a comparison with toluene permeabilization). When the increase in activity with temperature is taken into account, the ratios of specific activities are similar. The pLAC-X-GUS construct consistently had higher β-glucuronidase activity and thus a higher β- glucuronidase/β-galactosidase ratio, indicating that the B-variant TIS v2 might be acting as a weak transcriptional terminator or lowering the levels of the downstream reporter in an unidentified way (**Figure 4.18 B**). This result confirmed our expectation that the "X" (for excised) TIS would serve as a better no-terminator reference plasmid control.

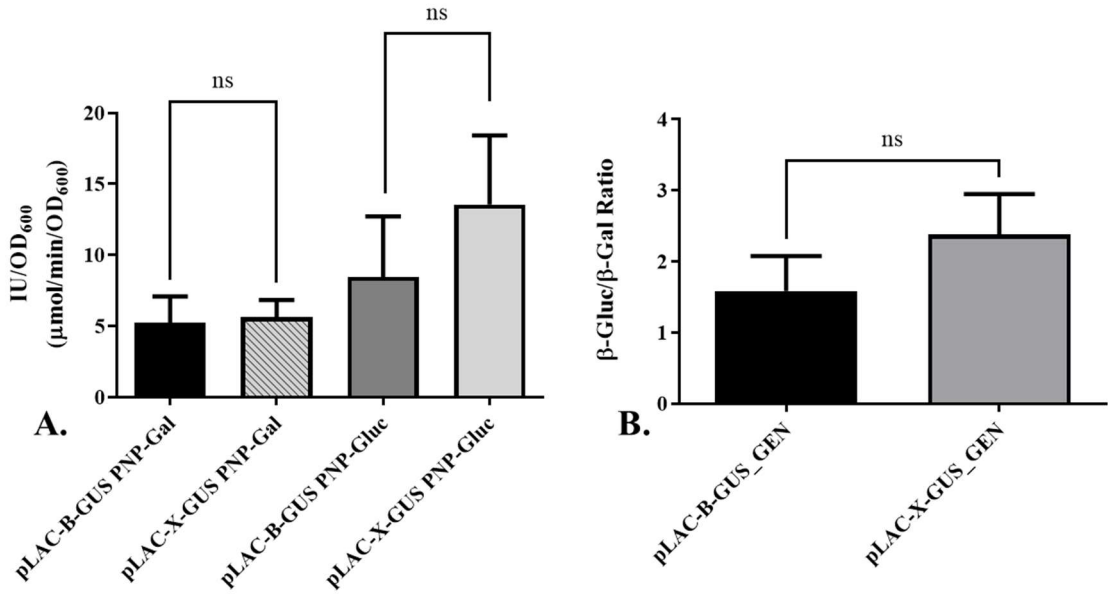


Figure 4.18 Comparison of reporter activities from pLAC-B-GUS and pLAC-X-GUS strains.

A. Activity is plotted as means of six measurements - biological replicates (N=3), technical replicates (N=2). PNP-Gal treatment: ns – not significant *p*-value = 0.6787

(Unpaired t-test); PNP-Gluc treatment: ns -not significant *p*-value = 0.0823 (Unpaired t-test). Error bars represent standard deviation. **B.** The ratio of activity is plotted as mean of three biological replicates (N=3) with two technical replicates each (N=2). ns – not significant *p*-value = 0.1409. Error bars represent standard deviation.

4.5.3 Activity measurement of dual enzymatic reporter system in *B. subtilis*

An advantage of the pIM13 origin used for working in *C. saccharoperbutylacetonicum* is that it also functions in *B. subtilis,* as do the selectable markers chosen, so we assessed whether there was scope to use the dual reporters in the model organism.

The four constructs (pLAC-X-GUS_GEN, pLAC-B-GUS_GEN, pGUS-X-LAC_GEN and pGUS-X-LAC_GEN) were transformed into *B. subtilis* str. 168 and overnight cultures were tested for enzyme activity. Overall, the downstream reporter had high measurable activity levels in the constructs and the activity ratio (DW/UP) of downstream/upstream reporters (β-glucuronidase/β-galactosidase for LAC-GUS or β-galactosidase/β- glucuronidase for GUS-LAC) were stable (**Table 4.3** and **Figure 4.19**) even though absolute expression levels varied owing to stochasticity and imperfections in our incubation and harvest protocol (overnight cultures in stationary phase as opposed to steady-state exponentially growing cultures).

Table 4.3 Activity measurements of dual enzymatic reporters in *B. subtilis.* Data shown is rounded to 3 significant figures. The ratio of activity is plotted as mean of three biological replicates (N=3) with two technical replicates each (N=2) except for pLAC-B-GUS_GEN for which activity was obtained only from a single biological replicate.

Strain	β-Galactosidase Activity(IU/OD_{600})	β-Glucuronidase Activity(IU/OD_{600})	Activity Ratio (DW/UP)
pGUS-B-LAC_GEN	1.45±0.434	0.27±0.12	5.52±0.61
pGUS-X-LAC_GEN	8.67±3.1	1.92±0.90	4.71±0.56
pLAC-X-GUS_GEN	0.92±0.55	5.73±3.62	6.17±0.31

pLAC-B-GUS_GEN	5.78	23.7	4.11

Interestingly, the pGUS-B-LAC_GEN had a slightly higher DW/UP ratio than pGUS-X-LAC_GEN, indicating that the B-variant TIS v2 does not have a big effect on the GUS-LAC construct in *B. subtilis*.

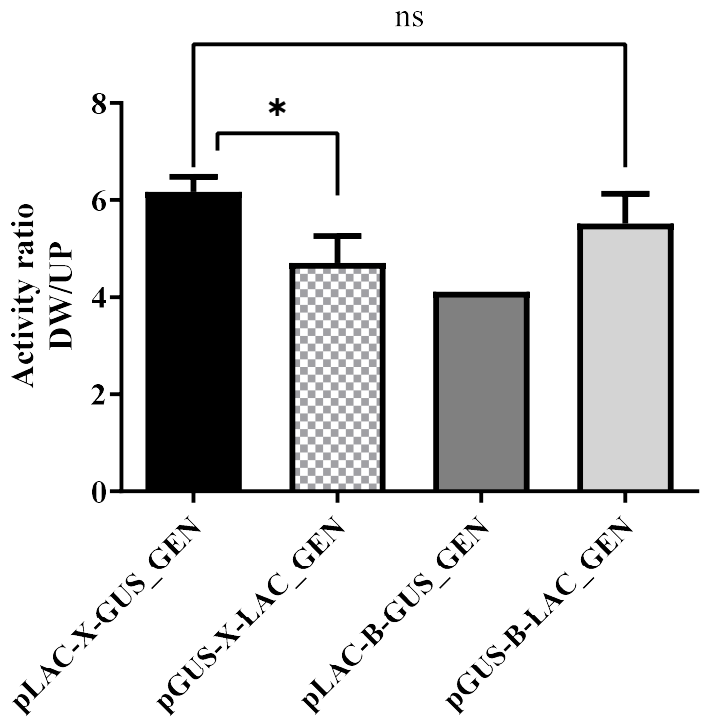


Figure 4.19 Ratios of activity measurements at at 37˚C of dual enzymatic reporters in *B. subtilis*. * - statistically significant *p*-value = 0.0161 (Unpaired t-test); ns – not significant, *p*-value = 0.1750 (Unpaired t-test). The ratio of activity is plotted as mean of three biological replicates (N=3) with two technical replicates each (N=2) except for pLAC-B-GUS_GEN for which activity was obtained only from a single biological replicate.

4.5.4 Activity measurement of dual enzymatic reporter system in *C. saccharoperbutylacetonicum*

Finally, the plasmid pLAC-X-GUS_GEN was transformed into *C. saccharoperbutylacetonicum* to serve as a reference plasmid for future terminator

assays. The activities of β-galactosidase and β-glucuronidase were measured (**Figure 4.20**). The former was measured using heat-treatment and incubated at 60ºC to minimize background activity found in *C. saccharoperbutylacetonicum* (see Materials and Methods and Chapter 3). The pLAC-X-GUS_GEN strain produced comparable levels of β-glucuronidase to the pGUS-85151 strains described in Chapter 3 (**Table 4.4**). The level of β-galactosidase activity we measured from pLAC-X-GUS_GEN was significantly higher than that from pLAC-85151_GEN (*p*-value <0.0001, Unpaired t-test) and was similar to that measured in some of our *E. coli* strains. It should be taken into account that the β-galactosidase values were obtained at an assay temperature of 60ºC and we expect them to be about 5-fold higher than at 37ºC based on results described in Chapter 3.

Table 4.4 Activities of β-glucuronidase and β-galactosidase expressed from plasmid pLAC-X-GUS_GEN in *C. saccharoperbutylacetobutylicum.* Data shown is rounded to 3 significant figures. Mean of 6 measurements biological replicates (N=3) and technical replicates (N=2).

Measurement	Mean	Standard Deviation	Relative Standard Deviation
β- Galactosidase Activity (IU/OD_{600})	9.83	2.1	21.3%
β- Glucuronidase Activity(IU/OD_{600})	0.24	0.098	41.4%
Ratio (UP/DW)	45.9	13.6	29.6%
Ratio (DW/UP)	0.023	0.007	29%

The resulting ratios of upstream/downstream reporter were quite high (**Table 4.4**) and were the opposite of the trends we observed in *E. coli* and *B. subtilis* where the downstream reporter had higher levels regardless of its identity in the dual enzymatic reporter series.

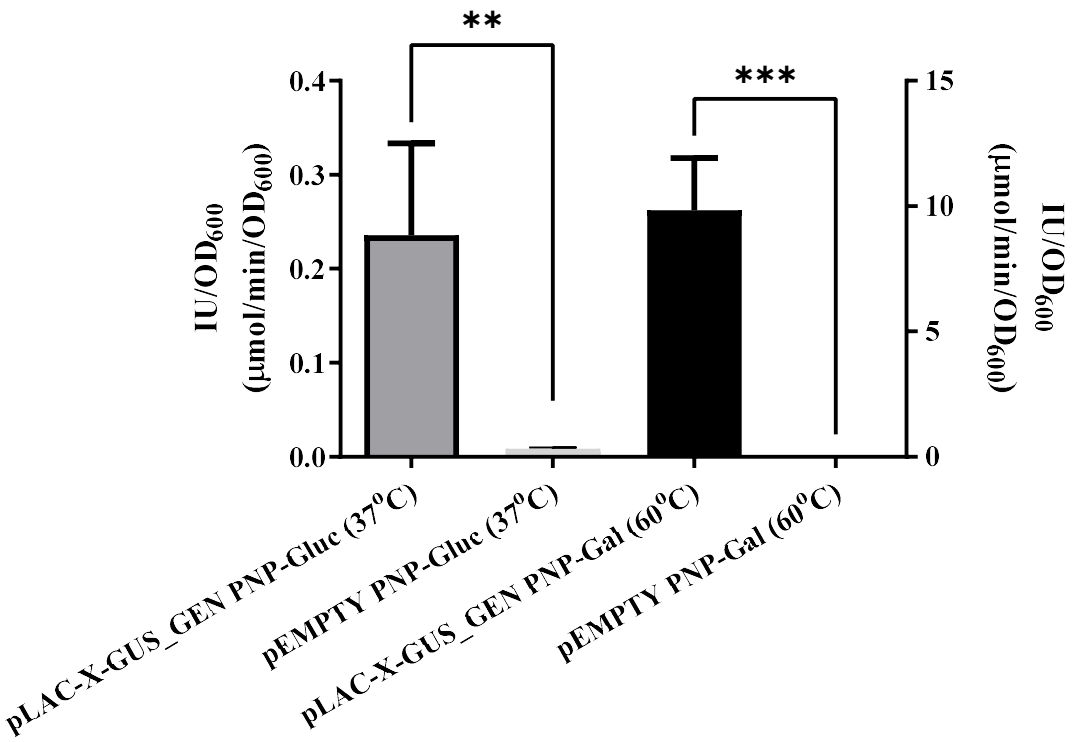


Figure 4.20 Activities of β-glucuronidase and β-galactosidase expressed from plasmid pLAC-X-GUS_GEN in *C. saccharoperbutylacetobutylicum*. Data shown are mean of biological replicates (N=3) and technical replicates (N=2), error bars are standard deviation (SD) to the mean; ** - significant *p*-value = 0.0059 (Unpaired t-test), *** - significant *p*-value = 0.0001 (Unpaired t-test). Note the different scale of left y-axis vs the right y-axis.

To conclude, the pLAC-X-GUS_GEN dual reporter was chosen for the work with the terminator library described in Chapter 5 in large part because of the higher activity of the β-D-glucuronidase reporter in *E. coli* and *B. subtilis*.

The key features of pLAC-X-GUS_GEN compared to the other constructs developed in this study and some of the constructs from the literature are non-overlapping suitably detectable levels of the two reporters, an insulator and spacer region to minimize false positives (termination inhibtion) and false negatives (translation inhibition). Furthermore, the system is on a shuttle vector and allows cross-species comparative characterization.

Chapter V

Comparative analysis of transcriptional terminators in *Escherichia coli*, *Bacillus subtilis* and *Clostridium saccharoperbutylacetonicum*

5. Comparative analysis of transcriptional terminators in *E. coli*, *B. subtilis* and *C. saccharoperbutylacetonicum*

In this final results chapter, the tools developed and tested in Chapters 3 and 4 are used to assess a range of natural and predicted terminators from a range of organisms for use in three phylogenetically distinct organisms that are widely used in molecular microbiology or industrial biotechnology.

5.1 Terminator Library Creation

To assess a wide range of terminators for application in *C. saccharoperbutylacetonicum* and other bacteria, we used a number of different approaches. These largely fall into two main classes. The first was to use the existing literature to construct a library of known terminators, both natural and synthetic, from a range of sources, including a set from Firmicutes. The second was to discover, using computational predictions, new sets of terminators from the solventogenic *Clostridium* species. In the following section, these various strategies are described and the terminators chosen for study are defined.

5.1.1 Literature survey

5.1.1.1 Selection of a set of E. coli Natural and BioBrick terminators from literature

In order to expand the *Clostridium* synthetic biology toolbox we started to create a library of transcriptional terminators which could be characterized using the dual enzymatic reporter system (described in Chapter 4). A survey of the literature returned several candidates derived from the *Escherichia coli* genome (ECK-series) that had been previously characterized by Chen *et al.* (2013)[148]. In order to validate the range and sensitivity of the reporter system seven *E.coli* natural terminators were chosen (Table 5.1). Additionally, 4 commonly used terminators from the BioBrick biological

part collection were selected (Table 5.1). We termed this group of terminators "Natural and BioBrick".

The previously measured termination efficiencies of these terminators ranged from weak (TE ≤ 65%) through medium (65% < TE≤80%) to strong (80% < TE ≤ 95%) and very strong (95% < TE ≤ 99%) (***Table 5.1*** for general features, ***Table 5.8*** for published strength), TE was calculated according to ***Equation 4.3***. The weak, medium, strong and very strong categories are subjective and specific to this work, to the author's knowledge exact values to group terminators into strength/efficiency categories have not been agreed upon. Chen *et al* (2013) defines strong terminators as having T_S>40 (TE > 97.5%) and weak ones as having T_S < 3 (TE < 70%) (T_S was calculated according to ***Equation 4.4***). Overall, most of the terminators selected in the previously measured group were expected to be more than 90% efficient. Two of the BioBrick terminators (BBa_B0053m and BBa_B0062m) we selected are used as the multiple-cloning site (MCS) flanking terminators in a very commonly used vector series for synthetic biology in *E. coli* – pSB**3. The terminator sequences we chose match the ones found in these vectors which differ from the sequence of the deposited parts under the same code name by one or more single nucleotide polymorphisms (SNPs) and thus the terminators are designated "m" for mutant. These terminators were selected with the goal of testing them in order to evaluate whether potential shuttle vectors derived from the aforementioned vector series would have well insulated MCSs in *Bacillus* or *Clostridium* species and thus function as expected in *E. coli.* In addition, testing in *E. coli* would provide information on whether the SNPs found in these terminators affect efficiency.

Table 5.1 Natural *E. coli* and BioBrick set of terminators. Minimum free energy (ΔG) was calculated using RNAfold[304].

Terminator Name	Length (bp)	Sequence (5'-3')	ΔG (kcal/mol)	Source Gene	Source Organism	Reference
BBa_B0010	80	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTC	-39.00	*rrnB* T1 loop	*E.coli* K12	148, 24, 305
Bba_B0015	129	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA	-60.00	*rrnB* T1 loop + TE T7	*E.coli* K12 + T7	305
BBa_B0053m	58	TCCGGCAAAAAAGGGCAAGGTGTCACCACCCTGCCCTTTTTCTTTAAAACCGAAAAGA	-19.10	*his* operon	*E.coli* K12	305
BBa_B0062m	45	ATTTCAGATAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCT	-12.00	*rrnC*	*E.coli* K12	305
ECK120010783	49	ACGAGCCAATAAAAATACCGGCGTTATGCCGGTATTTTTTTACGAAAGA	-15.90	*mdoGH*	*E.coli* K12	148
ECK120010800	44	AGTTTGTTCGCCCGGTAGTTGTGACGCTACCGGGTTCTTTTCGA	-17.10	*gntKU,gntRKU*	*E.coli* K12	148
ECK120015170	47	ACAATTTTCGAAAAAACCCGCTTCGGCGGGTTTTTTTATAGCTAAAA	-16.20	*rplM-rpsI*	*E.coli* K12	148
ECK120029600	90	TTCAGCCAAAAAACTTAAGACCGCCGGTCTTGTCCACTACCTTGCAGTAATGCGGTGGACAGGATCGGCGGTTTTCTTTTCTCTTCTCAA	-40.30	*spy*	*E.coli* K12	148
ECK120033127	45	TACTTCTTACTCGCCCATCTGCAACGGATGGGCGAATTTATACCC	-18.00	*sdaA*	*E.coli* K12	148
ECK120033737	57	GGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGG	-23.70	*thrLABC*	*E.coli* K12	148
ECK125108723	78	TAGCGTAAAAGCAAAACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTTTCGCCGTCATA	-13.60	*mgrR*	*E.coli* K12	148

5.1.1.2 Selection of a set of Synthetic terminators from literature

The same study which the majority of the Natural set of terminators were derived from also produced a series of synthetic terminators derived from strong natural *E. coli* terminator scaffolds via site-directed mutagenesis and component shuffling (referred to as the Synthetic set). Several of these synthetic terminators exhibited high termination efficiency and some were different enough to be recombination resistant (dissimilar amongst each other)[148]. In this case, recombination resistance (as opposed to recombination propensity) refers to the lack of or low frequency of observed homologous recombination between two sequences because of a sufficient lack of similarity (similarity is sufficient to promote homologous recombination[306]). We picked ten of these synthetic terminators (***Table 5.2***), again with variable previously measured strengths, but a majority of strong terminators (TE>90%). Terminators with the code names LxUxHx vary in the U-tract (Ux) and Hairpin (Hx) while Lx refers to the library number. The L1 library terminators were derived by combining designed synthetic U-tracts and designed hairpin stems (with varying free energy of folding) while the loop sequence was fixed. We picked two of those terminators. The L2 library terminators were created based on a fixed stem sequence with a variable loop and U-tract. We picked one of these terminators. Both the L1 and L2 libraries produced mostly weak to moderately strong terminators in the original study even when perfect U-tracts were used[148]. The terminators designated with the LxSxPx code names are all part of L3 library which vary in the strong natural terminator scaffold used (Sx) and mutant variant (Px).

Table 5. 2 Set of synthetic terminators. Minimum free energy (ΔG) was calculated using RNAfold[304].

Terminator Name	Length (bp)	Sequence (5'-3')	ΔG (kcal/mol)	Source Gene	Source Organism	Reference
L1U1H08	29	CCCGCATGTTCGCATGCGGGTTTTTTTTT	-13.20	synthetic library 1	synthetic	[148]
L1U1H09	29	CGACGATGTTCGCATCGTCGTTTTTTTTT	-10.50	synthetic library 1	synthetic	[148]
L2U2H09	28	ACGGCCCTCGCAAGGGCCGTTTTTTTGT	-14.10	synthetic library 2	synthetic	[148]
L3S1P11	50	GACGAACAATAAGGCCTCCCTTCGGGGGGGC CTTTTTTATTGATACAAAA	-19.40	synthetic library 3 - scaffold 1 - pheA-1	synthetic	[148]
L3S1P32	52	GACGAACAATAAGGCCTCCCAAATCGGGGGG CCTTTTTATTTTTCAACAAAA	-19.20	synthetic library 3 - scaffold 1 - pheA-1	synthetic	[148]
L3S1P51	52	AAAAAAAAAAAAGGCCTCCCAAATCGGGGGG CCTTTTTTATTGATAACAAAA	-22.10	synthetic library 3 - scaffold 1 - pheA-1	synthetic	[148]
L3S2P00	63	CTCGGTACCAAATTCCAGAAAAGAGGGGAGC GGGAAACCGCTCCCCTTTTTTCGTTTTGGTC C	-37.80	synthetic library 3 - scaffold 2 - ECK120034435	synthetic	[148]
L3S2P21	61	CTCGGTACCAAATTCCAGAAAAGAGGCCTCC CGAAAGGGGGGCCTTTTTTCGTTTTGGTCC	-34.30	synthetic library 3 - scaffold 2 - ECK120034435	synthetic	[148]
L3S2P36	57	CTCGGTACCAAATTCCAGAAAAGAGACGCTG AAAAGCGTCTTTTTTATTGATGGTCC	-20.90	synthetic library 3 - scaffold 2 - ECK120034435	synthetic	[148]
L3S2P56	57	CTCGGTACCAAATTTTCGAAAAAAGACGCTG AAAAGCGTCTTTTTTCGTTTTGGTCC	-26.60	synthetic library 3 - scaffold 2 - ECK120034435	synthetic	[148]

5.1.1.3 *Selection of a set of Firmicute-derived terminators previously used in Clostridium species*

In addition to the well-characterized natural, BioBrick and synthetic terminators derived from and used in *E. coli*, a list of terminators previously used in *Clostridium* engineering projects was compiled, referred to as '*Firmicutes'* set. A majority of these are derived from *Clostridium* species (8), while some are from *Lactobacilli* (2) and *B. subtilis* (3). Six of the total 13 chosen were part of a screen in *C. acetobutylicum* for terminators and have not been used in engineering projects; this study is to the author's knowledge the only *in vivo* study of terminators in a solventogen to date. One of the terminators picked (T*thl_Cspba*) was used in an unpublished study from our laboratory and is derived from the main thiolase gene of the *C. saccharoperbutylacetonicum* N1-4 (HMT) genome. There are two more thiolase terminators in this list (T*thlA_Cac* and T*thl_Cbeij*), one from acetoacetate decarboxylase (T*adc*) and ferredoxin (T*fdx*), thus highlighting the trend to choose terminators from genes encoding central metabolic enzymes. Several terminators are used in commonly used vectors - T_{CD0164} and T*fdx* in the pMTL80000 series and T*fdx* and T*slpA_Cdiff_ext* from the pRPF185 plasmid. Since the start of this work, we identified only three additional terminators that have been reported (see Introduction, Table 1.2).

Table 5. 3 Set of Firmicutes terminators. Minimum free energy (ΔG) was calculated using RNAfold[304].

Terminator Name	Length (bp)	Sequence (5'-3')	ΔG (kcal/mol)	Source Gene	Source Organism	Reference
T*tyrS*	45	ATAATCAATCGTCCCTTCGTGTAAACGAAGGGGCGTTTTTTATTT	-19.30	*tyrS*	*B. subtilis*	24
T*phiTD1*	61	AACAATCAAAAGAAAAGCCTATCGTCTGAGGAACGGTAGGCTCTTTTGTAGCATATAGTTG	-17.40	phage Φ29 late TD1	*B. subtilis*	24, 260, 261
T*gyrA*	54	AAGAAGAAGTGTGAAAAAGCGCAGCTGAAATAGCTGCGCTTTTTTGTGTCATAA	-20.50	*gyrA*	*B. subtilis*	24
T*thlA_Cac*	47	AAAAATAACTCTGTAGAATTATAAATTAGTTCTACAGAGTTATTTTT	-25.00	*thl*	*C. acetobutylicum* ATCC 824	218
T*adc*	83	TAATAAAAATAAGAGTTACCTTAAATGGTAACTCTTATTTTTTTAATATTGTTTCATAGTATTTCTTTCTAAACAGCCATGGG	-20.90	*adc*	*C. acetobutylicum* ATCC 824	259
T*thl_Cbeij*	101	ATATAAATTAAGATTTAAAAAGGTTACTATGATAATTCTCATGGTAACCTTTTTTTATTAAATAAGAGTATAAAATAAAGTTAAAAGAAGAAAATAGAAAT	-20.00	*thl*	*C. beijerinckii* NCIMB 8052	85
T*slpA_Cdiff*	43	AAATATAAAAAGACTTCTCAGATGAGAAGTCTTTTTTGTGAAA	-15.60	*slpA*	*C. difficile* DSM 27639	24
T_{CD0164}	54	ATAAAAAAATTGTAGATAAATTTTATAAAATAGTTTTATCTACAATTTTTTTTAT	-18.40	ORF0164	*C. difficile* 630	53
T*fdx*	42	ATAAAAATAAGAAGCCTGCATTTGCAGGCTTCTTATTTTTAT	-26.50	*fdx*	*C. pasteurianum*	214
T*thl_Cspba*	142	GTATACAAGTTCACATTCGCAACAAGTTACTATGATAAGATATATTATCATAGTAACTTTTTTATATAATAAAAATTTTAATTGATTCGGTATAAAAGAGTATACTAAAGAGGAAAATAGTATTTGAATTAGTGAGTGATTG	-25.30	*thl*	*C. saccharoperbutylacetonicum* N1-4 (HMT)	R. Hennessy *et al.*, unpublished
T*slpA_Lac*	33	TGAAAAAGGCAGAGCGAAAGCTCTGTCTTTTTT	-19.30	*slpA*	*L. acidophilus*	24
T*slpA_Cdiff*_ext	51	CTTTAAATAGAAAAAGGCTTCTCTCATGAGAAGTCTTTTTTATTTAAAATA	-20.10	*slpA*	*C. difficile* 630	228
T*pepN*	53	TAATTTATAAATAAAAATCACCTTTTAGAGGTGGTTTTTTTATTTATAAATTA	-19.40	*pepN*	*L. lactis*	24

5.1.2 Terminator predictions using published algorithms and comparative analysis

Apart from the list of terminators previously used in *Clostridium* species there is no equivalent database to RegulonDB[307] for *Clostridium* species, from which many of the natural *E. coli* set were derived[148].

Therefore, after compiling a library of published and previously used terminators for testing, an *in silico* approach was undertaken for the discovery of additional Clostridial terminators for screening. The conserved structure of intrinsic terminators, consisting of a hairpin-stem loop, almost invariably followed by a stretch of uracils (termed U-tract) makes them comparatively easy to detect in genomes[308]. Several bacterial intrinsic terminator prediction algorithms have been published to date. TransTermHP[255], RNIE[241] and WebGeSTer[257] as well as web-based tools such as ARNold[309] are some of the more commonly cited prediction algorithms. These algorithms differ in the method of putative terminator detection.

We selected the first three of the aforementioned programs for comparison of the overlap of their predictions. Five genomic sequences were used for predictions with the three algorithms using default settings. The results of the predictions were compared using the predicted terminators' coordinates and a Python script (courtesy of Steve Thorpe, University of York). The script compares the positions of the terminators and detects overlaps between the unique predictions of each algorithm. The results showed that while TransTermHP and WebGeSTer produced comparatively large numbers of predictions only between a third and a half of those overlapped in any way (Figure 5.1). All three algorithms did not share many predictions. The RNIE algorithm produced very

few total predictions and shared more with TransTermHP than WebGeSTer. An example of the kind of overlap produced by TransTermHP, WebGeSTer and RNIE is shown in Figure 5.2A. The example shown is of a convergent gene region from *C. acetobutylicum* and a putative bidirectional terminator. The output of the programs differs slightly - the TransTermHP program's predictions encompass the entire length of the bidirectional terminator (including the so-called A-tract, corresponding to a U-tract in the opposite direction) while RNIE and WebGeSTer do not include the A-tract in their predictions. Bidirectional symmetric terminators like the one shown form some of the overlap between the algorithms (Figure 5.2B). It is important to note that most symmetrical terminators still differ in their secondary structure due in part due to G:U base-pairing, the reverse-complementary A:C does not form a stable base pair.

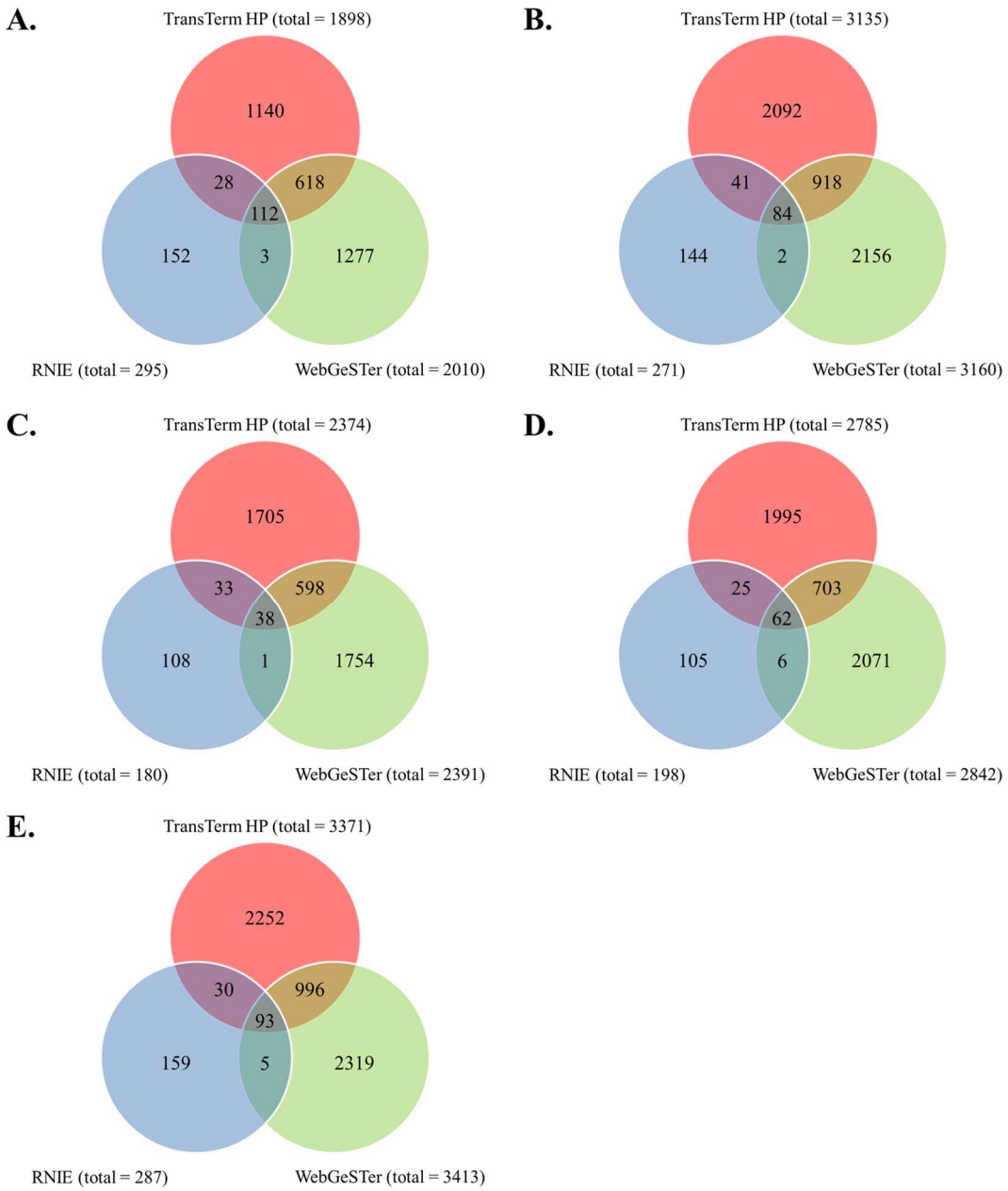


Figure 5. 1 Comparisons of terminator outputs from three popular prediction software

A. *C. acetobutylicum* ATCC 824 NC_003030 **B.** *C. beijerinckii* NCIMB 8052, accession CP000721 **C.** *C. pasteurianum* DSM 525, accession CP009268 **D.** *C. saccharobutylicum* DSM 13864, accession CP006721 **E.** *C. saccharoperbutylacetonicum* N1-4 (HMT), accession CP004121.

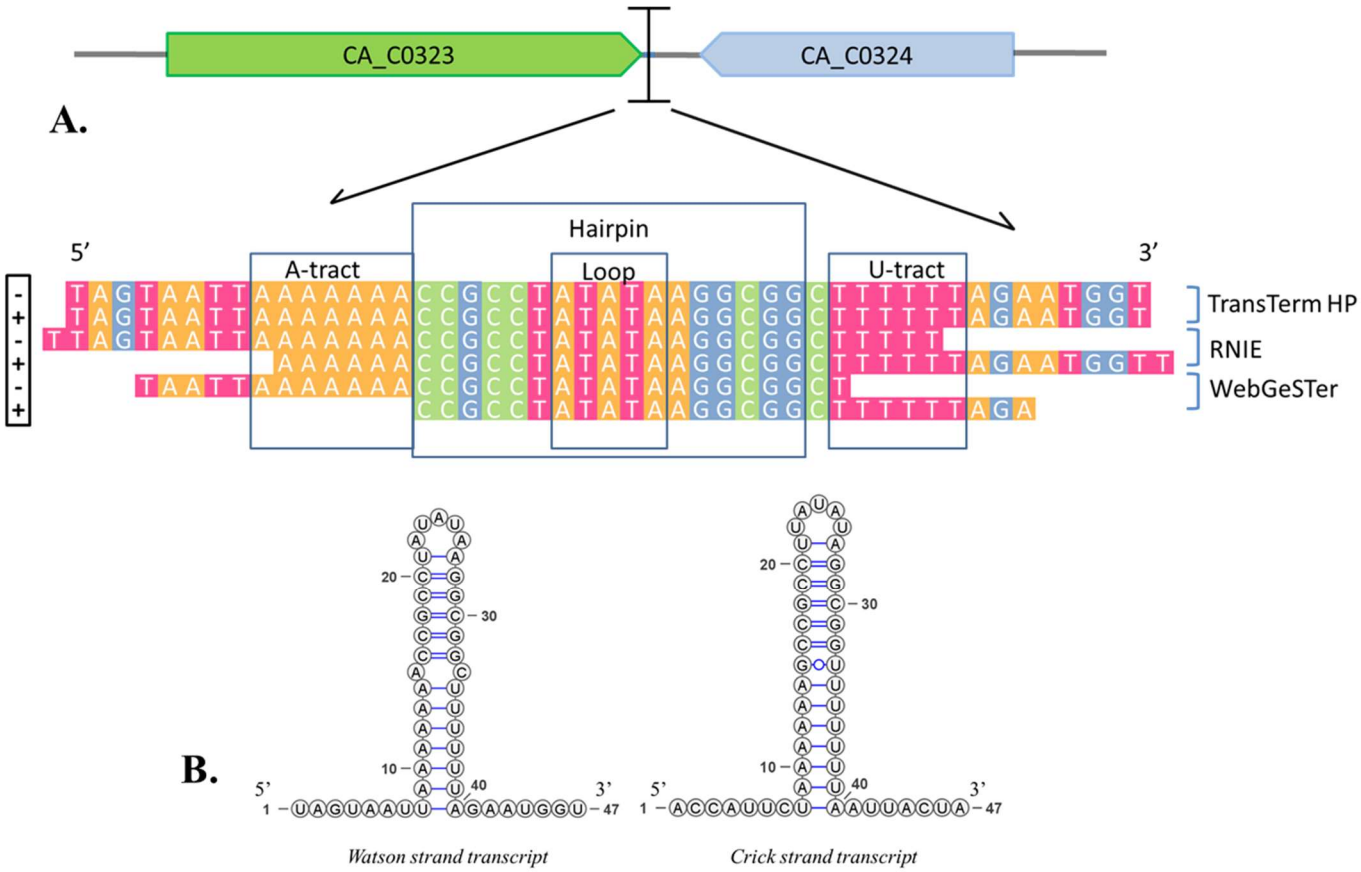


Figure 5. 2 Application of terminator prediction software to *C. acetobutylicum.*

A. An alignment of predictions of a putative bidirectional terminator (marked by T sign) by all three algorithms (note all three produce separate predictions for plus (+)and minus (-)) **B**. RNA structures predicted by RNAfold[304,310] for the TransTermHP sequence drawn with VARNA[311]; it is worth noting that the base-pairing between the A-tract and U-tract may not form until after termination[312].

The genomes of the five solventogenic *Clostridium* species (a total of eight replicons – 5 chromosomes and 3 plasmids) were tested with TransTermHP and RNIE using local installation of the programs. The total number of predictions from TransTermHP was 19184 while RNIE produced 1359. The number of predictions did not increase in a linear fashion with genome size.

Table 5. 4 List of accessions used for predictions and summary of results.

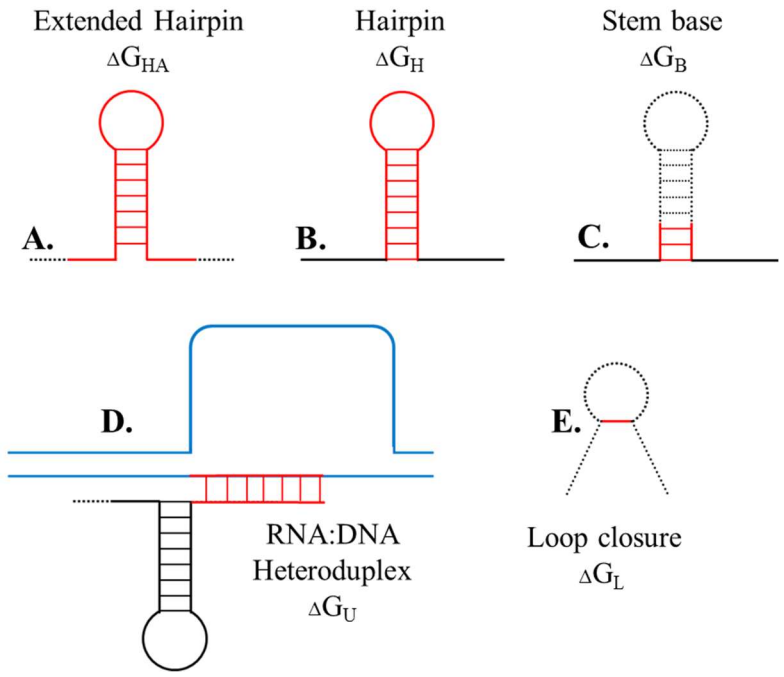
Replicon Name	Accession	Size	TransTermHP	RNIE
C. acetobutylicum ATCC 824 chromosome	NC_003030	3.94 Mb	4302	295
C. acetobutylicum ATCC824 megaplasmid pSOL1	NC_001988	192 kb	109	13
C. pasteurianum BC1 chromosome	CP003261	4.99 Mb	2347	288
C. pasteurianum BC1 plasmid pCLOPA01	CP003262	53 kb	25	2
C. saccharobutylicum DSM 13864 chromosome	CP006721	5.1 Mb	2380	198
C. saccharoperbutylacetonicum N1-4 (HMT) chromosome	CP004121	6.53 Mb	2996	287
C. saccharoperbutylacetonicum N1-4 (HMT) plasmid Csp_135p	CP004122	136 kb	34	5
C. beijerinckii NCIMB 8052 chromosome	CP000721	6.00 Mb	6991	271

As shown above, the algorithms produced thousands of predictions and while they provide a score of confidence there is no metric for predicted efficiency built into those programs. We therefore sought to increase the probability of discovering functional and strong terminators from the set of putative terminators.

5.1.3 *In silico* terminator scoring by implementation of a published biophysical model of transcription termination

There are several published biophysical models of transcription termination such as the Von Hippel model and the Thermes model. More recently Cambray *et al.*[313] and Chen *et al*[148] developed models for predicting transcription termination efficiency.

The Chen model is a kinetic model that incorporates free energy predictions of elements of the secondary structure of a terminator together with several fit parameters based on rate of hairpin folding, rate of RNAP progression, rate of RNAP progression after folding and rate of U-tract displacement. This model was shown to be an improvement over using only one of the commonly used free energy parameters or prediction algorithm confidence scores. The secondary structure features used for calculations are the Extended Hairpin (***Figure 5.3A***), the Hairpin (***Figure 5.3B***), the Stem base (***Figure 5.3C***), the Loop closure (***Figure 5.3E***) and the RNA:DNA heteroduplex (***Figure 5.3D***). The minimum free energies of these elements are used in Equation 5.1 (***Figure 5.3F***) to calculate predicted terminator strength (Ts).



F.

$$T_S = 1 + \frac{1}{B_1 e^{\beta_1 \Delta G_L} + B_4 e^{\beta_4 (\Delta G_B + \Delta G_A - \Delta G_u)} (1 + B_1 e^{\beta_1 \Delta G_L})} \qquad \textit{Equation 5.1}$$

Figure 5. 3 Schematic of different secondary structure elements that were considered in the Chen biophysical model of terminator strength[148].

A-E. The features used in the terminator model. F. The equation used for predicted terminator strength (T_S) calculation where $B_1 = 0.005$, $B_4 = 6.0$, $\beta_1 = 0.6$, and $\beta_4 = 0.45$ are fit parameters.

We chose to apply the Chen model to the dataset of putative terminators as its implementation was deemed more rigorous given the higher level of methodology documentation that was available than for other comparable models such as the Cambray model[313]. The workflow described in the original publication was adapted for ease-of-use and does not require custom-made software (**Figure 5.4**).

To start the analysis, the redundant terminators in the list were removed from the combined TransTermHP set, reducing the number of sequences from 19184 to 11577 terminators. From this set 11519 reverse complementary sequences were generated. A

total of 23096 sequences were analyzed – this set was referred to as "Unique Forward Plus Reverse". Next, the secondary structures of these sequences were generated using RNAfold of the Vienna RNA package[304]. From these overall secondary structure predictions, the location of the loop was recorded. The sequence of the terminator and position of the loop were then added to a custom Excel spreadsheet. Then using Excel formulas, the sequence downstream of the loop was parsed into 8bp fragments and these were used to identify the highest scoring U-tract downstream of the loop. The principle of U-tract scoring relies on the free energy of binding of the RNA U-tract and template DNA[244]. The scoring is done by breaking down the 8bp fragment into dinucleotides; the free energy contribution of each dinucleotide is looked up in a table[148,244]. The free energy contributions are summed together with the initiation term of RNA:DNA hybridization which is a constant to give ΔG_U[148]. Based on the start position of the U-tract we took the distance from the loop to the U-tract to be the right arm of the stem loop, an equal distance was taken to be the left arm of the stem loop. The resulting sequence was assumed to be the hairpin+U-tract (left arm+loop+right arm+U-tract). RNAeval[304] was used to determine the free energy of folding of the hairpin+U-tract (ΔG_H)[148]. Larger loops tend to have a higher (less favourable) free energy value and the energy is also heavily influenced by the identity of the closing basepair[314]. Then, the A-tract was taken to be the 8bp upstream of the left arm and the free energy of folding of the extended hairpin (ΔG_{HA}) was also determined using RNAeval[304].

A number of parameters were then obtained directly from the RNAeval output - the A-tract parameter ΔG_A is the difference of the extended hairpin ΔG_{HA} and ΔG_H. RNAeval was also used to obtain the energy of loop closure ΔG_L and the energy of the last three closing base pairs in the stem loop ΔG_B.

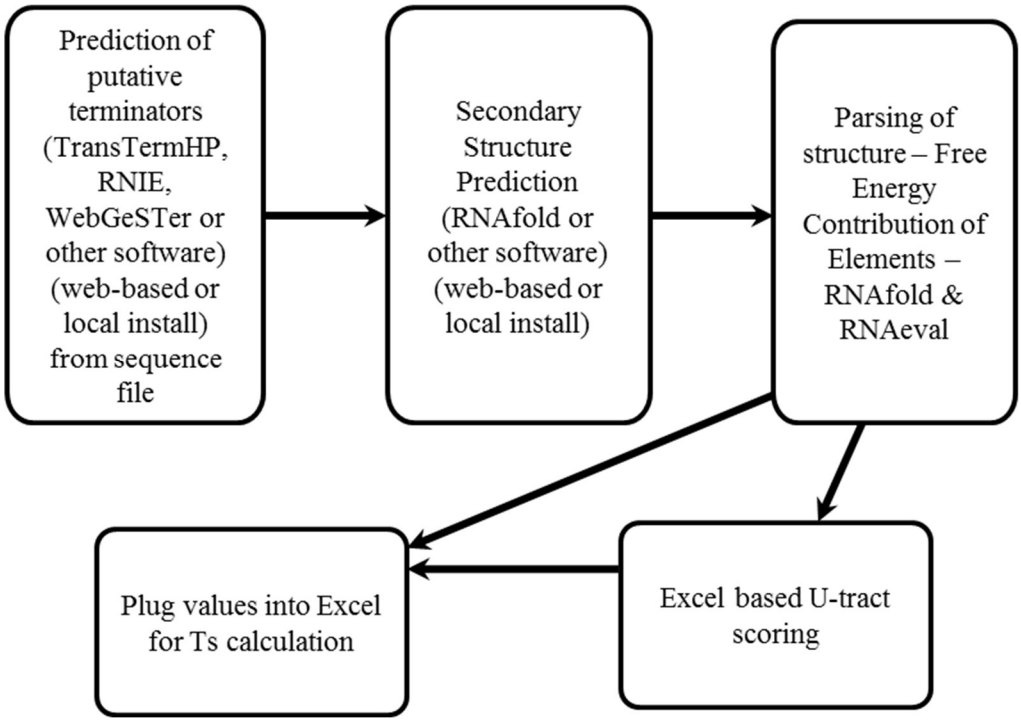


Figure 5. 4 Workflow for the prediction of the *Predicted* terminator set.

The results of the predicted terminator strength (Ts) resembled a truncated normal distribution (***Figure 5.5A***) with a cut-off at Ts of 1. This is because Ts of 1 indicates a TE of 0%. A value below a TE of 0% or Ts of 1 indicates promoter activity which this model is not able to predict. The distribution of predicted strength of the putative set was compared to that of the *E. coli* natural terminators from the Chen study[148] (***Figure 5.5B***). It is noticeable that even with a nearly 73-times greater sample size there were no terminators from the *Clostridium* predicted set that had a higher predicted Ts values than 48. In fact, only 210 predicted terminators (from either direction so this could contain reverse complementary terminators) score above a predicted Ts of 19 (equivalent to a predicted TE of 94.7%). In contrast, the *E. coli* natural set from Chen *et*

al. (2013) had over 5% of sequences scoring above a predicted Ts of 25 with a maximum Ts prediction nearly 3-times higher at 131.4 as opposed to 47.43 of the *Clostridium* set.

Possible reasons for this discrepancy might be the markedly lower genomic GC-content of Clostridia – around 30% or lower. While canonical terminators' hairpin stems are GC-rich, the hairpin stem-loops of the *Clostridium* set centred around 30% (***Fig. 5.6***).

The free energy of folding for the closure of the loop was also compared between the predicted set (***Fig. 5.7A***) and the natural *E. coli* set (***Fig. 5.7B***) – the former had a slightly higher median value than the main peak of the *E. coli* set. However the *E. coli* set appeared bi-modal with a second smaller peak with significantly lower energies (more favourable loop closure) which was completely absent from the *Clostridium* set.

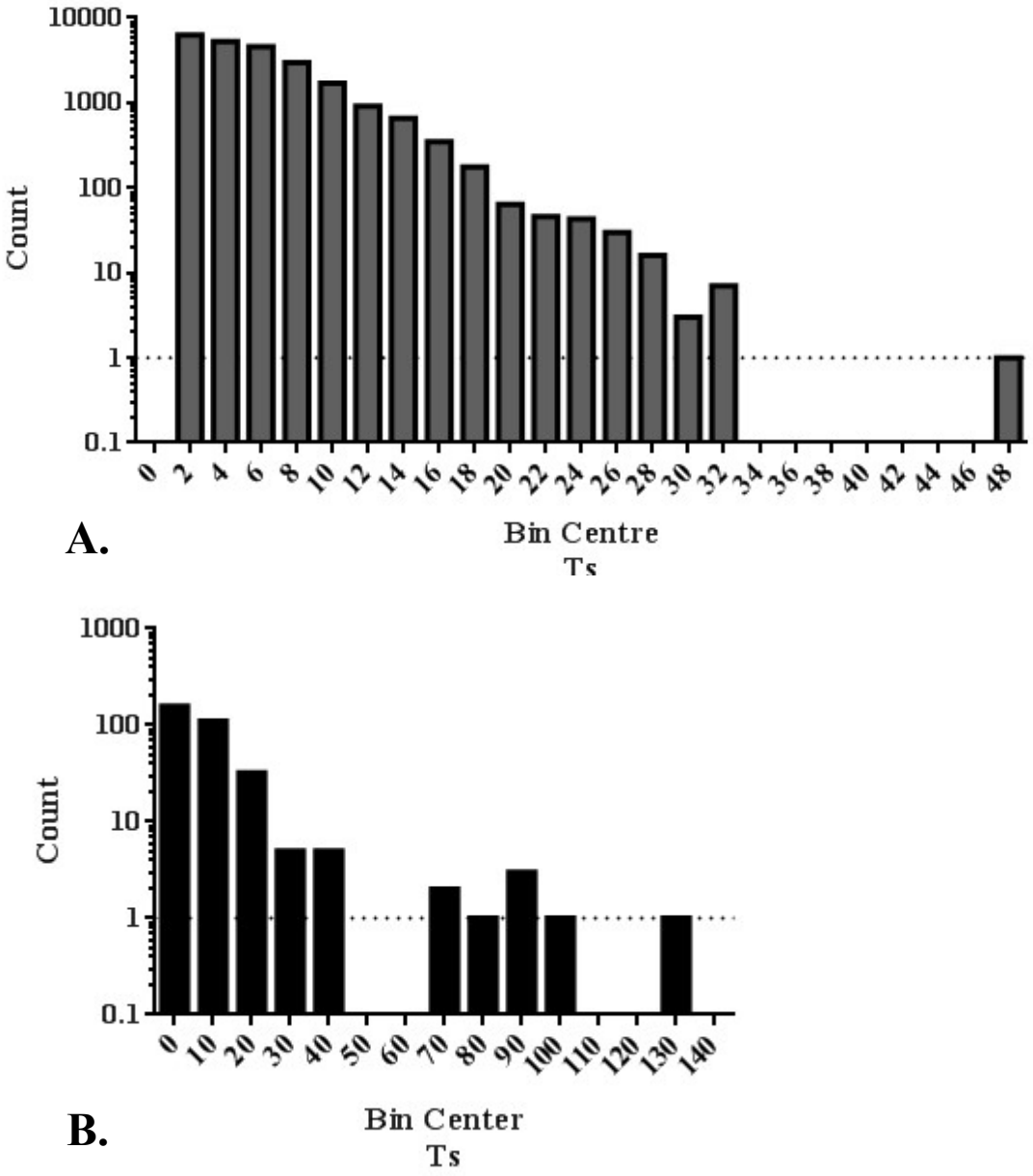


Figure 5.5 Distribution of predicted scores of the putative terminator set of "unique forward plus reverse" (n=23096) (**A**), compared to predicted scores of 'natural' set (n=317) from Chen *et al.* (2013)[148] (**B**).

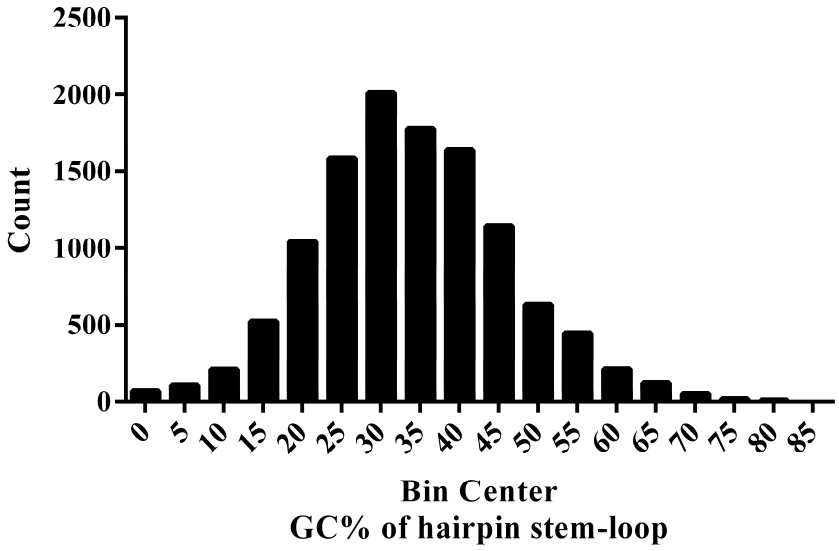


Figure 5.6 Distribution of the GC% of the hairpin stem-loops of the putative terminator set of "unique forward plus reverse" (n=23096) (see main text).

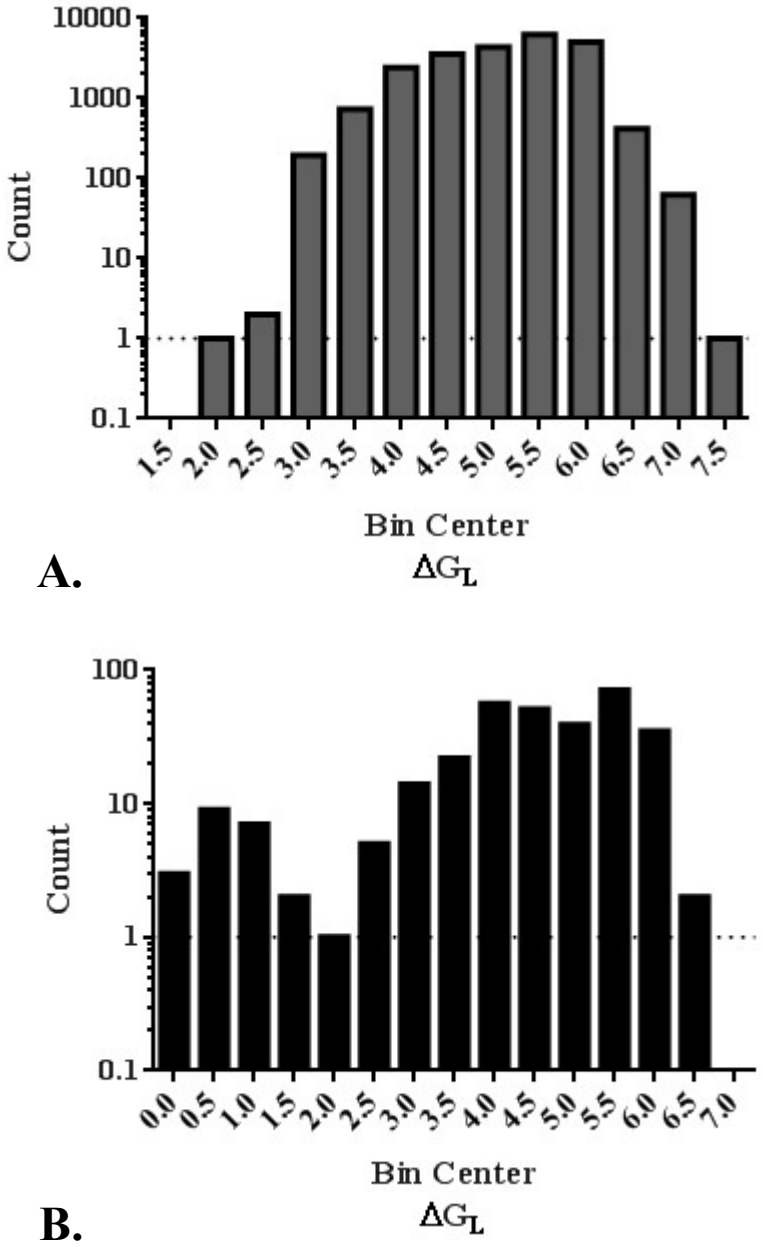


Figure 5.7 Distribution of folding energy of the loop (ΔG_L)

Distribution of folding energy of the loop (ΔG_L) of the putative terminator set of "unique forward plus reverse" (n=23096) (**A**), compared to that of the 'natural' set (n=317) from Chen *et al.* (2013)[148] (**B**).

The total number of non-duplicated terminators with a recognizable hairpin secondary structure in both directions was 10668.

We then sorted the scored putative terminators in descending order based on two parameters – the Reverse Strand and Forward Strand Predicted Ts. The Reverse and Forward Strand Predicted Ts scores were compared and the terminators sorted based on the Predicted Ts of the lower scoring strand. This resulted in the terminators being scored based on predicted bidirectionality – with the "least weak in either direction" terminators scored highest. From a total of 10667 tested, 415 terminators were found to have Predicted Ts for both the Forward and Reverse Strand of above 10 (TE of above 90%). We referred to this set of terminators as Predicted Strong Putative Bidirectional. Unsurprisingly, we found that these putative terminators exhibited symmetry in their sequence. In particular, the A-tract and U-tract were well conserved (Figure 5.8).

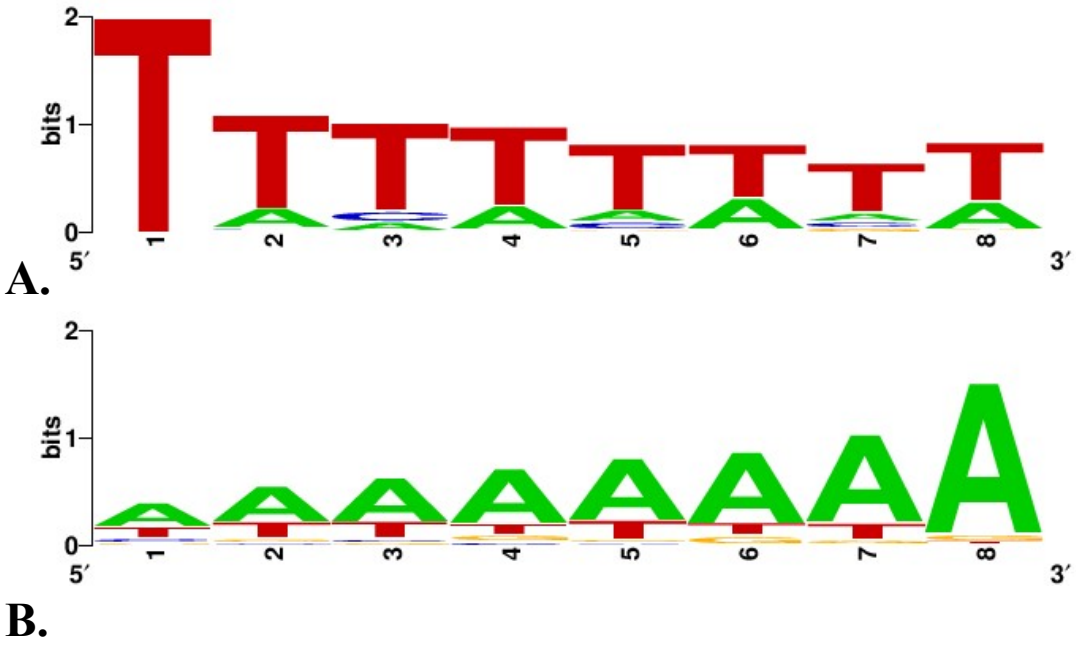


Figure 5. 8 Sequence logos of T-tract encoding the U-tract and the A-tract.

The U-tract (**A**) and the A-tract (**B**) of 415 putative terminators that were predicted to have strong (Termination Efficiency>90%) complementary strand terminators. The overall height of the stack represents sequence conservation at that position in bits whereas the height of the nucleotide represents the frequency of the nucleotide at the respective position. The starting position (1) is the first position in the U-tracts or A-

tracts of the putative terminators. The images were generated using WebLogo v2.8.2 [315,316].

While the predicted bidirectional set was not scored on maximal strength, the highest ranking predicted strong bidirectional terminator was 18th in absolute highest predicted score in either direction.

The 20 top scoring terminators of the Predicted Strong Putative Bidirectional set were chosen for synthesis in their predicted stronger orientation (**Table 5.5**), forming the library of terminators named *Predicted*. Interestingly the positions of the top 20 were analyzed (**Table 5.6**) and 60% of them were determined to be derived from convergently transcribed gene regions which is biologically relevant. Several terminators of the set were derived from genes coding for central metabolic enzymes (similar to the subjective strategy of terminator selection previously mentioned). Interestingly, the *C.acetobutylicum hydrogenase* gene T*hydA_Cac* terminator which was recently submitted as a BioBrick – BBa_K2715014 by others independently of the present study, was found in the top 20 list at position 2. However, as shown later in this chapter, this terminator was weaker than 70% in *E. coli* and even weaker in *B. subtilis* (under 40%).

Table 5. 5. Sequence and features of the *Predicted* set of terminators. Minimum free energy (ΔG) was calculated using RNAfold[304]. Italicized, underlined and bold nucleotide in T41_Csbu_TERM_674 was found to be deleted in T41m_Csbu_TERM_674 which was used for subsequent work (see main text).

Terminator Name	Length (bp)	Sequence (5'-3')	ΔG (kcal/mol)	Source Organism	Predicted Ts – synthesized	Predicted Ts reverse of synthesized
T35_Cac_TERM_3401	56	ATATTTAATTAAAAAGAAGCTGTCCTACTCTGGACAGCTTCTTTTTATGCAACTTA	-24.10	*C. acetobutylicum* ATCC 824	27.99	19.95
T36_Cac_TERM_42	81	GAGTATGGAGTAAAATAGTAAAATAAATGTGCCTCAACTTAGATGTTAAGGCACATTTATTTTATATATTATTCATGTTTT	-27.30	*C. acetobutylicum* ATCC 824	25.83	18.85
T37_Cspba_TERM_560	88	TTTTGTAACAGCTCATTTTTTTATATAATTAATGTGAGTTCTTAATTGAATTCACATTAATTATATAAGATAATATATTTGTATTAAG	-26.70	*C. saccharoperbutyl-acetonicum* N1-4 (HMT)	25.22	18.52
T38_Cbeij_TERM_4660	64	ATTAATTATCACATACATAAAAATGCCCCTTTATTGGGGCATTTTTATATTTAACAAAGGTCAT	-18.70	*C. beijerinckii* NCIMB 8052	26.85	18.45
T39_Csbu_TERM_1093	86	ATTTAAAATATATTTTAGAAAATAAAAGTACTCTTATTGCTACATTGCAATAAGAGTACTTTTATTTTCGGTTAGTTATAATATAA	-31.30	*C. saccharobutylicum* DSM 13864	24.49	18.13
T40_Csbu_TERM_20	74	TTCTTGTTAATATAATTAATAAAGGGATATATCTTTAAAATGATATATCCCTTTATTTATAGAATAAAAAATTC	-21.70	*C. saccharobutylicum* DSM 13864	25.29	18.12
T41_Csbu_TERM_674	52	TGACATTATCCTTAAGAAGCGCTAAAATTTGCGCTT***C***TTTCTCATTTGTATG	-14.30	*C. saccharobutylicum* DSM 13864	25.79	18.07
T42_Csbu_TERM_339	62	TTTTAATATCTATAAGAAGAGGCTATTTCATTTGAAATAGCCTCTTCTTATAGATAAACAAG	-34.30	*C. saccharobutylicum* DSM 13864	17.88	17.88
T43_Cac_TERM_3875	80	TGATAAAATCTCGTAATTATATAAAGAAAAAGCAAGAGAACTCTCTTGCTTTTTCTTTATATAGCTTCATATCCTATGGC	-27.10	*C. acetobutylicum* ATCC 824	17.75	17.75
T44_Csbu_TERM_2221	62	TTTCCTTTACAAAAATGAAGACTCTAGTGAAATCACTAAAGTCTTCATTTTTAAATGTGTAT	-20.10	*C. saccharobutylicum* DSM 13864	17.75	17.75

T45_Cac_TERM_1157	78	TAAAAATAAGAACCTATTAAAATAAATGGGC GCTGCTTTAATGCAGTGTCCATTTATTTTAA GTTTAAATAGAAACGT	-28.00	*C. acetobutylicum* ATCC 824	26.35	17.67
T46_Cspba_TERM_850	76	CCAAATTTACAAGGAGTTATAAAGTTAAACC ACTACGTTTGTAGTGGTTTAACTTTATAATA ATGAATAAGCTTAT	-26.90	*C. saccharoperbutyl-acetonicum* N1-4 (HMT)	17.65	17.65
T47_CA_plas_TERM_14	69	TTGGATGATGATGAATGATAAAAAAATAGGA CCTTTTGGTCCTATTTTTTTATATAAACTTA TATAATT	-19.60	*C. acetobutylicum* ATCC 824 pSOL1	24.54	17.61
T48_Cspba_TERM_1466	72	TATTGGTTAATAAAATAGGTACCTTAAAACT ACTATCTTGTAATTTTAAGATACCTATTTTG AAATTTACAA	-18.60	*C. saccharoperbutyl-acetonicum* N1-4 (HMT)	24.49	17.57
T49_Cspba_TERM_2009	56	CCATTTATGATAAAATAAGCACTAACTTTTG TTAGTGCTTATTTTATCATAAATAG	-30.50	*C. saccharoperbutyl-acetonicum* N1-4 (HMT)	24.31	17.49
T50_Cpa_TERM_1752	60	CTTTGGATAGAAGAATAAAGTGCATCTGTAA TTTGATGCACTTTATTTTTAATTTATTTA	-21.40	*C. pasteurianum* BC1	25.89	17.45
T51_Cac_TERM_293	56	TGCTGTCATAAAAAAGTACCAGCCATTAATT AGGCTGGTACTTTTTATCCATCATT	-21.10	*C. acetobutylicum* ATCC 824	18.84	17.41
T52_Cpa_TERM_1881	74	AATTTATTACCAAAATTTAAAAGTACATTCT TACTAAATTGTAAGAATGTACTTTTAAATTG TAAGGAAATAGA	-29.20	*C. pasteurianum* BC1	23.06	17.35
T53_Cbeij_TERM_773	74	AGGAACCATATAAAAGAAAGGATACTTAGTT GCTTGATATTGCAACTAAGTATCCTTTCTGT CTATTAACCGTT	-29.80	*C. beijerinckii* NCIMB 8052	21.25	17.24
T54_Csbu_TERM_2075	56	TTATCTATATAAAAATAAGGATGAACTATTG CTCATCCTTATTTTTATATTATAAT	-20.20	*C. saccharobutylicum* DSM 13864	23.76	17.21

Table 5.6 Locations of the 20 terminators from the *Predicted* set. The selected terminators are enriched for convergently transcribed regions on the *C. saccharoperbutylacetonicum* genome.

Terminator Name	Upstream CDS	Downstream CDS	Terminator Start	Terminator End	Convergent Genes
T35_Cac_TERM_3401	CA_C2944	CA_C2945	3077630	3077685	Yes
T36_Cac_TERM_42	PyrE CDS	HydA CDS	38033	38113	Yes
T37_Cspba_TERM_560	PurD CDS	Cspa_c11540 CDS (cysteine protease)	1280922	1281009	Yes
T38_Cbeij_TERM_4660	Cbei_3412 CDS (conserved hypothetical protein)	Cbei_3413 CDS (AraC protein, arabinose-binding/dimerisation)	3964772	3964835	Yes but 535 bp distance
T39_Csbu_TERM_1093	Cph CDS (phytochrome-like protein Cph2)	gltT2 CDS (proton/sodium-glutamate symport protein GltT)	2452997	2453082	Yes
T40_Csbu_TERM_20	CLSA_c00340 CDS (iron hydrogenase 1)	ykpA CDS (ABC transporter ATP-binding protein YkpA)	41199	41272	Yes
T41_Csbu_TERM_674	CLSA_c13640 CDS (coaBC: coenzyme A biosynthesis bifunctional protein CoaBC)	priA CDS (primosomal protein N')	1539232	1539283	No
T42_Csbu_TERM_339	TcrY CDS (sensor histidine kinase TcrY)	apbE2 CDS (thiamine biosynthesis lipoprotein ApbE)	777854	777915	No
T43_Cac_TERM_3875	CA_C3377	CA_C3378	3555038	3555117	Yes
T44_Csbu_TERM_2221	CLSA_c42890 CDS (RNA methyltransferase)	Nyk CDS	4814266	4814327	No
T45_Cac_TERM_1157	CA_C0885	CA_C0886	1014948	1015025	No

T46_Cspba_TERM_850	Cspa_c17710 CDS (Two component transcriptionalregulator, AraC family)	Cspa_c17720 CDS (cellobiose phosphorylase)	1960359	1960434	Yes
T47_CA_plas_TERM_14	CA_P0027 CDS	CA_P0028 CDS	29347	29415	Yes
T48_Cspba_TERM_1466	rpfG7 CDS	Cspa_c29200 CDS (arabinose efflux permease familyprotein)	3171648	3171719	No
T49_Cspba_TERM_2009	*trnF2* gene	Cspa_c40470 CDS (hypothetical protein)	4384820	4384875	Yes
T50_Cpa_TERM_1752	Clopa_3739 CDS	Clopa_3740 CDS	3760179	3760238	Yes
T51_Cac_TERM_293	CA_C0231	CA_C0230	262369	262424	No
T52_Cpa_TERM_1881	Clopa_3966 CDS (ribose 5-phosphate isomerase B)	Clopa_3965 CDS (ABC-type antimicrobial peptide transport system, ATPase component)	4000124	4000197	No
T53_Cbeij_TERM_773	Cbei_0575 CDS (Radical SAM N-terminal domain protein)	Cbei_0574 CDS (hypothetical protein)	684130	684203	Yes
T54_Csbu_TERM_2075	CLSA_c40300 CDS (hypothetical protein)	metQ CDS (D-methionine-binding lipoprotein MetQ)	4459233	4459288	No

5.1.4 Library construction

The terminators were synthesized as complimentary oligonucleotides with MoClo overhangs and assembled as described in Materials and Methods. Golden Gate assembly was used to insert the terminators in MoClo destination vector DVA_DE.

The terminators L3S2P00, L3S2P21 and L3S2P36 were synthesized with an extra A nucleotide between the 5' end and the 5'MoClo overhang in order to prevent the formation of a BsaI restriction site for the purpose of ensuring downstream Golden Gate compatibility. The sequences were designated L3S2P00m, L3S2P21m and L3S2P36m to denote the extra A.

Terminator T41 of the Predicted set was synthesized and cloned like the rest of the library, however, extensive screening by PCR, restriction digests and sequencing revealed numerous mutants containing various SNPs at different positions. While it is not clear whether this was of biological significance or due to poor quality oligonucleotides, a mutant was picked for further study – designated T41m. The mutant chosen has a single base deletion of a C nucleotide at position 37 (underlined, italicized and bold in ***Table 5.5***). Multiple attempts at isolating a correct clone of T41 in plasmid DVA_DE failed (screening of multiple colonies and assembly repeats)– each had one or more mutations (list of mutated sequences which were confirmed by sequencing - ***Table 5.7***), T41m was therefore chosen for further work as this was the clone with the least changes isolated. It is not clear whether this sequence presents a cloning challenge because of poor oligonucleotide quality, secondary structure preventing correct assembly or detrimental biological activity in the cloning host (see the promoter-like activity displayed by T41m in ***Table 5.11*** and ***Figure 5.12*** and discussed further in main text).

Table 5.7 Wild-type and isolated mutant sequences of T41_Csbu_TERM_674. The **-** symbol denotes gaps in the alignment of the sequences– in the wild-type this is due to an insertion in mutant clone #4, in all other mutants the symbol denotes deletions compared to wild-type. Insertions relative to the wild-type are displayed in italicized, underlined and bold font.

Terminator name	Sequence (5' to 3')
T41_Csbu_TERM_674 (WT)	TGACATTATCCTTAAGAAGCGCTAAAATTTGCGCTTCTTT-CTCATTTGTATG
T41m_Csbu_TERM_674 (mutant clone #1)	TGACATTATCCTTAAGAAGCGCTAAAATTTGCGCTT-TTT-CTCATTTGTATG
T41 mutant clone #2	TG-CATTATCCTTAAGAAGCGCTAAAATTTGCGCTTCTTT-CT-ATTTGTATG
T41 mutant clone #3	TG-CATTATCCTTAAGAAGCGCTAAAATTTGCGCTTCTTT-CTCATTTGTATG
T41 mutant clone #4	TGACATTATCCTTAAGAAGC-CTAAA--------TT-TTT***T***CTCATTTGTATG

Additionally, unpublished results from our laboratory indicate an inability to clone strong promoters without an accumulation of mutations in a DVA-derived vector in the default cloning-site orientation (as provided by the authors). In this cloning-site orientation, promoter-like activity is directed into the ColE1 origin and could generate transcripts reverse-complementary to the RNA II transcript which positively regulates replication[51]. RNA II is inhibited by RNA I - the native antisense regulator of this plasmid origin[51]. Thus, copy number may be drastically lowered or replication altogether inhibited. Non-replicating or lower than expected plasmid may fail to confer resistance at the required level by indirectly lowering the antiobiotic resistance protein expression levels. The DVA vector is made more susceptible by such read-through effects by the absence of a terminator between the ColE1 region and the cloning site.

The cloning of the terminators in DVA_DE was done using BpiI – there are two BpiI sites in T44_Csbu_TERM_2221 (T44) but cloning proceeded without issue. Likewise, the L3S2P36m (T3) terminator contains a single Esp3I restriction site but was sub-cloned successfully from DVA_DE into pLAC-B-GUS_GEN using Esp3I with the rest of the terminator library. The presence of two Type IIS sites in a naturally-occurring predicted terminator is intriguing as it bears similarity to synthetic cloning constructs such as ones used in this work. The BpiI recognition sites (which as typical Type IIS restriction enzyme recognition sites are not palindromic themselves) form an interrupted inverted repeat (interrupted palindrome) that corresponds to part of the predicted hairpin stemloop for T44. This arrangement is similar to the synthetic terminator insertions sites developed in this work – TIS v1 (BsaI sites), R-TIS v2 (BsaI sites) and B-TIS v2 (BsmBI/Esp3I sites). T44's naturally occurring BpiI restriction sites have their cut sites facing 'inward' whereas the synthetic TIS elements have theirs facing 'outward' (described in Section 4.2).

As mentioned, the library of 54 terminators was sub-cloned from DVA_DE into the dual enzymatic reporter of termination efficiency construct pLAC-B-GUS_GEN. Plasmids were verified by Sanger sequencing.

5.2 Terminator Library Characterisation in *E. coli*, *B. subtilis* and *C. saccharoperbutylacetonicum*

We began characterizing the sets of terminators first in *E. coli* and then in *B. subtilis* before selecting a final set of terminators for characterization in *C. saccharoperbutylacetonicum*. The reason to test for activity in *B. subtilis* was two-fold. First, the direct comparative analysis of terminator activity in the two model organisms using the present broad-host reporter construct would provide novel data on whether there are differences between terminator activity across species and taxa as has been documented for promoters[317]. Second, by including *B. subtilis* as a second member of the Firmicutes we were providing a much closer phylogenetic comparison to *Clostridium* than with *E. coli*, in a bacterial strain which is also widely used. Also, practically, *B. subtilis* has several advantages over Clostridia such as easily achievable transformation and robust aerobic growth. Its suitability as a proxy of Clostridia for terminator testing was also to be evaluated.

All 54 terminators were used for attempted transformation of *B. subtilis*, however, not all produced transformants that could be verified by colony PCR and gel electrophoresis. In the comparison panels these are listed as “N.D” for no data. Those were 10 total terminators T11, T12, T33, T34, T42, T45, T47, T50, T51 and T53. All of the PCR-verified transformants were attempted to be verified by a single round of Sanger sequencing of PCR-purified products (amplified with primers IG0130 and IG0081) with primer IG0130. Nevertheless, 7 terminator constructs were verified by

PCR-only due to poor quality reads –T2, T9, T15, T26, T27, T32 and T48. These constructs displayed activity and have been included in subsequent analysis. Terminators that were verified by Sanger sequencing but produced no activity in *B. subtilis* in either the upstream or the downstream reporter were excluded from analysis. These were 4 terminators - T6, T17, T19 and T21. In the following sections, data obtained in *B. subtilis* is presented for 40 terminator constructs (all verified by PCR and displaying activity), 33 of which have been fully verified.

5.2.1 Characterizing the *Natural* and *BioBrick* set of terminators in *E. coli* and *B. subtilis*

To assess this first set of well-used terminators, we grew *E. coli* DH5α and *B. subtilis* 168 cells and transformed them with the pLAC-GUS_GEN reporter constructs containing the *Natural* and *Biobricks* library set. All *E. coli* strains were grown, harvested and assayed and activity was calculated as described in Materials and Methods and Chapter 4. The data is summarized in ***Table 5.8*** and graphically presented in ***Figure 5.9***. In some cases the upstream reporter was not detected when assaying in *B. subtilis* and so data is missing for a small number of terminators. We also plotted the predicted and published terminator strengths for comparison (***Fig. 5.9***). The two terminators included for analysis in *E. coli* (ECK120033127 and ECK120010800) which had been selected as examples of weak and medium terminators based on their published Ts scores of 2.91 and 5.21, respectively (***Table 5.8***)[148]. These terminators turned out to have similar activity in this study, having average strengths of 2.13 and 1.71, respectively, in *E. coli* (***Table 5.8***). Data from *B. subtilis* for these two terminators is not available.

The terminators ECK120015170 and ECK120010783 both had higher measured strengths in the present system than previously reported, with the former's increase in

strength being about 1.7-fold (Ts=144) and the latter's about 7.27-fold (Ts=79.6). *B. subtilis* data for ECK120010783 is not available but ECK120015170 was weaker by large margin with a Ts of 30. Terminator ECK125108723 had a terminator strength that was nearly two-fold higher than the previously reported measurement, in *B. subtilis* this terminator was quite weak with a Ts of only 3.05.

The terminators ECK120033737 and ECK120029600, previously reported to be very strong with Ts over 300, were on average 5.6-fold and 14-fold weaker, the data from *B. subtilis* indicated even lower strengths of 10.5 and 16.5, respectively. It is worth noting that even with these weaker measurements, these terminators are ≥ 90% efficient. Notably, the values we measured are closer to the Predicted Ts.

Remarkably, the terminator BBa_B0062m, from the BioBrick collection (the reverse complement of the terminator from *rrnC*) which is used in pSB1C3-derived plasmids (pSB**3 series) plasmids as a MCS-flanking terminator and has a previously measured efficiency > 99% (Ts=111), did not have a measurable terminator activity (Fig. 5.9). This in fact agrees with its very low predicted strength of only 2.68. In this study the measured average Ts value was 0.87, which was also observed in *B. subtilis* (Ts of 0.847). To restate, a Ts of 1 is equivalent to TE=0%, therefore a Ts<1 corresponds to negative TE values and thus an increase in the expression of the downstream reporter relative to the control. This implies one or more of the following possibilities at the RNA level – increased transcription of the downstream portion of the mRNA driven by promoter-like activity from the putative terminator sequence, altered stability of the terminated transcript (decreased) or altered stability of the readthrough transcript (increased). A *de novo* promoter could also be generated at the junctions of the inserted terminator and the vector[151]. It is also possible that the putative terminator sequence increases the translation initiation rate of the downstream reporter gene from the

readthrough transcript. However, the large distance (115bp) between the terminator and the RBS of the downstream reporter sought to minimize this possibility (Figure 4.4).

The BBa_B0053m terminator is derived from the *E. coli his* operon and is also used within pSB1C3-derived plasmids as a MCS-flanking terminator. This terminator sequence was also tested in this study and measured to be very strong (Ts=35, TE=97%). It was, however, a lot weaker in *B. subtilis* (Ts of 6.3). To the author's knowledge, this is the first characterization of the vector-derived variant sequence of the *his* operon terminator. For comparison, the genomic copy had a measured strength of 3.27[148].

The BBa_B0010 (T1 loop of *rrnB*) terminator had an almost five-fold higher strength than previously reported, making it extremely strong in our reporter system (Ts of nearly 400). The measurement was similarly high in *B. subtilis* (however, data was only obtained from a single replicate in that organism). Consistent with these data, the strongest measured terminator in this set is BBa_B0015 with an average Ts score of 739, with a similarly high score in *B. subtilis.* This terminator is a dual terminator derived from BBa_B0010 (T1 loop of *rrnB*) and BBa_B0012 (T7 phage early terminator). These data suggest that this is a robust and very strong terminator that functions in both Gram-negative and Gram-positive bacteria.

Overall, the terminators tested in this set had terminator strength scores that significantly differed from the published scores. However, the two weakest terminators did not show a dramatic increase in apparent strength indicating that the effects observed are not artefacts of the system. Additionally, the scores did not differ in a linear manner, i.e. consistent overestimation or underestimation. The differences in measured termination efficiencies could arise from differences in the constructs used. Termination efficiency has been shown to vary with promoter strength[295]. To recall the design differences in the present study compared to other published work, the dual

enzymatic reporter system includes an insulator element in order to prevent interference with the translation of the downstream reporter by the strong secondary structures of the terminators. If previous measurements taken from constructs with shorter spacer regions between the terminator and the translation initiation region were affected by this, the measured efficiencies are expected to be lower if the spacing and thus a negative interaction is reduced. Additionally, the spacer between the upstream reporter's stop codon in the system described in the present study is 32bp (outside of inhibitory ribosomal footprint) whereas the reference data is from a construct with 9bp distance (well within the inhibitory ribosomal footprint). A potential relief of ribosomal repression would result in increased measured efficiencies. Constructs with intermediate design – our enzymatic reporter vector with the published intergenic spacers or the published reporter vector with the intergenic spacer regions from this study - could illuminate possible causes of measured efficiency discrepancies.

In the case of BBa_B0062m, possible explanations for the observed effect include the formation of a *de novo* spurious promoter at the interface of two genetic elements[151]. A spurious promoter could be increasing the downstream reporter levels counteracting terminator activity leading to the poor measured activity of the terminator part. It is also possible that neighbouring sequences are interacting with the terminator sequence in another way – such as influencing RNA secondary structure and reducing termination efficiency, *i.e.* spurious anti-terminator.

Several terminators also had marked differences in efficiency between *E. coli* and *B. subtilis*, and with the exception of the BioBrick BBa_B0015 (notably, containing the T1 loop of *rrnB*), all the measured terminators from this set had lower activity in *B.subtilis*. This trend could be of interest in the future development of terminators for use in *E.coli-B.subtilis* shuttle vectors.

Table 5.8 Set of natural terminators – terminator strengths measured and statistics. Data shown is mean of biological replicates (N=3) and technical replicates (N=2). SD – standard deviation to the mean of biological replicates ('-' denotes that data for only a single replicate was obtained). N.D. – no data or no transformants obtained in the duration of the study. N.D. – no data, no transformants obtained in the duration of the study. Data shown is rounded to 3 significant figures.

Terminator Name	Plasmid	*E. coli* Ts	E. coli Ts StDev	*E. coli* TE	E. coli TE StDev	*B. subtilis* Ts	*B. subtilis* Ts StDev	*B. subtilis* TE	*B. subtilis* TE StDev	Published Ts	Published Ts StDev	Published TE	Predicted Ts
ECK120033127	pLAC-GUS_T12	2.14	0.579	0.511	0.118	N.D.	N.D.	N.D.	N.D.	2.91	0.712	0.656	2.93
ECK120010800	pLAC-GUS_T11	1.71	0.163	0.411	0.0539	N.D.	N.D.	N.D.	N.D.	5.21	0.845	0.808	5.57
ECK120010783	pLAC-GUS_T17	79.6	33.4	0.986	0.0049	N.D.	N.D.	N.D.	N.D.	10.9	1.4	0.909	11.5
BBa_B0053m	pLAC-GUS_T27	35	18.8	0.966	0.0173	6.3	0.441	0.841	0.0112	N.D.	N.D.	N.D.	12.1
ECK125108723	pLAC-GUS_T29	43.5	15.2	0.975	0.0092	3.05	0.43	0.668	0.0434	21.3	6.05	0.953	15.4
BBa_B0010	pLAC-GUS_T30	400	65.8	0.998	0.0004	581	-	0.998	-	83.6	11.5	0.988	22.2
ECK120015170	pLAC-GUS_T15	144	44.7	0.993	0.0026	30	7.33	0.965	0.00791	85.8	6.71	0.988	99.6
BBa_B0062m	pLAC-GUS_T14	0.866	0.214	-0.198	0.265	0.847	0.136	-0.2	0.178	111	36	0.991	2.68
ECK120033737	pLAC-GUS_T25	55.9	25.2	0.979	0.0104	10.5	0.427	0.905	0.00387	313	117	0.997	15.1
ECK120029600	pLAC-GUS_T32	27	14.2	0.953	0.0296	16.5	5.78	0.933	0.0269	378	114	0.997	22.7
BBa_B0015	pLAC-GUS_T4	739	272	0.999	0.0007	968	420	0.999	0.00059	N.D.	N.D.	N.D.	N.D.

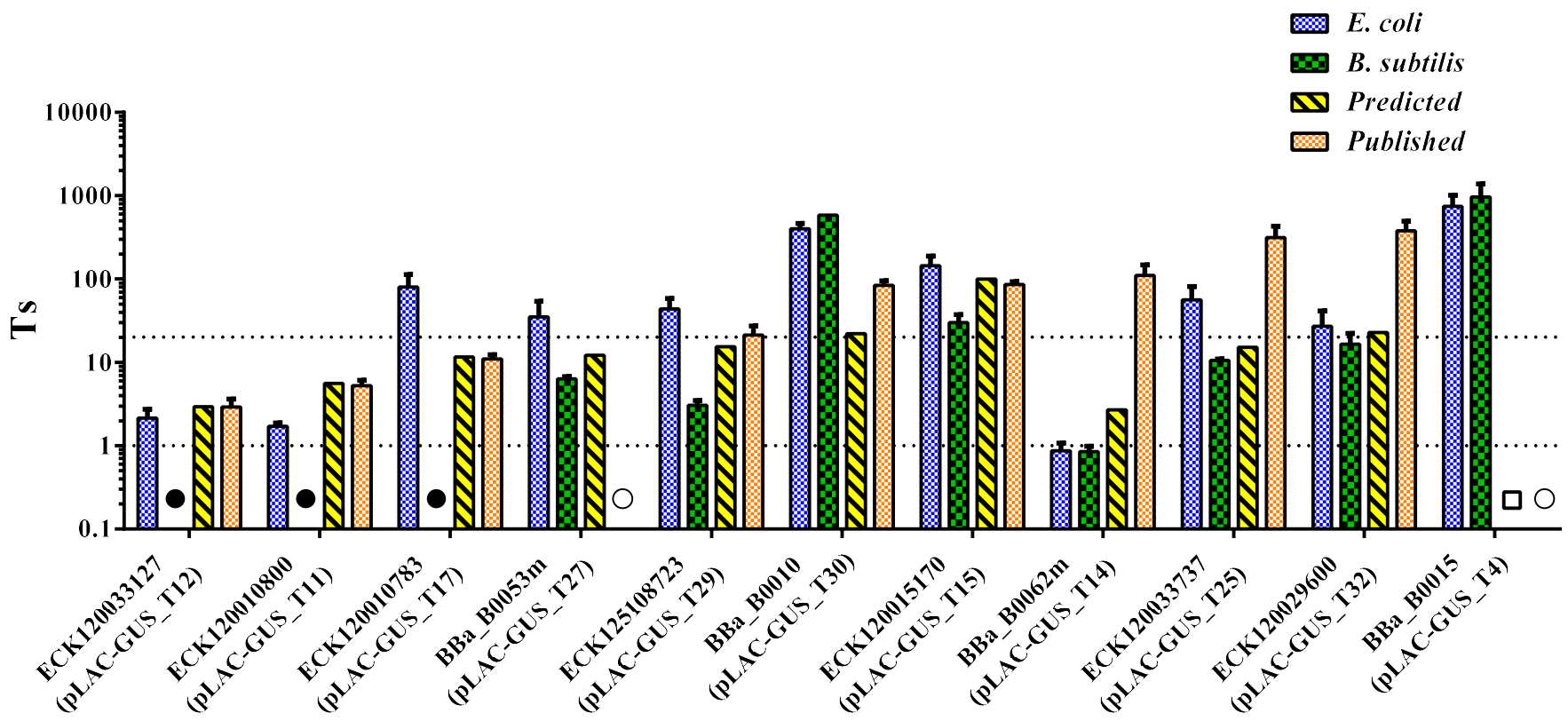


Figure 5. 9 Terminator strengths of the *Natural* set of terminators. Plotted data (*E. coli* and *B. subtilis*) represents mean of 6 measurements (biological replicates N=3 and technical replicates N=2), published *E. coli* data (N=3). Error bars - SD – standard deviation of the mean of biological replicates. The dotted lines represent Ts of 1 (TE = 0%) and Ts of 20 (TE = 95%) which mark the terminator cut-off(values below it display promoter-like activity) and the very strong terminator cut-off, respectively. Black circles indicate no data collected for *B.subtilis* condition. Open circles indicate published data not available or not shown. Open squares- predicted data not available.

5.2.2 Characterizing the *Synthetic* set of terminators in *E. coli* and *B. subtilis*

The *Synthetic* set of previously used terminators was transformed in *E. coli* DH5α and *B. subtilis* 168. Briefly, *E. coli* DH5α and *B. subtilis* cells were transformed with the pLAC-GUS_GEN reporter constructs containing this library set. The strains were grown, harvested and assayed, and activity was calculated as described in Materials and Methods and Chapter 4. The data is summarized in ***Table 5.9*** and graphically presented in ***Figure 5.10***.

Three weak terminators were tested (L1U1H09, L1U1H08, L2U2H09) and while L1U1H09 had a mild increase in strength it remained weak in in *E. coli* using the reporter system described in the present study. L1U1H08 and L2U2H09 were 12.6 and 2.5 times stronger than previously reported, respectively. Data for L1U1H08 in *B. subtilis* was not measurable due to lack of activity for the upstream reporter, however both L1U1H09 and L2U2H09 were much stronger in *B. subtilis* than in *E.* coli.

The measured strength of L3S2P36m was very close to that previously reported and the data also demonstrates that this is equally strong in *B. subtilis*.

Terminators L3S1P11 and L3S2P56 were the only two terminators from this set that were weaker in the measurement reported here than the published values. In *B. subtilis* their strength was markedly different; L3S1P11 was remarkably weak while L3S2P56 had slight increase in strength over that measured in *E. coli*. The L3S1P32 terminator was slightly stronger in *E. coli* in the construct described in the present study than the published data and in *B. subtilis* it had a substantial increase in strength.

Terminators L3S2P00m and L3S1P51 are two very strong terminators that measured slightly stronger in *E. coli* using the system described in the present study. Data isn't available for L3S1P11 in *B. subtilis*, however, L3S2P00m's Ts score in *B. subtilis* closely matched that from the literature for *E. coli*[148].

L3S2P21m was the strongest terminator in this set. Its *E. coli* strength in the current system was similar to the literature value while in *B. subtilis* it was slightly higher again, giving another robust terminator with the potential to work across multiple phyla of bacteria.

Most of the terminators in this set were measured to be stronger in the system described in the present study than in previous reports and were stronger still or maintained strength in *B. subtilis*. However, L3S1P11 was quite interesting as it was unexpectedly weak in *B. subtilis*. It should be noted again, that as previously mentioned, the LxSxPx terminators used here comprise two sub-groups of related sequences as they are derived from only two scaffolds and the three LxUxHx terminators chosen here share sequence similarities in the loop, hairpin and U-tract.

Table 5.9 Terminator strengths and accompanying statistical analyses for the *Synthetic* set of terminators. Data shown is mean of biological replicates (N=3) and technical replicates (N=2). SD – standard deviation to the mean of biological replicates. N.D. – no data or no transformants obtained in the duration of the study.N.D. – no data, no transformants obtained in the duration of the study. Data shown is rounded to 3 significant figures.

Terminator Name	Plasmid	*E. coli* Ts	E. coli Ts StDev	*E. coli* TE	E. coli TE StDev	*B. subtilis* Ts	*B. subtilis* Ts StDev	*B. subtilis* TE	*B. subtilis* TE StDev	Published Ts	Published Ts StDev	Published TE	Predicted Ts
L1U1H09	pLAC-GUS_T7	5.47	1.05	0.812	0.0404	40.7	5.59	0.975	0.00371	2.54	0.4	0.606	9.85
L1U1H08	pLAC-GUS_T6	47.1	21.1	0.976	0.0101	N.D.	N.D.	N.D.	N.D.	3.64	0.725	0.725	15.5
L2U2H09	pLAC-GUS_T5	13.3	6.43	0.907	0.0552	73.9	64.5	0.965	0.0418	5.38	2.28	0.814	6.73
L3S2P36m	pLAC-GUS_T3	41.9	18.1	0.973	0.0097	57.6	18.7	0.981	0.0071	37.4	4.42	0.973	27.4
L3S1P32	pLAC-GUS_T20	78.5	22.9	0.986	0.0044	287	222	0.995	0.00352	46.3	1.93	0.978	14.1
L3S1P11	pLAC-GUS_T18	28.9	4.71	0.965	0.0062	3.59	0.363	0.719	0.0296	108	15.5	0.991	41
L3S2P00m	pLAC-GUS_T1	193	70.2	0.994	0.0025	139	33.1	0.992	0.00208	127	21.5	0.992	133
L3S1P51	pLAC-GUS_T21	316	91.3	0.997	0.0011	N.D.	N.D.	N.D.	N.D.	237	96.2	0.996	17.1
L3S2P56	pLAC-GUS_T26	84.9	39.7	0.986	0.0065	109	16.6	0.991	0.0014	354	3.04	0.997	36.9
L3S2P21m	pLAC-GUS_T2	350	35.1	0.997	0.0003	488	150	0.998	0.00082	382	18.4	0.997	123

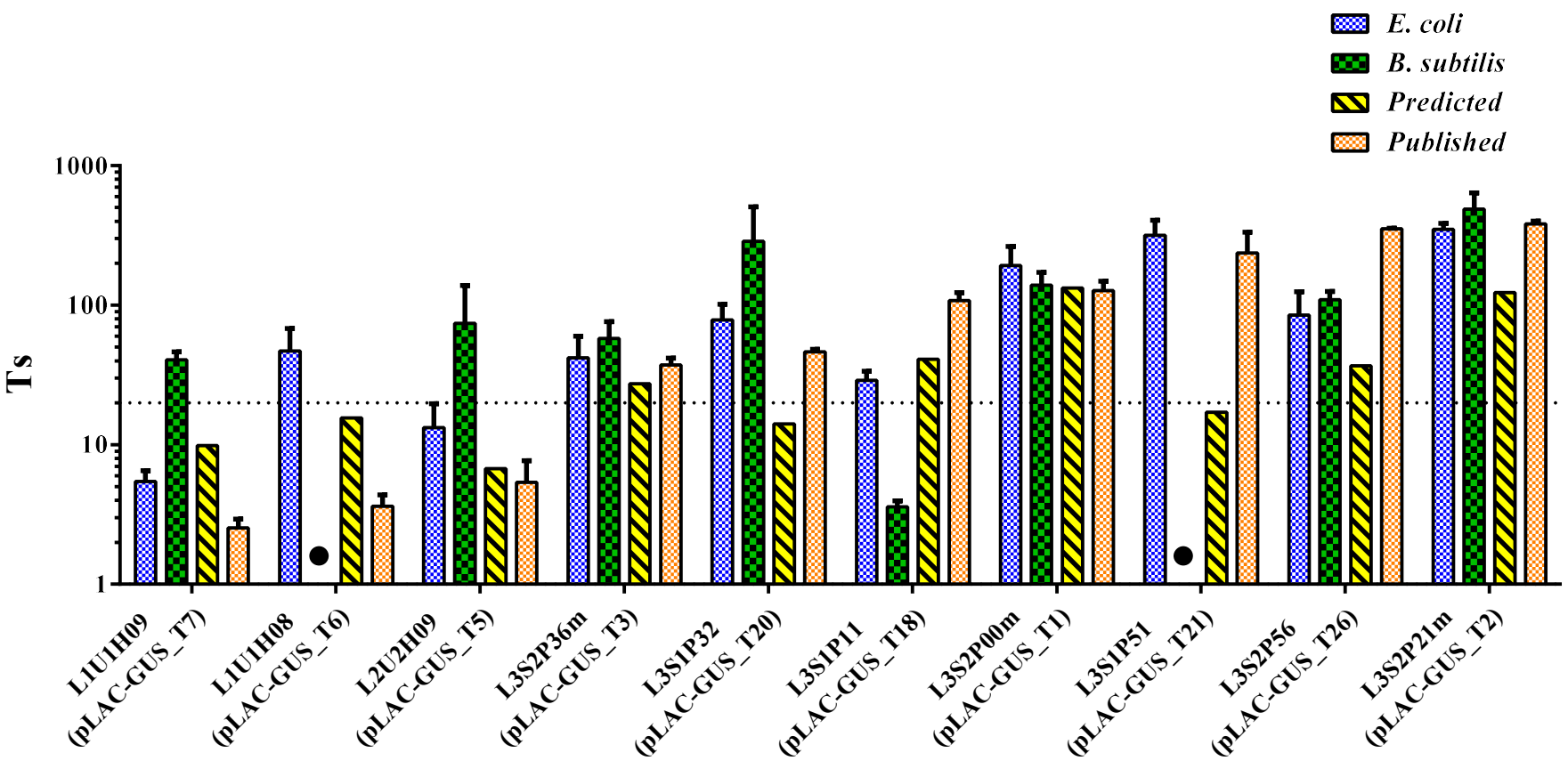


Figure 5. 10 Terminator strengths of the *Synthetic* set of terminators. Plotted data (*E. coli* and *B. subtilis*) represents mean of 6 measurements (biological replicates N=3 and technical replicates N=2), published *E. coli* data (N=3). Error bars - SD – standard deviation of the mean of biological replicates. The dotted lines represent Ts of 1 (TE = 0%) and Ts of 20 (TE = 95%) which mark the terminator cut-off(values below it display promoter-like activity) and the very strong terminator cut-off, respectively. Black circles indicate no data collected for *B.subtilis* condition.

5.2.3 Characterizing the *Firmicutes* set of terminators in *E. coli* and *B. subtilis*

The *Firmicutes* set of previously used terminators was transformed in *E. coli* DH5α and *B. subtilis* 168. Briefly, the strains were assayed as described previously in Materials and Methods and Chapter 4. The data produced is summarized in ***Table 5.10*** and graphically presented in ***Figure 5.9***.

The biophysical model was also applied to this set of terminators. The Pearson correlation coefficient *r* of Predicted Ts with measured *E. coli* Ts was *r*=0.364. The correlation coefficient of Predicted Ts with measured *B. subtilis* Ts was only *r*=-0.0519, while the correlation coefficient of *B. subtilis* Ts and *E. coli* Ts was *r*=0.393. There appears to be a trend of substantial terminator strength increase in the *B. subtilis* host over the *E. coli* measurement, most notably for T*fdx*, T*tyrS*, T*adc* and T*gyrA*.

Two of the putative terminators in this set displayed promoter-like activity (Ts<1) in *E. coli* (T_{CD0164} and T*thl_Cspba*), while no data was obtained for T*thl_Cspba* in *B. subtilis*, T_{CD0164} retained a promoter-like activity in *B. subtilis* albeit weaker.

The only terminators that were weaker in *B. subtilis* than in *E. coli* were T*slpA_Cdiff*, T*pepN*, and T_{phiTD1}. This trend was interesting as it could indicate clade-specific terminator adaptations, underlying the contrasting results with the *Firmicutes*, *Natural* and *Synthetic* terminator sets.

Table 5.10 Terminator strengths and accompanying statistical analyses for sequences derived from Firmicutes. Data shown is mean of biological replicates (N=3) and technical replicates (N=2). SD – standard deviation to the mean of biological replicates. N.D. – no data or no transformants obtained in the duration of the study.N.D. – no data, no transformants obtained in the duration of the study. Data shown is rounded to 3 significant figures.

Terminator Name	Plasmid	*E. coli* Ts	E. coli Ts StDev	*E. coli* TE	E. coli TE StDev	*B. subtilis* Ts	*B. subtilis* Ts StDev	*B. subtilis* TE	*B. subtilis* TE StDev	Predicted Ts
T*slpA_Lac*	pLAC-GUS_T8	24	11.9	0.948	0.0325	167	38.3	0.994	0.0015	26.9
T*fdx*	pLAC-GUS_T9	26.8	8.95	0.96	0.0117	742	299	0.998	0.00064	15.6
T*slpA_Cdiff*	pLAC-GUS_T10	8.66	1.09	0.883	0.0138	3.02	0.316	0.667	0.0331	13.4
T*tyrS*	pLAC-GUS_T13	6.95	1.8	0.849	0.0436	649	158	0.998	0.00038	5.76
T*thlA_Cac*	pLAC-GUS_T16	1.97	0.418	0.477	0.105	14.4	1.89	0.93	0.00973	28.2
T*slpA_Cdiff*_ext	pLAC-GUS_T19	12.2	5.92	0.906	0.0387	N.D.	N.D.	N.D.	N.D.	19.7
T*pepN*	pLAC-GUS_T22	5.44	1.07	0.811	0.038	5.03	3.44	0.744	0.126	6.09
T*gyrA*	pLAC-GUS_T23	38.1	7.49	0.973	0.0059	220	130	0.994	0.0032	6.68
T_{CD0164}	pLAC-GUS_T24	0.445	0.086	-1.3	0.398	1.05	0.241	0.0143	0.205	12.2
T_{phiTD1}	pLAC-GUS_T28	1.42	0.187	0.287	0.1	1.18	0.235	0.131	0.168	5.87
T*adc*	pLAC-GUS_T31	56.4	31.1	0.979	0.009	314	66.9	0.997	0.00075	25.3
T*thl_Cbeij*	pLAC-GUS_T33	16.5	4.03	0.937	0.0153	N.D.	N.D.	N.D.	N.D.	8.84
T*thl_Cspba*	pLAC-GUS_T34	0.436	0.19	-1.56	0.927	N.D.	N.D.	N.D.	N.D.	5.09

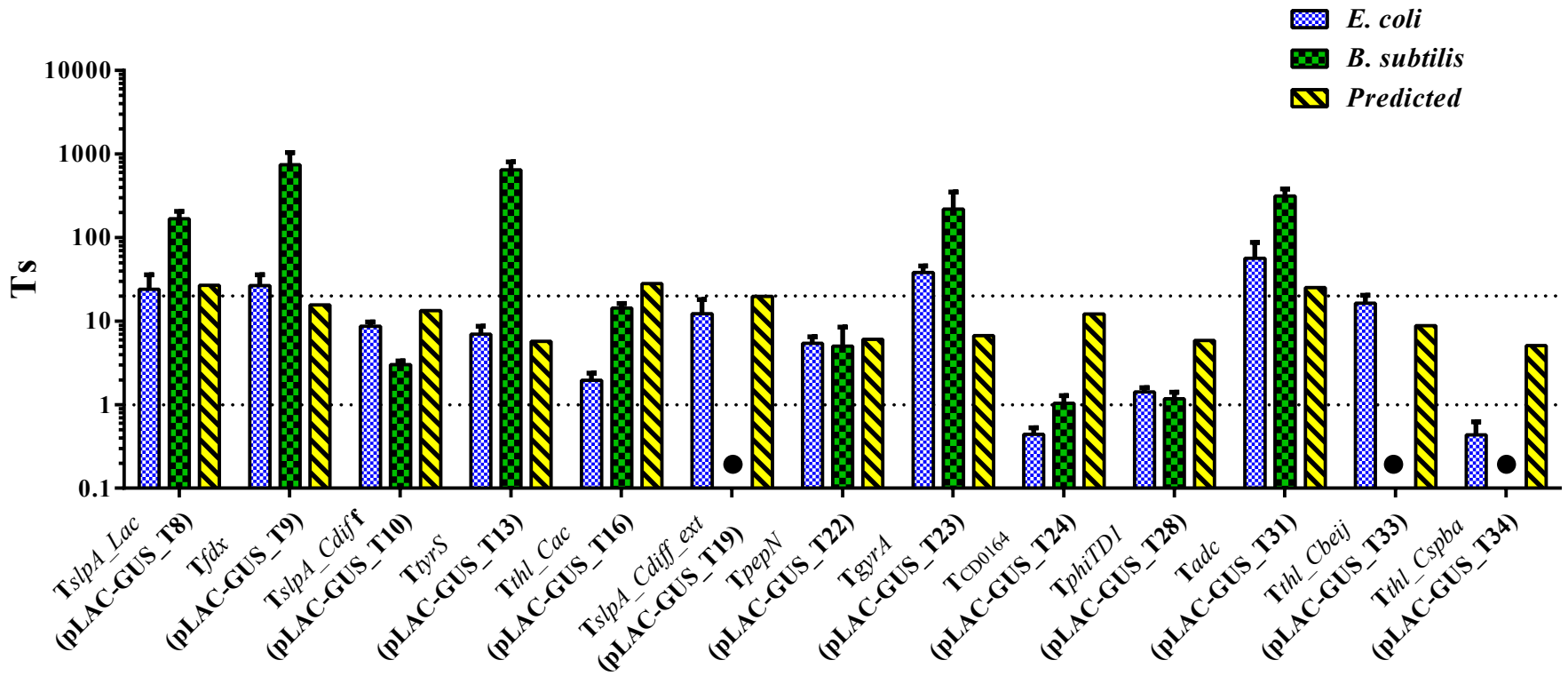


Figure 5. 11 Terminator strengths of the *Firmicutes* set of terminators. Plotted data represents mean of 6 measurements (biological replicates N=3 and technical replicates N=2). Error bars - SD – standard deviation of the mean of biological replicates. The dotted lines represent Ts of 1 (TE = 0%) and Ts of 20 (TE = 95%) which mark the terminator cut-off (values below it display promoter-like activity) and the very strong terminator cut-off, respectively. Black circles indicate no data collected for *B.subtilis* condition.

5.2.4 Characterizing the *Predicted* set of terminators in *E. coli* and *B. subtilis*

The *Predicted* set of previously used terminators was transformed in *E. coli* DH5α and *B. subtilis* 168. Briefly, the strains were assayed as described previously in Materials and Methods and Chapter 4. The data produced is summarized in ***Table 5.11*** and graphically presented in ***Figure 5.10***.

Due to the selection criteria used to define the *Predicted* set, the predicted strengths are consistently high across the full set (all have predicted Ts >17), although there are very large differences in the actual experimental values for this set. The most striking measurement is that for T35 in *B. subtilis*, which was extremely strong – in fact the strongest measured in the whole study in any of the organisms used. More generally, none of the terminators function particularly well in *E. coli* and in fact none were higher than the predicted Ts's, with only four terminators (T35, T38, T40 and T42) having Ts values over 10 (TE > 90%).

As well as T35 being very strong in *B. subtilis*, the two other very strong *B. subtilis* terminators (T39 and T49) functioned poorly in *E. coli*.

Surprisingly, T36, which encompasses the T*hydA_Cac* terminator (includes flanking sequence) was weak in both *E. coli* and *B. subtilis*. Terminator T41m (for the cloning history of T41m see Section 5.1.4 and ***Table 5.7***) displayed very high downstream reporter activity, indicating the presence of a strong promoter-like activity. The TE calculation for T41m produced a score of -9.26 instead of a positive number and fraction of 1 (indicative of terminator-like activity). A negative TE score (equivalent to a Ts below 1) is only informative insofar as it highlights instances of promoter-like activity in the data set. T41m had a 10.62 times larger ratio of DW/UP reporter activity than the no-terminator control. In addition, T37, T48, T49 and T54 also display

promoter-like activity in *E. coli*. In *B. subtilis* T37 and T48 retained promoter-like activity. T36 and T54 became weak terminators in *B. subtilis*. T49 on the other hand was very strong in *B. subtilis* (Ts=108).

When the entire *Predicted* set of terminators is considered, the correlation coefficient of Predicted Ts with the measured *E. coli* Ts was r=0.153. However, when the 5 putative promoter-like terminators are removed the correlation coefficient increases to r=0.328 which is similar to the correlation for the *Firmicutes* set (in that set when the 2 putative promoters are removed the correlation coefficient drops to r=0.298 from 0.364).

Overall, 25% of this set can be considered strong in *E. coli* (Ts > 10), one of these terminators (T35) and 3 others have a Ts > 20 (very strong) in *B. subtilis* – 25% of total. Therefore, the success rate of strong terminator discovery is considered satisfactory; however, interestingly the candidates that performed desirably in both organisms were a small minority, highlighting the necessity of *in vivo* characterization.

Terminators like T49 may have potential for alternative synthetic biology applications, in that they could offer species-specific constitutive repression. For example, in *E. coli* this element would not function as a terminator, but would be a very strong one in *B. subtilis*. This would allow the expression of a gene in *E. coli* but its repression in *B. subtilis* if the gene is placed under the control of a species-specific terminator. Likewise, terminators L3S1P11 (T18) and ECK125108723 (T29) from the previously described sets appear to be a functional opposite of T49 and thus could be used to keep a gene off in *E. coli* but allow its expression in *B. subtilis*.

Table 5.11 Terminator strengths and accompanying statistical analyses for sequences derived from *Clostridium*. Data shown is mean of biological replicates (N=3) and technical replicates (N=2). SD – standard deviation to the mean of biological replicates. N.D. – no data or no transformants obtained in the duration of the study. Data shown is rounded to 3 significant figures.

Plasmid	*E. coli* Ts	E. coli Ts SD	*E. coli* TE	E. coli TE SD	*B. subtilis* Ts	*B. subtilis* Ts SD	*B. subtilis* TE	*B. subtilis* TE SD	Predicted
pLAC-GUS_T35	16.4	10.3	0.924	0.0359	1590	218	0.999	0.0000945	28
pLAC-GUS_T36	3.48	1.43	0.678	0.129	1.56	0.334	0.338	0.151	25.8
pLAC-GUS_T37	0.242	0.189	-4.79	3.21	0.324	0.0669	-2.17	0.625	25.2
pLAC-GUS_T38	14.7	5.09	0.925	0.0327	10.5	0.857	0.904	0.00804	26.8
pLAC-GUS_T39	2.47	0.261	0.592	0.0422	122	59	0.991	0.0043	24.5
pLAC-GUS_T40	17.4	7.43	0.934	0.0305	4.05	0.607	0.75	0.0346	25.3
pLAC-GUS_T41m	0.272	0.318	-9.26	10.4	1.78	0.313	0.424	0.111	25.8
pLAC-GUS_T42	13.1	7.33	0.909	0.0384	N.D.	N.D.	N.D.	N.D.	17.9
pLAC-GUS_T43	3.97	2.21	0.692	0.157	8.58	0.988	0.882	0.0134	17.8
pLAC-GUS_T44	1.01	0.236	-0.0369	0.279	5.91	2.61	0.81	0.0735	17.8
pLAC-GUS_T45	3.82	1.33	0.718	0.0841	N.D.	N.D.	N.D.	N.D.	26.3
pLAC-GUS_T46	3.08	0.888	0.654	0.111	4.08	1.09	0.741	0.0813	17.6
pLAC-GUS_T47	1.29	0.314	0.192	0.21	N.D.	N.D.	N.D.	N.D.	24.5
pLAC-GUS_T48	0.381	0.064	-1.68	0.48	0.917	0.13	-0.106	0.17	24.5
pLAC-GUS_T49	0.938	0.111	-0.0769	0.131	108	48.4	0.99	0.00384	24.3
pLAC-GUS_T50	4.25	1.45	0.748	0.0733	N.D.	N.D.	N.D.	N.D.	25.9
pLAC-GUS_T51	6	3.22	0.803	0.0855	N.D.	N.D.	N.D.	N.D.	18.8
pLAC-GUS_T52	2.63	0.961	0.582	0.156	6.06	1.46	0.827	0.0465	23.1
pLAC-GUS_T53	3.73	0.935	0.718	0.0809	N.D.	N.D.	N.D.	N.D.	21.2
pLAC-GUS_T54	0.288	0.123	-3.14	2.35	1.11	0.28	0.0617	0.209	23.8

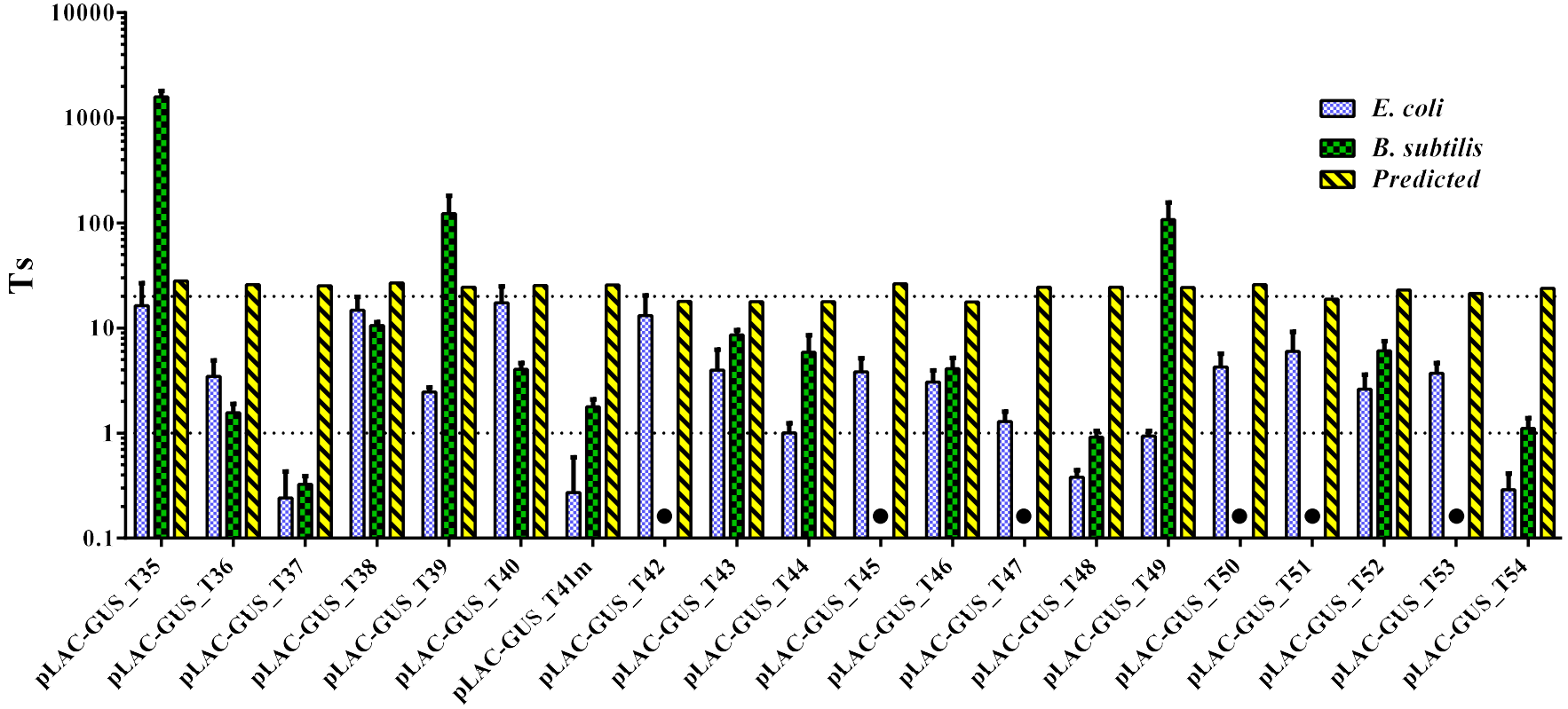


Figure 5. 12 Terminator strengths of the *Predicted* set of terminators. Plotted data represents mean of 6 measurements (biological replicates N=3 and technical replicates N=2). Error bars - SD – standard deviation of the mean of biological replicates. The dotted lines represent Ts of 1 (TE = 0%) and Ts of 20 (TE = 95%) which mark the terminator cut-off(values below it display promoter-like activity) and the very strong terminator cut-off, respectively. Black circles indicate no data collected for *B.subtilis* condition.

5.2.5 Characterizing the *Curated* set of terminators in *C. saccharoperbutylacetonicum*

Based on the results described earlier in this section a set of 12 terminators was selected for transformation in *C. saccharoperbutylacetonicum* N1-4(HMT). Technical difficulties in obtaining efficient robust transformations in York prevented a larger set being used as this methodology could only be done in the labs of Green Biologics.

Therefore, a set of terminators was assembled - this list is referred to as the *Curated* set. The *Curated* set contained two of terminators with differential activity in *E. coli* and *B. subtilis* - two terminators with high terminator strength in *E. coli* that were weak in *B. subtilis*- natural *E. coli*-derived ECK125108723 (T29) and synthetic L3S1P11 (T18), a single terminator with low promoter-like activity in *E. coli* that was strong in *B. subtilis* – T49 (from *Predicted* set) and a terminator with weak terminator strength in *E. coli* that was very strong in *B. subtilis* – T39 (from *Predicted* set). Four terminators that had moderate to high strength in *E. coli* but were very or extremely strong in *B. subtilis* were also picked – the *Firmicutes* terminators T*adc*, T*fdx*, T*tyrS* and the *Predicted* T35. Finally, three terminators were selected that were consistently strong and extremely strong in *E. coli* and *B. subtilis*. Those were the synthetic terminators L3S2P36m (T3) and L3S2P21m (T2) as well as the BioBrick double terminator BBa_B0015 (T4). Finally, the T*thl_Cspba* (T34) was also chosen as it was of interest due to its use in historical constructs produced at the Thomas Laboratory. Nine of the 12 terminators were successfully transformed into *C. saccharoperbutylacetonicum* N1-4 (HMT) – the T4, T39 and T49 were not transferred successfully during this study. Presence of the plasmid was verified by gel electrophoresis of PCR amplification with primers IG0130 and IG0081 after colony purification. Subsequently, the PCR products were purified and sequenced using the Sanger method with the IG0130 primer. Sanger sequencing

results confirmed the identity of the terminators, however, the sequence traces of T13 and T34 were not of

high quality and in the sequence trace of T35 there is an extra T at position 41 (relative to start of terminator as shown in ***Table 5.5***), potentially extending the U-tract, the activities from this sequence are reported here as T35m (pLAC-GUS_T35m) since the sequencing was done after activity was recorded. However, with a single sequencing reaction we cannot exclude an artifact introduced during PCR or an erroneous base call.

The strains were grown and assayed as described for *C. saccharoperbutylacetonicum* in Chapter 4 and Materials and Methods.

The control strain containing the reporter construct had high activity of the upstream reporter LacZ_Tts (***Figure 5.13, Table 5.12***). The ten terminator constructs can be separated into several sub-groups based on the level of upstream reporter activity detected. Terminators L3S2P00m (T2), L3S2P36m (T3), T*fdx* (T9) and T35m have high β-galactosidase activity similar to pLAC-X-GUS_GEN (***Table 5.12***). Terminators T*tyrS* (T13) and ECK125108723 (T29) have slightly lower β -galactosidase activity (***Table 5.12***). Terminator L3S1P11 (T18) has an intermediate level of β -galactosidase activity while T*adc* (T31) had similar levels of low but detectable β -galactosidase activity (***Table 5.1***2). Terminator T*thl_Cspba* (T34) had no detectable β -galactosidase activity above the level of background (***Table 5.12***) and since it did not have detectable β-glucuronidase activity either it was excluded from further analysis. Statistical comparisons between the upstream reporter activity and the negative control revealed that only T18, T31 and T34 did not exhibit a statistically significant increase (***Table 5.13***). On the other hand, a comparison for downstream reporter activity (against the negative control) resulted in significant increases confirmed only in the positive dual reporter control pLAC-X-GUS_GEN and ECK125108723 (T29) (***Table 5.14***).

Table 5.12 Activities of *C. saccharoperbutylacetonicum* strains.

β-GAL - β-galactosidase, β -GLUC - β-glucuronidase. Data shown is mean of 6 measurements (biological replicates N=3 and technical replicates N=2). SD – standard deviation. Data shown is rounded to 3 significant figures.

Construct name	β-GAL Activity (IU/OD_{600})	SD of β-GAL Activity	β-GLUC Activity (IU/OD_{600})	SD of β-GAL Activity
pEMPTY	0.01	0.0005	0.0085	0.0013
pMTL85151	0.0224	0.0215	0.0021	0.0035
pLAC-X-GUS_GEN	8.19	1.92	0.197	0.0909
L3S2P21m (pLAC-GUS_T2)	12.2	2.66	0.0145	0.0034
L3S2P36m (pLAC-GUS_T3)	9.5	1.87	0.0306	0.0111
T*fdx* (pLAC-GUS_T9)	18.3	2.66	0.0301	0.0223
TtyrS (pLAC-GUS_T13)	5.49	1.26	0.0156	0.0025
L3S1P11 (pLAC-GUS_T18)	0.834	0.43	0.0199	0.0061
ECK125108723 (pLAC-GUS_T29)	6.1	1.43	0.106	0.0254
T*adc* (pLAC-GUS_T31)	0.27	0.0668	0.0138	0.0052
T*thl_Cspba*(pLAC-GUS_T34)	0.0119	0.0026	0.0125	0.0065
T35m (pLAC-GUS_T35m)	15.4	3.79	0.0362	0.0196

Table 5.13 Summary of Dunett's multiple comparisons test after Ordinary one-way ANOVA.

Data for β-galactosidase activity from *C. saccharoperbutylacetonicum* strains (data summary in Table 5.12) were compared against the negative control pEMPTY. Mean Diff. – mean difference, CI of diff. – confidence interval.

Dunnett's multiple comparisons test	*Mean Diff.*	*95.00% CI of diff.*	*Below threshold?*	*Summary*	*Adjusted P- Value*
pEMPTY vs. pMTL85151	-0.01242	-4.412 to 4.387	No	ns	>0.9999
pEMPTY vs. pLAC-X-GUS_GEN	-8.180	-12.58 to -3.780	Yes	***	0.0001
pEMPTY vs. L3S2P21m (pLAC-GUS_T2)	-12.18	-16.58 to -7.777	Yes	****	<0.0001
pEMPTY vs. L3S2P36m (pLAC-GUS_T3)	-9.489	-13.89 to -5.089	Yes	****	<0.0001
pEMPTY vs. Tfdx (pLAC-GUS_T9)	-18.24	-22.64 to -13.84	Yes	****	<0.0001
pEMPTY vs. TtyrS (pLAC-GUS_T13)	-5.483	-9.883 to -1.084	Yes	**	0.0095
pEMPTY vs. L3S1P11 (pLAC-GUS_T18)	-0.8239	-5.223 to 3.575	No	ns	0.9969
pEMPTY vs. ECK125108723 (pLAC-GUS_T29)	-6.086	-10.49 to -1.686	Yes	**	0.0036
pEMPTY vs. Tadc (pLAC-GUS_T31)	-0.2603	-4.660 to 4.139	No	ns	0.9997
pEMPTY vs. Tthl_Cspba (pLAC-GUS_T34)	-0.001905	-4.401 to 4.397	No	ns	>0.9999
pEMPTY vs. T35m (pLAC-GUS_T35m)	-15.38	-19.78 to -10.98	Yes	****	<0.0001

Table 5.14 Summary of Dunett's multiple comparisons test after Ordinary one-way ANOVA.

Data for β-glucuronidase activity from *C. saccharoperbutylacetonicum* strains (data summary in Table 5.12) were compared against the negative control pEMPTY. Mean Diff. – mean difference, CI of diff. – confidence interval.

Dunnett's multiple comparisons test	*Mean Diff.*	*95.00% CI of diff.*	*Below threshold?*	*Summary*	*Adjusted P-Value*
pEMPTY vs. pMTL85151	0.006374	-0.06375 to 0.07649	No	ns	0.9996
pEMPTY vs. pLAC-X-GUS_GEN	-0.1881	-0.2582 to -0.1179	Yes	****	<0.0001
pEMPTY vs. L3S2P21m (pLAC-GUS_T2)	-0.005994	-0.07611 to 0.06413	No	ns	0.9997
pEMPTY vs. L3S2P36m (pLAC-GUS_T3)	-0.02206	-0.09218 to 0.04806	No	ns	0.9530
pEMPTY vs. Tfdx (pLAC-GUS_T9)	-0.02164	-0.09176 to 0.04848	No	ns	0.9579
pEMPTY vs. TtyrS (pLAC-GUS_T13)	-0.007076	-0.07720 to 0.06304	No	ns	0.9996
pEMPTY vs. L3S1P11 (pLAC-GUS_T18)	-0.01135	-0.08147 to 0.05877	No	ns	0.9993
pEMPTY vs. ECK125108723 (pLAC-GUS_T29)	-0.09714	-0.1673 to -0.02702	Yes	**	0.0035
pEMPTY vs. Tadc (pLAC-GUS_T31)	-0.005257	-0.07538 to 0.06486	No	ns	0.9997
pEMPTY vs. Tthl_Cspba (pLAC-GUS_T34)	-0.003958	-0.07408 to 0.06616	No	ns	0.9998
pEMPTY vs. T35m (pLAC-GUS_T35m)	-0.02765	-0.09777 to 0.04247	No	ns	0.8512

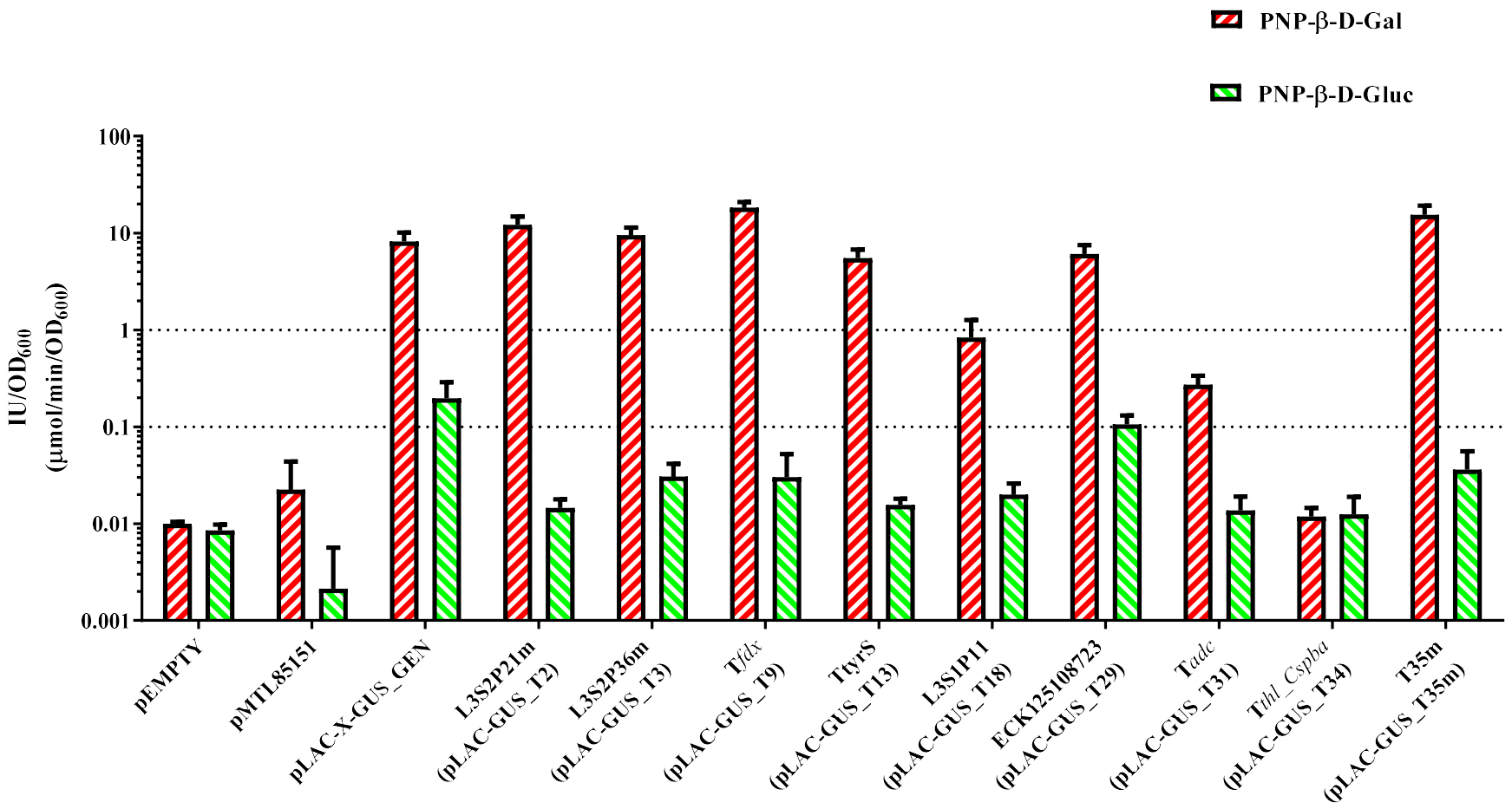


Figure 5. 13 Activities of *C. saccharoperbutylacetonicum* strains with terminator constructs.

Plotted values represent the mean of 6 measurements - biological replicates (N=3) and technical replicates (N=2), error bars represent standard deviation (SD). The dotted lines represent $IU/OD_{600} = 0.1$ and $IU/OD_{600} = 1$ and were added to enhance visual comprehension.

Terminator strength (Ts) and termination efficiency (TE) were calculated as described in Chapter 4. The results showed that several terminators were strong in the *Clostridium* host – terminators L3S2P21m, Tfdx, TtyrS and T35m had a Ts > 13 (TE > 92%) (***Figure 5.14*** and ***Table 5.15***). Surprisingly, terminator T*adc* which was very strong in *B. subtilis* did not have high Ts values in this organism. These calculations may be skewed by the very low upstream reporter activity detected from this construct. Terminators L3S1P11 and ECK125108723 were weak in *B.subtilis* and were also weak in *C. saccharoperbutylacetonicum*. Notably, ECK125108723 had high upstream reporter expression levels, discounting a false positive scenario (no expression from promoter). Terminator L3S2P36m, which was consistent across the previously tested species, had lower strength in *C. saccharoperbutylacetonicum*. Overall, 4 new strong terminators were added to the *Clostridium* toolbox in addition to a single moderately strong terminator (***Figure 5.14***). Demonstrating poor termination activity is also important as it once again highlights the necessity of further characterization and library expansion for Clostridia. Finally, interesting terminators such as ECK125108723 and L3S1P11 with differential activity may find applications in the regulation of expression of genetic devices in the three species studied.

Table 5.15 Terminator strength measured in *C. saccharoperbutylacetonicum.*

Data shown is mean of 6 measurements (biological replicates N=3 and technical replicates N=2). SD – standard deviation. T_S – terminator strength, TE – termination efficiency. Data shown is rounded to 3 significant figures.

Construct name	Ts	SD of Ts	TE	SD of TE
L3S2P21m (pLAC-GUS_T2)	32.6	14.9	0.963	0.0221
L3S2P36m (pLAC-GUS_T3)	8.84	0.724	0.886	0.0097
T*fdx* (pLAC-GUS_T9)	41.3	41.2	0.945	0.0552
TtyrS (pLAC-GUS_T13)	13.1	7.5	0.9	0.0643
L3S1P11 (pLAC-GUS_T18)	1.31	0.594	0.102	0.465
ECK125108723 (pLAC-GUS_T29)	1.36	0.365	0.219	0.242
T*adc* (pLAC-GUS_T31)	0.919	0.629	-0.533	1.06
T35m (pLAC-GUS_T35m)	13.2	7.53	0.909	0.0394

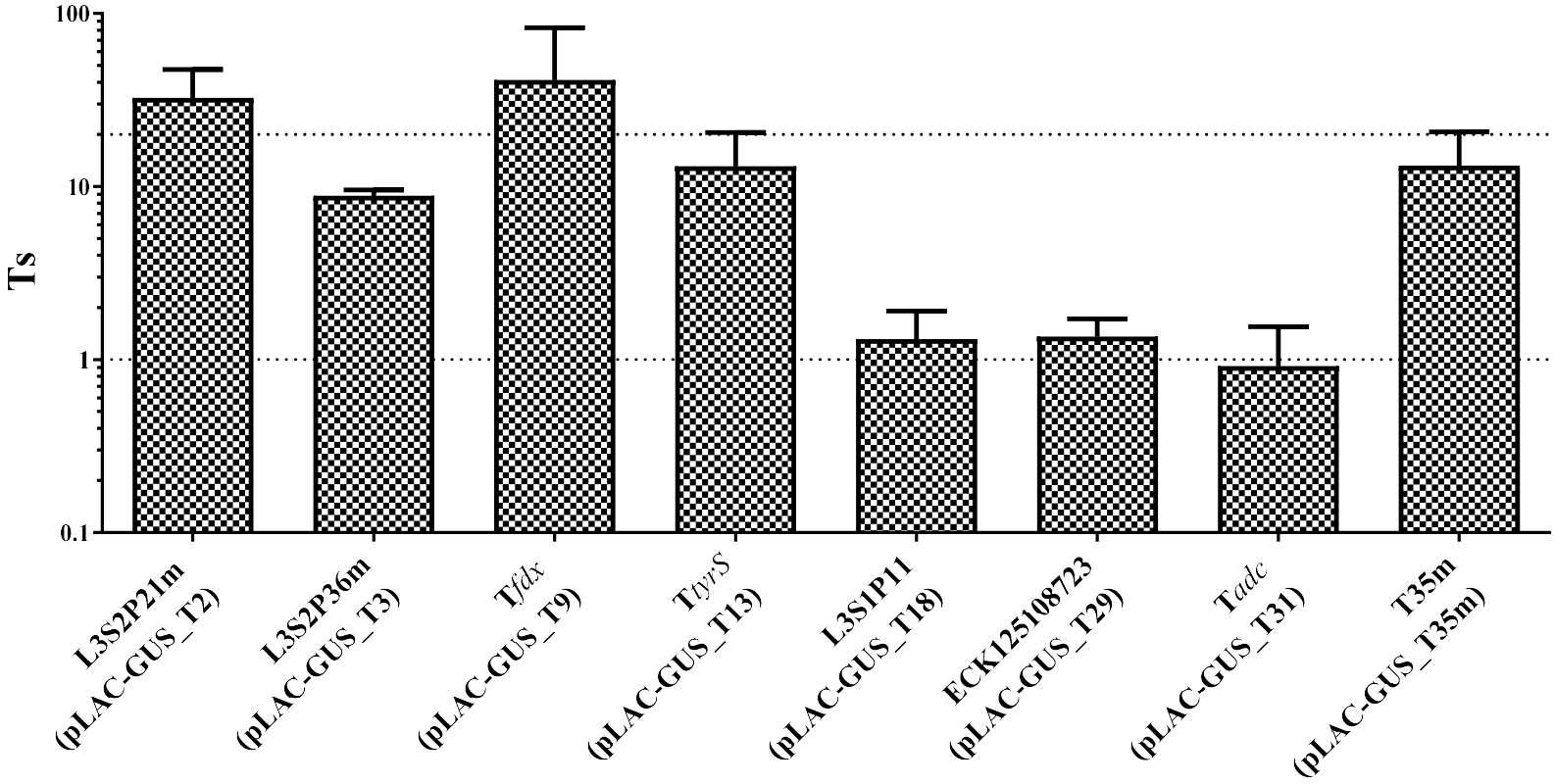


Figure 5.14 Terminator strength measured in *C. saccharoperbutylacetonicum.*

Plotted values represent the mean of biological replicates (N=3), error bars represent standard deviation (SD). The dotted lines represent T_S = 1 (TE = 0%) which marks the terminator cut-off (values below it display promoter-like activity) and T_S = 20 (TE = 97.5%) which marks the very strong terminator cut-off.

In conclusion, the terminators tested had a higher level of correlation between the measurements taken in the different species compared to the predicted terminator strength. This highlights the need for experimental validation of terminators. It is also noteworthy that we were able to discover a novel terminator (T35) that was strong in *E.coli* and very strong in *B. subtilis*. We inadvertently tested a variant T35m in *C. saccharoperbutylacetonicum* instead of T35 and found that it also was strong in *C. saccharoperbutylacetonicum*. Future work should involve comparisons of T35 and T35m in that organism. We also discovered interesting cases which had differential activity – such complex regulatory elements could be exploited as synthetic biology parts. For example, species-specific repression could be a promising alternative to inducible promoters by switching off the expression of genes encoding toxic gene products in the cloning strain (species), particularly for industrial strains in large-scale fermentations where induction can become costly.

Overall, the designed terminators from the synthetic set of terminators had similar activities in our system to the published values[148]. This indicates that they might be less-context dependent than the other terminators in our test set. Whereas the natural set chosen for testing here had more variable levels of termination compared to the published results. It is interesting whether terminators best explained by different models of termination (such as hybrid-shearing or forward translocation, see Section 1.8.2) have differing levels of context-dependence or whether they tend to be stronger in one taxonomic group of bacteria over another. However, further *in vitro* characterization might be necessary to discern between the most plausible termination model for each terminator.

As mentioned in the Introduction, species-specific de-repression functionality has already been developed for *C. difficile* using the P*tcdB* promoter and the specialized

genome-encoded TcdR sigma factor in *C. difficile*-transposon mutagenesis[121]. The system works by keeping the transposase gene expression off in the cloning host and having constitutive expression in the target organism. Using a species-specific terminator would provide a minimal functional equivalent to that system that would also be factor-independent and thus applicable to a broader range of Clostridia (which do not have TcdR) or Firmicutes. Since the diverging activities were conserved in *B. subtilis* and *C.saccharoperbutylacetonicum*, two distantly-related Firmicutes, perhaps a more appropriate name would be clade-specific terminator, however, further testing is warranted to establish the existence and utility of clade-specific terminators.

Chapter VI

Discussion

6. Discussion and Future Work

6.1 Flavin-binding fluorescent proteins – weak signal and complex regulation

The Flavin-binding fluorescent protein phiLOV2.1, although appearing to work in some published studies[68,104,108], proved difficult to detect in *Clostridium saccharoperbutylacetonicum* in this work. Its detectable fluorescence levels were also low in *E. coli*. To increase the activity of this reporter for future work, two areas are considered herein: limits to the expression level and the availability of the co-factor FMN.

Approaches to remedy poor expression should investigate transcript stability as well as the rate of translation and transcription. Interestingly, a recent study conducted with a Flavin mononucleotide-based fluorescent protein (FbFP) in *C. beijerinckii* using FACS to isolate a highly-fluorescent sub-population identified genetic changes in the expression plasmid that led to high activity[68]. A complex inverted repeat of the C-terminal coding sequence was found that did not contribute to the polypeptide chain as it was not predicted to be translated; however, it is likely that it formed extensive secondary structure at the 3' end. Additionally, also at the 3' end, a SNP mutation was found within the terminator of the construct. SNPs were also discovered in the replication protein RepL of the plasmid and hypothesized to contribute to increased copy number[68]. Thus, while vector copy number increases would result in higher levels of expression, it is possible that the transcript of that particular FbFP was unstable prior to the duplication due to degradation by processive exoribonucleases. Indeed, secondary structures such as hairpins at the 3'end have been demonstrated to improve transcript stability[192]. The FbFP-coding sequences are most often entirely synthetic and derived

through codon-optimization algorithms that do not account for RNA stability and exonuclease resistance.

While multiple codon-optimization algorithms have been developed in recent years, it is likely that until experimental mutagenesis and fine-tuning are performed in the target organism, translation will be sub-optimal. In the case of translation initiation, undesirable secondary structure at the 5' of the coding sequence can have negative effects on expression levels[168,303].

The second possibility – lack of available co-factor FMN - is more challenging to troubleshoot. The exogenous addition of co-factor precursors such as riboflavin is one possibility. Interestingly, *E. coli* is unable to import riboflavin and requires complementation with a heterologous transporter in order to do so[318]. The *Bacillus/Clostridium* species clade possess riboflavin biosynthesis operons and transporters, however expression appears tightly regulated[319,320]. In the case of *C. acetobutylicum* which has been used as a riboflavin producer strain[27], iron-limitation has been shown to induce riboflavin biosynthesis – in particular, the addition of a chelating agent - 2,2'-bipyridine[321]. A more recent report demonstrated that in the commonly used CGM medium (see Materials and Methods for composition) *C. acetobutylicum* does not produce detectable levels of riboflavin, however the addition of another chelating agent - 2-(4-morpholino)ethanesulfonic acid (MES) - induced riboflavin biosynthesis[27]. Subsequently, a PhD thesis reported that the use of a MES-buffered minimal medium resulted in increased FbFP-mediated fluorescence in *C. acetobutylicum* over CGM, addition of riboflavin to *C. acetobutylicum* expressing FbFPs resulted in increases in FbFP-mediated fluorescence[101]. However, they also detected increases in the autofluorescence levels of controls and of the supernatant of washed cells, attributed to riboflavin itself rather than FbFPs. Citing poor water-

solubility of riboflavin and a report demonstrating riboflavin production in *C. acetobutylicum*, the author discontinued its use. In this study, we did not attempt the addition of riboflavin or chelating agents to specifically induce to riboflavin production. However, we attempted to grow *C. saccharoperbutylacetonicum* in a Minimal *Clostridium* Medium (Materials and Methods) in order to minimize autofluorescence in the green channel which is known to be influenced by Flavin-rich yeast extracts (common ingredient in complex media)[279] as we had done for *E. coli* with C medium (Chapter 3). However, we did not observe fluorescence (data not shown) and we also found growth to be variable due to culture crashes (data not shown). Incidentally, the Minimal *Clostridium* Medium used in the present study included a chelating agent (sodium citrate) to prevent precipitation of metal ion additives. However, this specific chelating agent (citrate) increases the bioavailability of iron in the form of ferric citrate to species such as *E. coli*[322] and *B. subtilis*[323] and may have a similar function in *C. saccharoperbutylacetonicum*. Future experiments with FbFPs in *C. saccharoperbutylacetonicum* should test whether the addition of MES (or other chelators), riboflavin or combinations of the two to growth media increase FbFP-mediated fluorescence in this species.

Interestingly, in *B. subtilis*, FMN suppresses riboflavin biosynthesis but riboflavin itself does not[320]. Similarly, a study in *E. coli* demonstrated that expression of the FbFP PpSB2-LOV resulted in co-factor trapping of FMN and overproduction of FMN by the strain[324]. Free FMN has some overlap in the fluorescence emission profile with FbFPs and its overproduction could influence autofluorescence levels. If this is the case then the detected fluorescence would not directly correlate with the FbFP expression levels. Another possible and unexpected source of autofluorescence, particularly in Tet-inducible constructs, is the fluorescence of the TetR repressor protein bound to

tetracycline derivatives which fluoresces with excitation at 405nm and emission at 505nm in certain conditons[325]. Whether TetR actually contributes to *in vivo* fluorescence in FbFP-expressing constructs is currently unknown.

Overall, the tight regulation of the fluorescent proteins' cofactor and its central metabolic importance combined with their dimness and low photostability make the utility of FbFPs as genetic reporters somewhat tenuous.

6.2 mCherry and GFP-family fluorescent proteins – promising reporters

The GFP-family of fluorescent proteins is well established as a source of genetic reporters in a variety of organisms. We were able to detect mCherry fluorescence in the anaerobic *C. saccharoperbutylacetonicum*. While YFP, CFP and GFP have been expressed and detected in several *Clostridium* species (see Introduction), there are many more variants in this family some of which have improved pH-stability and much improved brightness over the variants tested so far. The mCherry protein would be particularly useful in characterizing the copy number and segregational stability of the different Gram-positive origins. We tested it with two replicons (pCD6 and pIM13) and found some observable differences. Though the aerobic fluorescence recovery (AFR) step kills the cell, it is possible to extract DNA for sequencing from paraformaldehyde fixed cells[326], therefore FACS is still a useful technique in conjunction with an mCherry reporter.

6.3 Room for improvement in the β-glucuronidase assay

We used the *E. coli* DH5α strain for testing of β-glucuronidase activity. This strain is a K-12 derivative and is not a knockout for *gusA* (*uidA*). However, the expression of the gene is tightly repressed. Moreover, the K-12 *gusB* transporter is an inactive mutant

thus most inducers and substrates of the operon are not imported[327]. However, this strain is not suitable for blue-white (or magenta-white) screening with substrates such as X-Gluc and Magenta-Gluc as these substrates are able to penetrate the cell and induce *gusA* expression[327]. This is similar to X-Gal which enters the cell in the absence of the permease LacY, however X-Gal is not an inducer of the *lac* operon. For blue-white screening, the BW25141 strain[328] would be a desirable host strain, alternatively a Δ*gusA* deletion mutant of another more commonly used strain could be constructed. While, X-Gluc and Magenta-Gluc require O_2 for color development and are not suitable for work with anaerobes, one solution could be the use of the replica plating technique and exposing one of the plates to O_2. Alternatively, it might be possible to use the 4-MUG substrate instead. The 4-MU product has a UV-inducible fluorescence and it may be possible to screen living colonies using a handheld black lamp. In *C. saccharoperbutylacetonicum* the existing background β-glucuronidase activity of the strain may be induced, thus limiting the applicability of the screen. However, the model solventogen *C. acetobutylicum* does not have background activity and would thus be a more suitable screening host[125,134].

The *E. coli* β-glucuronidase GusA_Ec has some measurable background activity towards the β-galactosidase substrate p-nitrophenyl- β-D-galactopyranoside[329]. However, engineered variants of this enzyme have been developed that are more specific and lack this non-specific activity[330]. Such variants could be used to construct an improved dual enzymatic reporter system.

6.4 Thermostable β-galactosidase

The thermostable β-galactosidase from *Thermoanaerobacterium thermosulfurigenes* - LacZ_Tts was shown to be sensitive to perturbations of its C-terminal region in this work and a published study[134]. The *E. coli* enzyme LacZ is very commonly used in α-

complementation studies where an N-terminally truncated enzyme's activity is restored by the expression of the N-terminal peptide sequence (the α-peptide). In addition to the commonly employed α-complementation, Ω-complementation, which involves supplying the C-terminal region of the *E. coli* enzyme *in trans*, has also been reported[331]. It is conceivable that the loss of activity in C-terminally truncated LacZ_Tts could be complemented in a similar fashion.

The substrate X-Gal, like the aforementioned X-Gluc, requires O_2 for color development. For use in anaerobes, a novel substrate – S-Gal – has been developed which develops a black color independently of O_2 and could be further assessed. As with β-glucuronidase, the background β-galactosidase activity may be induced by the addition of the substrate to the medium. *C. acetobutylicum* does not have background β-galactosidase activity either and would thus be a more suitable screening host.

Additionally, it is worth noting that while the original *lacZ_Tts* sequence was deposited in Genbank as M57579.1[281], the latter correction[134] did not result in a database update. Given the importance of this reporter in *Clostridium* studies, the author's intent is to deposit the fully corrected sequence reported here to a suitable sequence database (such as Genbank[332]) and the reporter vector pLAC-85151 to a physical DNA repository (such as Addgene[333]).

6.5 Dual Reporter System – Drawbacks and Future work

In this work, a significant amount of effort was put into creating a dual reporter, over a simpler single downstream reporter system. Future work includes the transfer of the dual enzymatic reporter system and the library to *C. acetobutylicum* in order to expand the terminator characterization into more solventogenic species. While the system has only been used for characterization so far, it has the potential to be used for screening of

terminator activity in suitable genetic backgrounds of *E. coli* (Δ*gusA* Δ*lacZ*), *B. subtilis* and possibly *C. acetobutylicum*. Remembering that the pIM13 origin functions in both *B. subtilis* and Clostridial sp., it is also likely to function in other Gram-positive species, thereby further expanding the use of the library. The genetic screen provides a high-throughput alternative for terminator selection to computational predictions and literature surveys.

To improve terminator characterization with the dual reporter system described in this study, the use of more sensitive substrates such as 4-MUG *in vitro* should be explored. Furthermore, β-galactosidase and β-glucuronidase substrates that release substantially different products could be used in a single reaction, thus simplifying sample processing.

In some cases, terminators measured with the system used in this study were calculated to have high termination efficiency, *i.e.* act as strong terminators. While these terminators were indeed the strongest tested, it is possible that the termination efficiency estimate would benefit from a more sensitive assay.

In addition, variations in the level of the upstream reporter (referred to here as the upstream effect or δTE) can influence the scores obtained. For example, increases in the upstream reporter expression can result in terminator strength overestimation; conversely, decreases in upstream reporter expression can result in strength underestimation. This is important in the cases where this variation in the upstream reporter is caused by the insertion of the terminator. However, to experimentally determine the upstream effect (δTE) with the measurements taken so far would be challenging because of fluctuations between measurements. A solution would be to measure a third reporter – a distal upstream reporter – which could be used to normalize

the data. A suitable distal upstream reporter in the system described in the present study is the CAT enzyme encoded by the resistance marker *catP*. CAT activity can be measured using a DTNB assay and following the release of TNB by monitoring the absorbance at 412nm. This system may work well in *E. coli* (unpublished work); however, *C. saccharoperbutylacetonicum* is likely to have a high background activity level because it is already resistant to intermediate doses of thiamphenicol. The mechanism of resistance has not been experimentally verified but the strain contains two genes predicted to code for CAT enzymes (Cspa_c32480, Cspa_c34740 (*catB*)), one or both the genes may be contributing to the observed resistance.

6.6 Effects of terminators

The aforementioned upstream effects of some terminators have been documented in bacterial species including *E. coli* and *C. acetobutylicum*. The effects we observe would have to be validated in one of several ways – direct measurement of transcript levels – Northern blot or qPCR combined with RNA half-life measurements after transcription inhibition with rifampicin (*C. saccharoperbutylacetonicum* is resistant) and also a change of sequence context – an obvious choice is to use the system with the alternative reporter order – GUS-LAC. This system has the added advantage in *C. saccharoperbutylacetonicum* of reduced downstream reporter background activity by heat treatment. We measured substantial increases in absolute β-galactosidase activity over the single reporter β-galactosidase construct in the control dual reporter pLAC-X-GUS and in several terminator constructs. It is of interest whether pGUS-X-LAC would have a similar increase in β-glucuronidase activity compared to its single reporter control construct. Further work is required to determine the cause of the observed activity increases.

6.7 Terminator strength prediction model and scoring

Terminator prediction algorithms produce lots of putative terminator hits from *Clostridium* data sets because *Clostridium* intrinsic terminators appear to be predominantly of the canonical structure and almost invariably have a recognizable U-tract and which is easily identifiable[308].

However, even though these terminators appear to be more conserved, terminator strength prediction models still perform less well. The model[148] that was applied in this work for prediction of terminator strength does not have high predictive power. For example, a maximal correlation coefficient of 0.364 of predicted terminator strength with measured strength in *E. coli* was obtained with the *Firmicutes* terminator set derived from the literature. The correlation with predicted strength was even lower for the *Predicted* set of terminators. The terminators sampled from the work of Chen *et al.* (2013)[148] study had higher agreement with predicted strength.

The alternative models such as the linear sequence-function model published by *Cambray* et al[313] achieved a high correlation coefficients on a small training data set that excluded context-confounding sequences surrounding the minimal terminator; however, when the set was expanded the correlation coefficient dropped. Accurate terminator strength prediction models are still a work in progress and may benefit from data in diverse bacterial species that could illuminate shared features and points of divergence between intrinsic terminators.

References

1. Moon, H. G. *et al.* One hundred years of clostridial butanol fermentation. *FEMS Microbiol. Lett.* **363**, fnw001 (2016).

2. Breed, R. S., Murray, E. G. D. & Smith, N. R. *Bergey's manual of determinative bacteriology*. (Williams & Wilkins Co., 1957).

3. Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiol. Rev.* **50**, 484–524 (1986).

4. Lepage, C., Fayolle, F., Hermann, M. & Vandecasteele, J. P. Changes in membrane lipid composition of *Clostridium acetobutylicum* during acetone-butanol fermentation: effects of solvents, growth temperature and pH. *J. Gen. Microbiol.* **133**, 103–110 (1987).

5. Baer, S. H., Blaschek, H. P. & Smith, T. L. Effect of butanol challenge and temperature on lipid composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **53**, 2854–2861 (1987).

6. Keis, S., Bennett, C. F., Ward, V. K. & Jones, D. T. Taxonomy and phylogeny of industrial solvent-producing clostridia. *Int. J. Syst. Bacteriol.* **45**, 693–705 (1995).

7. Keis, S., Shaheen, R. & Jones, D. T. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int. J. Syst. Evol. Microbiol.* **51**, 2095–2103 (2001).

8. Hellinger, E. *Clostridium aurantibutyricum* (n.sp.): a pink butyric acid

Clostridium. *J. Gen. Microbiol.* **1**, 203–210 (1947).

9. Holt, R. A., Cairns, A. J., Morris, J. G. & Morris, J. G. Production of butanol by *Clostridium puniceum* in batch and continuous culture. *Appl Microbiol Biotechnol* **27**, 319–324 (1988).

10. Abd-Alla, M. H., Zohri, A.-N. A., El-Enany, A.-W. E. & Ali, S. M. Acetone–butanol–ethanol production from substandard and surplus dates by Egyptian native *Clostridium* strains. *Anaerobe* **32**, 77–86 (2015).

11. Sarchami, T., Munch, G., Johnson, E., Kießlich, S. & Rehmann, L. A review of process-design challenges for industrial fermentation of butanol from crude glycerol by non-biphasic *Clostridium pasteurianum*. *Fermentation* **2**, 13 (2016).

12. Tashiro, Y., Yoshida, T., Noguchi, T. & Sonomoto, K. Recent advances and future prospects for increased butanol production by acetone-butanol-ethanol fermentation. *Eng. Life Sci.* **13**, 432–445 (2013).

13. Gheshlaghi, R., Scharer, J. M., Moo-Young, M. & Chou, C. P. Metabolic pathways of clostridia for producing butanol. *Biotechnol. Adv.* **27**, 764–781 (2009).

14. Poehlein, A. *et al.* Microbial solvent formation revisited by comparative genome analysis. *Biotechnol. Biofuels* **10**, 58 (2017).

15. Udaondo, Z., Duque, E. & Ramos, J. L. The pangenome of the genus *Clostridium*. *Environ. Microbiol.* (2017).

16. Shaheen, R., Shirley, M. & Jones, D. T. Comparative fermentation studies of industrial strains belonging to four species of solvent-producing clostridia. *J.*

Mol. Microbiol. Biotechnol. **2**, 115–124 (2000).

17. Jang, Y.-S. *et al.* Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *MBio* **3**, e00314-12 (2012).

18. Pyne, M. E. *et al.* Disruption of the reductive 1,3-propanediol pathway triggers production of 1,2-propanediol for sustained glycerol fermentation by *Clostridium pasteurianum*. *Appl. Environ. Microbiol.* AEM.01354-16 (2016).

19. Xiao, H. *et al.* Metabolic engineering of d-xylose pathway in *Clostridium beijerinckii* to optimize solvent production from xylose mother liquid. *Metab. Eng.* **14**, 569–578 (2012).

20. Xiao, H. *et al.* Confirmation and elimination of xylose metabolism bottlenecks in glucose phosphoenolpyruvate-dependent phosphotransferase system-deficient *Clostridium acetobutylicum* for simultaneous utilization of glucose, xylose, and arabinose. *Appl. Environ. Microbiol.* **77**, 7886–7895 (2011).

21. Mingardon, F., Chanal, A., Tardif, C. & Fierobe, H.-P. P. The issue of secretion in heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **77**, 2831–2838 (2011).

22. Chanal, A., Mingardon, F., Bauzan, M., Tardif, C. & Fierobe, H.-P. P. Scaffoldin modules serving as ‘cargo’ domains to promote the secretion of heterologous cellulosomal cellulases by *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **77**, 6277–6280 (2011).

23. Kovács, K. *et al.* Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824.

Biotechnol. Biofuels **6**, 117 (2013).

24. Willson, B. J. *et al.* Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* **9**, 109 (2016).

25. Collas, F. *et al.* Simultaneous production of isopropanol, butanol, ethanol and 2,3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains. *AMB Express* **2**, 45 (2012).

26. González-Pajuelo, M. *et al.* Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1,3-propanediol from glycerol. *Metab. Eng.* **7**, 329–336 (2005).

27. Cai, X. & Bennett, G. N. Improving the *Clostridium acetobutylicum* butanol fermentation by engineering the strain for co-production of riboflavin. *J. Ind. Microbiol. Biotechnol.* **38**, 1013–1025 (2011).

28. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).

29. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y).* **39**, 783–791 (1985).

30. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–9 (2011).

31. ABAS. Classification of prokaryotes (bacteria and archaea) into risk groups

(TRBA 466). *Bundesarbeitsblatt. http://www.baua.de* 19 (2015).

32. Johnson, J. L. & Chen, J.-S. Taxonomic relationships among strains of *Clostridium acetobutylicum* and other phenotypically similar organisms. *FEMS Microbiol. Rev.* **17**, 233–240 (1995).

33. Máté de Gérando, H. *et al.* Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. *BMC Genomics* **19**, 242 (2018).

34. Nölling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).

35. Wang, Y., Li, X., Mao, Y. & Blaschek, H. P. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. *BMC Genomics* **13**, 102 (2012).

36. Mermelstein, L. D., Welker, N. E., Bennett, G. N. & Papoutsakis, E. T. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Nat. Biotechnol.* **10**, 190–195 (1992).

37. Mermelstein, L. D. & Papoutsakis, E. T. *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **59**, 1077–1081 (1993).

38. Reid, S. J., Allcock, E. R., Jones, D. T. & Woods, D. R. Transformation of *Clostridium acetobutylicum* protoplasts with bacteriophage DNA. *Appl. Environ. Microbiol.* **45**, 305–307 (1983).

39. Oultram, J. D. *et al.* Introduction of genes for leucine biosynthesis from *Clostridium pasteurianum* into *C. acetobutylicum* by cointegrate conjugal transfer. *Mol. Gen. Genet. MGG* **214**, 177–179 (1988).

40. Oultram, J. D. *et al.* Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiol. Lett.* **56**, 83–88 (1988).

41. Reysset, G., Hubert, J., Podvin, L. & Sebald, M. Transfection and transformation of *Clostridium acetobutylicum* strain N1-4081 protoplasts. *Biotechnol. Tech.* **2**, 199–204 (1988).

42. Nakotte, S., Schaffer, S., Böhringer, M. & Dürre, P. Electroporation of, plasmid isolation from and plasmid conservation in *Clostridium acetobutylicum* DSM 792. *Appl. Microbiol. Biotechnol.* **50**, 564–567 (1998).

43. Herman, N. A. *et al.* Development of a high-efficiency transformation method and implementation of rational metabolic engineering for the industrial butanol hyperproducer *Clostridium saccharoperbutylacetonicum* strain N1-4. *Appl. Environ. Microbiol.* **83**, AEM.02942-16 (2017).

44. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnol. Biofuels* **6**, 50 (2013).

45. Grosse-Honebrink, A., Schwarz, K. M., Wang, H., Minton, N. P. & Zhang, Y. Improving gene transfer in *Clostridium pasteurianum* through the isolation of rare hypertransformable variants. *Anaerobe* **48**, 203–205 (2017).

46. Lesiak, J. M., Liebl, W. & Ehrenreich, A. Development of an *in vivo* methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis

of two endonuclease mutants. *J. Biotechnol.* **188**, 97–99 (2014).

47. Kolek, J., Sedlar, K., Provaznik, I. & Patakova, P. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnol. Biofuels* **9**, 14 (2016).

48. Oh, Y. H. *et al.* Optimized transformation of newly constructed *Escherichia coli-Clostridia* shuttle vectors into *Clostridium beijerinckii*. *Appl. Biochem. Biotechnol.* **177**, 226–236 (2015).

49. Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A. & Chou, C. P. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnol. Adv.* **32**, 623–641 (2014).

50. Minton, N. P. *et al.* A roadmap for gene system development in *Clostridium*. *Anaerobe* **41**, 104–112 (2016).

51. del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–64 (1998).

52. Lilly, J. & Camps, M. Mechanisms of theta plasmid replication. in *Plasmids* vol. 3 33–44 (ASM Press, 2015).

53. 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).

54. Projan, S. J., Monod, M., Narayanan, C. S. & Dubnau, D. Replication properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis*, and of its

close relative pE5, a plasmid native to *Staphylococcus aureus*. *J. Bacteriol.* **169**, 5131–9 (1987).

55. Garnier, T. & Cole, S. T. Identification and molecular genetic analysis of replication functions of the bacteriocinogenic plasmid plP404 from *Clostridium perfringens*. *Plasmid* **19**, 151–160 (1988).

56. 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**, 439–52 (2002).

57. Lee, S.-H. H. *et al.* A new shuttle plasmid that stably replicates in *Clostridium acetobutylicum*. *J. Microbiol. Biotechnol.* **25**, 1702–8 (2015).

58. 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).

59. Ehsaan, M. *et al.* Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* **9**, 4 (2016).

60. Li, Y., Blaschek, H. P. & Tan, D. Molecular characterization and utilization of the CAK1 filamentous viruslike particle derived from *Clostridium beijerinckii*. *J. Ind. Microbiol. Biotechnol.* **28**, 118–26 (2002).

61. Nakayama, S. *et al.* New host-vector system in solvent-producing *Clostridium saccharoperbutylacetonicum* strain N1-4. *J. Gen. Appl. Microbiol.* **53**, 53–6 (2007).

62. Maguin, E., Duwat, P., Hege, T., Ehrlich, D. & Gruss, A. New thermosensitive plasmid for gram-positive bacteria. *J. Bacteriol.* **174**, 5633–8 (1992).

63. Molitor, B., Kirchner, K., Henrich, A. W., Schmitz, S. & Rosenbaum, M. A. Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Sci. Rep.* **6**, (2016).

64. Lindow, J. C., Britton, R. A. & Grossman, A. D. Structural maintenance of chromosomes protein of *Bacillus subtilis* affects supercoiling *in vivo*. *J. Bacteriol.* (2002).

65. Snapper, S. B. *et al.* Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc. Natl. Acad. Sci. U. S. A.* (1988).

66. Panas, M. W. *et al.* Noncanonical SMC protein in *Mycobacterium smegmatis* restricts maintenance of *Mycobacterium fortuitum* plasmids. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13264–13271 (2014).

67. Azeddoug, H., Hubert, J. & Reysset, G. Stable inheritance of shuttle vectors based on plasmid pIM13 in a mutant strain of *Clostridium acetobutylicum*. *J. Gen. Microbiol.* **138**, 1371–1378 (1992).

68. Seo, S. O., Lu, T., Jin, Y. S. & Blaschek, H. P. Development of an oxygen-independent flavin mononucleotide-based fluorescent reporter system in *Clostridium beijerinckii* and its potential applications. *J. Biotechnol.* (2018).

69. Heap, J. T. *et al.* Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res.* **40**, e59 (2012).

70. Gu, P. *et al.* A rapid and reliable strategy for chromosomal integration of gene(s)

with multiple copies. *Sci. Rep.* **5**, 9684 (2015).

71. Kuhlman, T. E. & Cox, E. C. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38**, e92 (2010).

72. Bryant, J. A., Sellars, L. E., Busby, S. J. W. & Lee, D. J. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* **42**, 11383–11392 (2014).

73. Berger, M., Gerganova, V., Berger, P., Rapiteanu, R. & Lisicovas, V. Genes on a wire : the nucleoid- associated protein HU insulates transcription units in Escherichia coli. *Nat. Publ. Gr.* **6**, 1–12 (2016).

74. Sauer, C. *et al.* Effect of genome position on heterologous gene expression in *Bacillus subtilis*: an unbiased analysis. *ACS Synth. Biol.* (2016).

75. Anuchin, A. M., Goncharenko, A. V., Demidenok, O. I. & Kaprelyants, A. S. Histone-like proteins of bacteria. *Appl. Biochem. Microbiol.* **47**, 580–585 (2011).

76. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2009).

77. Li, Y. *et al.* Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).

78. Jiang, Y. *et al.* Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).

79. 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).

80. Croux, C. *et al.* Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. *Biotechnol. Biofuels* **9**, 23 (2016).

81. Yoo, M., Croux, C., Meynial-Salles, I. & Soucaille, P. Elucidation of the roles of *adhE1* and *adhE2* in the primary metabolism of *Clostridium acetobutylicum* by combining in-frame gene deletion and a quantitative system-scale approach. *Biotechnol. Biofuels* **9**, 92 (2016).

82. 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**, 8112–8121 (2012).

83. Zhang, N. *et al.* I-SceI-mediated scarless gene modification via allelic exchange in *Clostridium*. *J. Microbiol. Methods* **108**, 49–60 (2015).

84. Dong, H., Tao, W., Gong, F., Li, Y. & Zhang, Y. A functional *recT* gene for recombineering of *Clostridium*. *J. Biotechnol.* **173**, 65–67 (2014).

85. 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).

86. Wang, Y. *et al.* Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J. Biotechnol.* **200**, 1–5 (2015).

87. Li, Q. *et al.* CRISPR-based genome editing and expression control systems in

Clostridium acetobutylicum and *Clostridium beijerinckii*. *Biotechnol. J.* **11**, 961–972 (2016).

88. 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).

89. 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**, AEM.00233-17 (2017).

90. 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).

91. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005).

92. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12111–12116 (2003).

93. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).

94. Drepper, T. *et al.* Reporter proteins for *in vivo* fluorescence without oxygen. *Nat. Biotechnol.* **25**, 443–445 (2007).

95. Jean, D., Briolat, V. & Reysset, G. Oxidative stress response in *Clostridium*

perfringens. *Microbiology* **150**, 1649–1659 (2004).

96. 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).

97. Wakabayashi, Y. *et al.* An enhanced green fluorescence protein (EGFP)-based reporter assay for quantitative detection of sporulation in *Clostridium perfringens* SM101. *Int. J. Food Microbiol.* **291**, 144–150 (2019).

98. Ransom, E. M., Williams, K. B., Weiss, D. S. & Ellermeier, C. D. Identification and characterization of a gene cluster required for proper rod shape, cell division, and pathogenesis in *Clostridium difficile*. *J. Bacteriol.* **196**, 2290–300 (2014).

99. Ransom, E. M., Ellermeier, C. D. & Weiss, D. S. Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in *Clostridium difficile*. *Appl. Environ. Microbiol.* **81**, 1652–60 (2015).

100. Ransom, E. M., Weiss, D. S. & Ellermeier, C. D. Use of mCherryOpt fluorescent protein in *Clostridium difficile*. *Methods Mol. Biol.* **1476**, 53–67 (2016).

101. Schulz, F. Fluoreszenzproteine in *Clostridium acetobutylicum* - ein neues *in vivo* Reportersystem. (Universität Rostock, 2013).

102. Li, Y. *et al.* Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation abolishment and carbon alleviation. *Biotechnol. Biofuels* **7**, 25 (2014).

103. 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).

104. Cheng, C. *et al.* Development of an *in vivo* fluorescence based gene expression reporter system for *Clostridium tyrobutyricum*. *J. Biotechnol.* **305**, 18–22 (2019).

105. Chapman, S. *et al.* The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20038–43 (2008).

106. Christie, J. M. *et al.* Structural tuning of the fluorescent protein iLOV for improved photostability. *J. Biol. Chem.* **287**, 22295–22304 (2012).

107. Gawthorne, J. A. *et al.* Express your LOV: an engineered flavoprotein as a reporter for protein expression and purification. *PLoS One* **7**, e52962 (2012).

108. Buckley, A. M. *et al.* Lighting up *Clostridium difficile*: reporting gene expression using fluorescent LOV domains. *Sci. Rep.* **6**, 23463 (2016).

109. Holzer, W., Penzkofer, A., Fuhrmann, M. & Hegemann, P. Spectroscopic characterization of flavin mononucleotide bound to the LOV1 domain of Phot1 from *Chlamydomonas reinhardtii*. *Photochem. Photobiol.* **75**, 479–87 (2002).

110. Mukherjee, A. *et al.* Engineering and characterization of new LOV-based fluorescent proteins from *Chlamydomonas reinhardtii* and *Vaucheria frigida*. *ACS Synth. Biol.* **4**, 371–377 (2015).

111. Khrenova, M. G., Nemukhin, A. V. & Domratcheva, T. Theoretical characterization of the flavin-based fluorescent protein iLOV and its Q489K mutant. *J. Phys. Chem. B* **119**, 5176–5183 (2015).

112. Davari, M. D. *et al.* Photophysics of the LOV-based fluorescent protein variant iLOV-Q489K determined by simulation and experiment. *J. Phys. Chem. B* **120**, 3344–3352 (2016).

113. Crivat, G. & Taraska, J. W. Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol.* **30**, 8–16 (2012).

114. Cassona, C. P., Pereira, F., Serrano, M. & Henriques, A. O. Fluorescent reporter for single cell analysis of gene expression in Clostridium difficile. in *Methods in Molecular Biology* vol. 1476 69–90 (Humana Press, New York, NY, 2016).

115. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **21**, 86–89 (2003).

116. Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15**, 128–136 (2008).

117. 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 (2018).

118. Streett, H. E., Kalis, K. M., Papoutsakis, E. T. & Kivisaar, M. A strongly fluorescing anaerobic reporter and protein-tagging system for *Clostridium* organisms based on the fluorescence-activating and absorption-shifting tag protein (FAST). *Appl. Environ. Microbiol.* **85**, 622–641 (2019).

119. Plamont, M. A. *et al.* Small fluorescence-activating and absorption-shifting tag for tunable protein imaging *in vivo*. *Proc. Natl. Acad. Sci.* **113**, E1412–E1412 (2016).

120. Bullifent, H. L., Moir, A. & Titball, R. W. The construction of a reporter system and use for the investigation of *Clostridium perfringens* gene expression. *FEMS Microbiol. Lett.* **131**, 99–105 (1995).

121. Zhang, Y., Grosse-Honebrink, A. & Minton, N. P. A universal *mariner* transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One* **10**, e0122411 (2015).

122. Shaw, W. V. V. Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods Enzymol.* **43**, 737–755 (1975).

123. Roberts, M. C., Swenson, C. D., Owens, L. M. & Smith, A. L. Characterization of chloramphenicol-resistant *Haemophilus influenzae*. *Antimicrob. Agents Chemother.* **18**, 610–615 (1980).

124. Tummala, S. B., Welker, N. E. & Papoutsakis, E. T. Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **65**, 3793–9 (1999).

125. 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).

126. Moberg, L. J. Fluorogenic assay for rapid detection of *Escherichia coli* in food. *Appl. Environ. Microbiol.* **50**, 1383–7 (1985).

127. Sabathé, F., Croux, C., Cornillot, E. & Soucaille, P. *amyP*, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol. Lett.* **210**, 93–98 (2002).

128. Paiva, A. M. O., Friggen, A. H., Hossein-Javaheri, S. & Smits, W. K. The signal sequence of the abundant extracellular metalloprotease PPEP-1 can be used to secrete synthetic reporter proteins in *Clostridium difficile*. *ACS Synth. Biol.* **5**, 1376–1382 (2016).

129. Quixley, K. W. & Reid, S. J. Construction of a reporter gene vector for *Clostridium beijerinckii* using a *Clostridium* endoglucanase gene. *J. Mol. Microbiol. Biotechnol.* **2**, 53–7 (2000).

130. Oh, Y. H. *et al.* Construction of heterologous gene expression cassettes for the development of recombinant *Clostridium beijerinckii*. *Bioprocess Biosyst. Eng.* **39**, 555–563 (2016).

131. Gilham, D. & Lehner, R. Techniques to measure lipase and esterase activity in vitro. *Methods* **36**, 139–147 (2005).

132. Hassan, E. A., Abd-Alla, M. H., Bagy, M. M. K. & Morsy, F. M. In situ hydrogen, acetone, butanol, ethanol and microdiesel production by *Clostridium acetobutylicum* ATCC 824 from oleaginous fungal biomass. *Anaerobe* **34**, 125–131 (2015).

133. Edwards, A. N. *et al.* An alkaline phosphatase reporter for use in *Clostridium difficile*. *Anaerobe* **32**, 98–104 (2015).

134. Feustel, L., Nakotte, S. & Dürre, P. Characterization and development of two reporter gene systems for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **70**, 798–803 (2004).

135. Waidmann, M. S., Bleichrodt, F. S., Laslo, T. & Riedel, C. U. Bacterial luciferase reporters: the swiss army knife of molecular biology. *Bioeng. Bugs* **2**,

1–9 (2011).

136. Phillips-Jones, M. K. Use of a *lux* reporter system for monitoring rapid changes in α-toxin gene expression in *Clostridium perfringens* during growth. *FEMS Microbiol. Lett.* **188**, 29–33 (2000).

137. Potzkei, J. *et al.* Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor. *BMC Biol.* **10**, 28 (2012).

138. Martin, L., Che, A. & Endy, D. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS One* **4**, (2009).

139. Kelly, J. R. *et al.* Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3**, 4 (2009).

140. Haima, P., van Sinderen, D., Bron, S. & Venema, G. An improved β-galactosidase α-complementation system for molecular cloning in *Bacillus subtilis*. *Gene* **93**, 41–47 (1990).

141. Knipfer, N., Nooruddin, L. & Shrader, T. E. Development of an alpha-complementation system for mycobacterial promoter analysis. *Gene* **217**, 69–75 (1998).

142. Volkov, I. Y., Lunina, N. A., Berezina, O. V., Velikodvorskaya, G. A. & Zverlov, V. V. *Thermoanaerobacter ethanolicus* gene cluster containing the α- and β-galactosidase genes *melA* and *lacA* and properties of recombinant LacA. *Mol. Biol.* **39**, 799–805 (2005).

143. Huber, R. E., Roth, N. J. & Bahl, H. Quaternary structure, Mg2+ interactions, and some kinetic properties of the β-galactosidase from *Thermoanaerobacterium*

thermosulfurigenes EM1. *J. Protein Chem.* **15**, 621–629 (1996).

144. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **110**, 14024–14029 (2013).

145. 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**, 1137–1142 (2012).

146. Köbbing, S., Blank, L. M. & Wierckx, N. Characterization of context-dependent effects on synthetic promoters. *Front. Bioeng. Biotechnol.* **8**, 1–13 (2020).

147. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39**, 1131–1141 (2011).

148. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).

149. 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).

150. Nielsen, A. A. K., Segall-Shapiro, T. H. & Voigt, C. A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.* **17**, 878–892 (2013).

151. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360

(2013).

152. 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).

153. Li, R., Zhang, Q., Li, J. & Shi, H. Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. *Nucleic Acids Res.* **44**, gkv1285- (2015).

154. Shultzaberger, R. K., Bucheimer, R. E. E., Rudd, K. E. & Schneider, T. D. Anatomy of *Escherichia coli* ribosome binding sites. *J. Mol. Biol.* **313**, 215–228 (2001).

155. Yang, Y. *et al.* Improving the performance of solventogenic clostridia by reinforcing the biotin synthetic pathway. *Metab. Eng.* **35**, 121–128 (2016).

156. Yang, G. *et al.* Rapid generation of universal synthetic promoters for controlled gene expression in both gas-fermenting and saccharolytic *Clostridium* species. *ACS Synth. Biol.* acssynbio.7b00155 (2017).

157. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine – Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* **22**, 4953–4957 (1994).

158. Hager, P. W. & Rabinowitz, J. C. Translational specificity in *Bacillus subtilis*. *Mol. Biol. bacilli* **2**, 1–32 (1985).

159. Band, L. & Henner, D. J. *Bacillus subtilis* requires a 'stringent' Shine-Dalgarno region for gene expression. *DNA* **3**, 17–21 (1984).

160. McLaughlin, J. R., Murray, C. L. & Rabinowitz, J. C. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* beta-lactamase gene. *J. Biol. Chem.* **256**, 11283–91 (1981).

161. Abolbaghaei, A., Silke, J. R. & Xia, X. How changes in anti-SD sequences would affect SD sequences in *Escherichia coli* and *Bacillus subtilis*. *G3 Genes, Genomes, Genet.* **7**, 1607–1615 (2017).

162. Dunn, J. J. & Studier, F. W. Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol.* **148**, 303–330 (1981).

163. Ringquist, S. *et al.* Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **6**, 1219–29 (1992).

164. Garnier, T., Canard, B. & Cole, S. T. Cloning, mapping, and molecular characterization of the rRNA operons of *Clostridium perfringens*. *J. Bacteriol.* **173**, 5431–5438 (1991).

165. Ma, J., Campbell, A. & Karlin, S. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733–5745 (2002).

166. He, J., Sakaguchi, K. & Suzuki, T. Determination of the ribosome-binding sequence and spacer length between binding site and initiation codon for efficient protein expression in *Bifidobacterium longum* 105-A. *J. Biosci. Bioeng.* **113**, 442–444 (2012).

167. de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7668–7672 (1990).

168. Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42 (2011).

169. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).

170. Reeve, B., Martinez-Klimova, E., Jonghe, J. de, Leak, D. J. & Ellis, T. The *Geobacillus* plasmid set: a modular toolkit for thermophile engineering. (2016).

171. Tian, L. *et al.* Metabolic engineering of *Clostridium thermocellum* for n-butanol production from cellulose. *Biotechnol. Biofuels* **12**, 1–13 (2019).

172. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).

173. Sharp, P. M. & Li, W.-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).

174. Baca, A. M. & Hol, W. G. J. J. Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.* **30**, 113–118 (2000).

175. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of *E.coli* for enhanced protein expression. *Innovations* **12**, 1–3 (2001).

176. Mondal, S. K., Kundu, S., Das, R. & Roy, S. Analysis of phylogeny and codon usage bias and relationship of GC content, amino acid composition with expression of the structural *nif* genes. *J. Biomol. Struct. Dyn.* **1102**, 1–18 (2015).

177. Musto, H., Romero, H. & Zavala, A. Translational selection is operative for synonymous codon usage in *Clostridium perfringens* and *Clostridium acetobutylicum*. *Microbiology* vol. 149 855–863 (2003).

178. Sastalla, I., Chim, K., Cheung, G. Y. C., Pomerantsev, A. P. & Leppla, S. H. Codon-optimized fluorescent proteins designed for expression in low-GC gram-positive bacteria. *Appl. Environ. Microbiol.* **75**, 2099–2110 (2009).

179. Wright, F. The 'effective number of codons' used in a gene. *Gene* **87**, 23–29 (1990).

180. Welch, M. *et al.* Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One* **4**, e7002 (2009).

181. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of expression in *Escherichia coli*. *Science (80-.).* **324**, 255–258 (2009).

182. Xia, X. An improved implementation of codon adaptation index. *Evol. Bioinforma.* **3**, 53–58 (2007).

183. Xia, X. A major controversy in codon-anticodon adaptation resolved by a new codon usage index. *Genetics* **199**, 573–579 (2015).

184. Lanza, A. M., Curran, K. A., Rey, L. G. & Alper, H. S. A condition-specific codon optimization approach for improved heterologous gene expression in

Saccharomyces cerevisiae. *BMC Syst. Biol.* **8**, 33 (2014).

185. Welch, M., Villalobos, A., Gustafsson, C. & Minshull, J. You're one in a googol: optimizing genes for protein expression. *J. R. Soc. Interface* **6**, (2009).

186. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).

187. Carrier, T. A. T. A. & Keasling, J. D. D. Controlling messenger RNA stability in Bacteria: strategies for engineering gene expression. *Biotechnol. Prog.* **13**, 699–708 (1997).

188. Carrier, T. A. T. A. & Keasling, J. D. D. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog.* **15**, 58–64 (1999).

189. Condon, C. & Putzer, H. The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res.* **30**, 5339–46 (2002).

190. Durand, S., Tomasini, A., Braun, F., Condon, C. & Romby, P. sRNA and mRNA turnover in gram-positive bacteria. *FEMS Microbiology Reviews* vol. 39 316–330 (2015).

191. Lehnik-Habrink, M. *et al.* RNase Y in *Bacillus subtilis*: a natively disordered protein that is the functional equivalent of RNase E from *Escherichia coli*. *J. Bacteriol.* **193**, 5431–41 (2011).

192. Lee, J., Jang, Y.-S. S., Papoutsakis, E. T. & Lee, S. Y. Stable and enhanced gene expression in *Clostridium acetobutylicum* using synthetic untranslated regions with a stem-loop. *J. Biotechnol.* **230**, 40–43 (2016).

193. Desai, R. P. & Papoutsakis, E. T. Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **65**, 936–945 (1999).

194. Song, J.-W. *et al.* 3′-UTR engineering to improve soluble expression and fine-tuning of activity of cascade enzymes in *Escherichia coli*. *Sci. Rep.* **6**, 29406 (2016).

195. Tummala, S. B., Junne, S. G. & Papoutsakis, E. T. Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium acetobutylicum* fermentations. *J. Bacteriol.* **185**, 3644–3653 (2003).

196. Pyne, M., Moo-Young, M., Chung, D. & Chou, C. Antisense-RNA-mediated gene downregulation in *Clostridium pasteurianum*. *Fermentation* **1**, 113–126 (2015).

197. Storz, G., Altuvia, S. & Wassarman, K. M. An abundance of RNA regulators. *Annu. Rev. Biochem.* **74**, 199–217 (2005).

198. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B. & Arkin, A. P. Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454 (2012).

199. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841–847 (2004).

200. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: *de-novo*-designed regulators of gene expression. *Cell* **159**, 925–939 (2014).

201. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

202. Mellin, J. R. *et al.* Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31**, 150–6 (2015).

203. Wittmann, A. & Suess, B. Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* **586**, 2076–2083 (2012).

204. Saad, N. Y. *et al.* Riboswitch (T-box)-mediated control of tRNA-dependent amidation in *Clostridium acetobutylicum* rationalizes gene and pathway redundancy for asparagine and asparaginyl-tRNA^{Asn} synthesis. *J. Biol. Chem.* **287**, 20382–94 (2012).

205. André, G. *et al.* S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* **36**, 5955–69 (2008).

206. Kim, P. B., Nelson, J. W. & Breaker, R. R. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **57**, 317–328 (2015).

207. Browning, D. D. F. & Busby, S. J. W. S. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).

208. Helmann, J. D. & Chamberlin, M. J. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **57**, 839–872 (1988).

209. Sauer, U. *et al.* Sigma factor and sporulation genes in *Clostridium*. *FEMS Microbiol. Rev.* **17**, 331–340 (1995).

210. Estrem, S. T., Gaal, T., Ross, W. & Gourse, R. L. Identification of an UP element consensus sequence for bacterial promoters. *Biochemistry* **95**, 9761–9766 (1998).

211. Kumar, A. *et al.* The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an 'extended minus 10' promoter. *J. Mol. Biol.* **232**, 406–18 (1993).

212. Sinoquet, C., Demey, S. & Braun, F. Large-scale computational and statistical analyses of high transcription potentialities in 32 prokaryotic genomes. *Nucleic Acids Res.* **36**, 3332–3340 (2008).

213. Katayama, S., Matsushita, O., Jung, C. M., Minami, J. & Okabe, A. Promoter upstream bent DNA activates the transcription of the *Clostridium perfringens* phospholipase C gene in a low temperature-dependent manner. *EMBO J.* **18**, 3442–3450 (1999).

214. Graves, M. C. & Rabinowitz, J. C. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for 'extended' promoter elements in gram-positive organisms. *J. Biol. Chem.* **261**, 11409–15 (1986).

215. Pyne, M. E. Development of genetic tools for metabolic engineering of *Clostridium pasteurianum*. (University of Waterloo, 2014).

216. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Appl Env. Microbiol* **54**, 2717–2722 (1988).

217. Stim-Herndon, K. P., Petersen, D. J. & Bennett, G. N. Characterization of an acetyl-CoA C-acetyltransferase (thiolase) gene from *Clostridium acetobutylicum* ATCC 824. *Gene* **154**, 81–85 (1995).

218. Bormann, S. *et al.* Engineering *Clostridium acetobutylicum* for production of kerosene and diesel blendstock precursors. *Metab. Eng.* **25**, 124–130 (2014).

219. Perret, S. *et al.* Production of heterologous and chimeric scaffoldins by *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* **186**, 253–257 (2004).

220. Mingardon, F. *et al.* Heterologous production, assembly, and secretion of a minicellulosome by *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **71**, 1215–1222 (2005).

221. Wietzke, M. & Bahl, H. The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* **96**, 749–761 (2012).

222. Zhang, L. *et al.* Redox-responsive repressor Rex modulates alcohol production and oxidative stress tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* **196**, 3949–3963 (2014).

223. Novichkov, P. S. *et al.* RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).

224. Wang, Y. *et al.* Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. *Biotechnol. Bioeng.* **113**, 2739–2743 (2016).

225. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Expansion of the genetic toolkit for metabolic engineering of *Clostridium pasteurianum*: chromosomal gene disruption of the endogenous CpaAI restriction enzyme. *Biotechnol. Biofuels* **7**, 163 (2014).

226. Geissendörfer, M. & Hillen, W. Regulated expression of heterologous genes in

Bacillus subtilis using the Tn10 encoded *tet* regulatory elements. *Appl. Microbiol. Biotechnol.* **33**, 657–663 (1990).

227. Corrigan, R. M. & Foster, T. J. An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. *Plasmid* **61**, 126–129 (2009).

228. Fagan, R. P. & Fairweather, N. F. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* **286**, 27483–27493 (2011).

229. Garcia, H. G. *et al.* Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep.* **2**, 150–161 (2012).

230. Zong, Y. *et al.* Insulated transcriptional elements enable precise design of genetic circuits. *Nat. Commun.* **8**, (2017).

231. An, W. & Chin, J. W. Synthesis of orthogonal transcription-translation networks. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8477–8482 (2009).

232. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* (1986).

233. William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* (1990).

234. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Biochemistry* **82**, 1074–1078 (1985).

235. Lamberte, L. E. *et al.* Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.* **2**, 16249 (2017).

236. Heap, J. T. *et al.* The ClosTron: mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* **80**, 49–55 (2010).

237. Fontaine, L. *et al.* Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* **184**, 821–30 (2002).

238. Ciampi, M. S. Rho-dependent terminators and transcription termination. *Microbiology* **152**, 2515–28 (2006).

239. Liu, C. C. *et al.* An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods* **9**, 1088–1094 (2012).

240. Banerjee, S., Chalissery, J., Bandey, I. & Sen, R. Rho-dependent transcription termination: more questions than answers. *J. Microbiol.* **44**, 11–22 (2006).

241. Gardner, P. P., Barquist, L., Bateman, A., Nawrocki, E. P. & Weinberg, Z. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845–52 (2011).

242. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).

243. Toulokhonov, I. & Landick, R. The Flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol. Cell* **13**, 299 (2004).

244. Sugimoto, N. *et al.* Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **34**, 11211–11216 (1995).

245. Yarnell, W. S. & Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science (80-.).* **284**, 611–615 (1999).

246. Bellecourt, M. J., Ray-Soni, A., Harwig, A., Mooney, R. A. & Landick, R. RNA polymerase clamp movement aids dissociation from DNA but is not required for RNA release at intrinsic terminators. *J. Mol. Biol.* **431**, 696–713 (2019).

247. Naville, M. & Gautheret, D. Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol.* **11**, R97 (2010).

248. Ravagnani, A. *et al.* Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Mol. Microbiol.* **37**, 1172–1185 (2000).

249. Lin, M.-T. T. *et al.* Novel utilization of terminators in the design of biologically adjustable synthetic filters. *ACS Synth. Biol.* **5**, 365–374 (2016).

250. Abe, H., Abo, T. & Aiba, H. Regulation of intrinsic terminator by translation in *Escherichia coli*: transcription termination at a distance downstream. *Genes Cells* **4**, 87–97 (1999).

251. Kelly, C. L., Taylor, G. M., Šatkutė, A., Dekker, L. & Heap, J. T. Transcriptional terminators allow leak-free chromosomal integration of genetic constructs in cyanobacteria. *Microorganisms* **7**, (2019).

252. Park, Y., Espah Borujeni, A., Gorochowski, T. E., Shin, J. & Voigt, C. A. Precision design of stable genetic circuits carried in highly-insulated E. coli genomic landing pads. *Mol. Syst. Biol.* **16**, (2020).

253. Postle, K. & Good, R. F. A bidirectional rho-independent transcription terminator

between the *E. coli tonB* gene and an opposing gene. *Cell* **41**, 577–585 (1985).

254. Lesnik, E. A. *et al.* Prediction of Rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* **29**, 3583–94 (2001).

255. Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* **8**, R22 (2007).

256. Mitra, A., Angamuthu, K., Jayashree, H. V. & Nagaraja, V. Occurrence, divergence and evolution of intrinsic terminators across eubacteria. *Genomics* **94**, 110–6 (2009).

257. Mitra, A., Kesarwani, A. K., Pal, D. & Nagaraja, V. WebGeSTer DB--a transcription terminator database. *Nucleic Acids Res.* **39**, D129-35 (2011).

258. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

259. Gerischer, U. & Durre, P. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. *J. Bacteriol.* **172**, 6907–6918 (1990).

260. Barthelemy, I., Salas, M. & Mellado, R. P. *In vivo* transcription of bacteriophage Φ29 DNA: transcription termination. *J. Virol.* **61**, 1751–1755 (1987).

261. Pulido, D., Jiménez, A., Salas, M. & Mellado, R. P. A *Bacillus subtilis* phage φ29 transcription terminator is efficiently recognized in *Streptomyces lividans*. *Gene* **56**, 277–282 (1987).

262. Gorwa, M. F., Croux, C. & Soucaille, P. Molecular characterization and

transcriptional analysis of the putative hydrogenase gene of *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* **178**, 2668–2675 (1996).

263. Bardhan, S. K., Gupta, S., Gorman, M. E. & Haider, M. A. Biorenewable chemicals: feedstocks, technologies and the conflict with food production. *Renew. Sustain. Energy Rev.* **51**, 506–520 (2015).

264. Papoutsakis, E. T. Engineering solventogenic clostridia. *Curr. Opin. Biotechnol.* **19**, 420–9 (2008).

265. Hayashida, S. & Ahn, B. K. Isolation and characteristics of an acetone-butanol-negative, ethanol-isovaleric acid-producing mutant of *Clostridium saccharoperbutylacetonicum* N1-4 ATCC 13564. *Agric. Biol. Chem.* **54**, 343–351 (1990).

266. Wagner, A. O. *et al.* Medium preparation for the cultivation of microorganisms under strictly anaerobic/anoxic conditions. *J. Vis. Exp.* **2019**, (2019).

267. Iverson, S., Haddock, T. L., Beal, J. & Densmore, D. CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enables rapid combinatorial design for synthetic and traditional biology. *ACS Synth. Biol.* **5**, 99–103 (2015).

268. Strand, T. A., Lale, R., Degnes, K. F., Lando, M. & Valla, S. A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS One* **9**, e90372 (2014).

269. Pósfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science (80-.).* **312**, 1044–1046 (2006).

270. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* (2012).

271. Gyulev, I. S. *et al.* Part by part: synthetic biology parts used in solventogenic Clostridia. *ACS Synth. Biol.* **7**, 311–327 (2018).

272. Müh, U., Pannullo, A. G., Weiss, D. S. & Ellermeier, C. D. A xylose-inducible expression system and a CRISPR interference plasmid for targeted knockdown of gene expression in *Clostridioides difficile*. *J. Bacteriol.* **201**, (2019).

273. Khader, H., Solodushko, V., Al-Mehdi, A. B., Audia, J. & Fouty, B. Overlap of doxycycline fluorescence with that of the redox-sensitive intracellular reporter roGFP. *J. Fluoresc.* **24**, 305–311 (2014).

274. Chang, W.-B., Zhao, Y.-B., Ci, Y.-X. & Hu, L.-Y. Spectrofluorimetric determination of tetracycline and anhydrotetracycline in serum and urine. *Analyst* **117**, 1377 (1992).

275. Kim, Y. *et al.* Single amino acid replacement transforms mCherry to a far-red fluorescent protein. *Biotechnol. Bioprocess Eng.* (2016).

276. Truffaut, N., Hubert, J. & Reysset, G. Construction of shuttle vectors useful for transforming *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* **58**, 15–19 (1989).

277. Lin-Chao, S., Chen, W.-T. & Wong, T.-T. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol. Microbiol.* **6**, 3385–3393 (1992).

278. Mukherjee, A., Walker, J., Weyant, K. B. & Schroeder, C. M. Characterization of flavin-based fluorescent proteins: an emerging class of fluorescent reporters. *PLoS One* **8**, e64753 (2013).

279. Billinton, N. & Knight, A. W. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical Biochemistry* (2001).

280. Tan, Y., Liu, Z. Y., Liu, Z., Zheng, H. J. & Li, F. L. Comparative transcriptome analysis between *csrA*-disruption *Clostridium acetobutylicum* and its parent strain. *Mol. Biosyst.* (2015).

281. Burchhardt, G. & Bahl, H. Cloning and analysis of the β-galactosidase-encoding gene from *Clostridium thermosulfurogenes* EM1. *Gene* (1991).

282. Rosenberg, M., Chepelinsky, A. & McKenney, K. Studying promoters and terminators by gene fusion. *Science (80-.).* **222**, 734–739 (1983).

283. Deuschle, U., Kammerer, W., Gentz, R. & Bujard, H. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J.* **5**, 2987–94 (1986).

284. Brunner, M. & Bujard, H. Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO J.* (1987).

285. Mooney, R. A., Artsimovitch, I. & Landick, R. Information processing by RNA polymerase: recognition of rregulatory signals during RNA chain elongation. *J. Bacteriol.* **180**, 3265–3275 (1998).

286. Warburton, N., Boseley, P. G. & Porter, A. G. Increased expression of a cloned

gene by local mutagenesis of its promoter and ribosome binding site. *Nucleic Acids Res.* **11**, 5837–5854 (1983).

287. Adhya, S. & Gottesman, M. Control of transcription termination. *Annu. Rev. Biochem.* **47**, 967–996 (1978).

288. Henkin, T. M. Control of transcription termination in prokaryotes. *Annu. Rev. Genet.* **30**, 35–57 (1996).

289. Klotsky, R. A. & Schwartz, I. Measurement of *cat* expression from growth-rate-regulated promoters employing β-lactamase activity as an indicator of plasmid copy number. *Gene* **55**, 141–146 (1987).

290. Rudge, T. J. *et al.* Characterization of intrinsic properties of promoters. *ACS Synth. Biol.* (2015).

291. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* **297**, (2002).

292. Honigman, A. *et al.* Plasmid vectors designed for the analysis of transcription termination signals. *Gene* **36**, 131–141 (1985).

293. Amarelle, V., Sanches-Medeiros, A., Silva-Rocha, R. & Guazzaroni, M. E. Expanding the toolbox of broad host-range transcriptional terminators for Proteobacteria through metagenomics. *ACS Synth. Biol.* (2019).

294. Henkin, T. M. Analysis of tRNA-directed transcription antitermination in the T-box system *in vivo*. in *Methods in molecular biology (Clifton, N.J.)* vol. 540 281–290 (2009).

295. Yoo, J. & Kang, C. Variation of *in vivo* efficiency of the bacteriophage T7

terminator depending on terminator-upstream sequences. *Mol. Cells* **6**, 352–358 (1996).

296. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. (2006).

297. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7090–5 (2000).

298. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3' -end chronicles. **412**, 793–813 (2011).

299. Raghunathan, N. *et al.* Genome-wide relationship between R-loop formation and antisense transcription in *Escherichia coli*. *Nucleic Acids Res.* **46**, 3400–3411 (2018).

300. Belogurov, G. A. & Artsimovitch, I. Regulation of transcript elongation. *Annu. Rev. Microbiol.* **69**, 49–69 (2015).

301. Dar, D. & Sorek, R. High-resolution RNA 3′-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res.* **46**, 6797–6805 (2018).

302. Wang, X. *et al.* Processing generates 3′ ends of RNA masking transcription termination events in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4440–4445 (2019).

303. Rennig, M. *et al.* TARSyn: tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. *ACS Synth. Biol.*

acssynbio.7b00200 (2018).

304. Lorenz, R. *et al.* ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

305. parts.igem.org. http://parts.igem.org/Main_Page.

306. Mézard, C., Pompon, D. & Nicolas, A. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. *Cell* (1992).

307. Santos-Zavaleta, A. *et al.* RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).

308. Unniraman, S., Prakash, R. & Nagaraja, V. Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.* **30**, 675–84 (2002).

309. ARNold, finding terminators at I2BC - Web Server. http://rssf.i2bc.paris-saclay.fr/toolbox/arnold/index.php#Results.

310. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70-4 (2008).

311. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

312. Wilson, K. S. & Hippel, P. H. Von. Transcription termination at intrinsic terminators: the role of the RNA hairpin (*Escherichia coli*/RNA polymerase/Rho-independent termination). *Biochemistry* **92**, 8793–8797 (1995).

313. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription

terminators. *Nucleic Acids Res.* **41**, 5139–5148 (2013).

314. Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A. & Turner, D. H. RNA hairpin loop stability depends on closing base pair. *Nucleic Acids Res.* **21**, 3845–3849 (1993).

315. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* 1188–1190 (2004).

316. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–100 (1990).

317. Henkin, T. M. & Sonenshein, A. L. Mutations of the *Escherichia coli lacUV5* promoter resulting in increased expression in *Bacillus subtilis*. *Mol Gen Genet* **209**, 467–474 (1987).

318. Mathes, T., Vogl, C., Stolz, J. & Hegemann, P. *In vivo* generation of flavoproteins with modified cofactors. *J. Mol. Biol.* (2009).

319. Vogl, C. *et al.* Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J. Bacteriol.* (2007).

320. Vitreschak, A. G. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* (2002).

321. Hickey, R. J. The inactivation of iron by 2, 2'-bipyridine and its effect on riboflavin synthesis by *Clostridium acetobutylicum*. *Arch. Biochem.* **8**, 439–447 (1945).

322. Welz, D. & Braun, V. Ferric citrate transport of *Escherichia coli*: functional

regions of the FecR transmembrane regulatory protein. *J. Bacteriol.* **180**, 2387–2394 (1998).

323. Ollinger, J., Song, K.-B., Antelmann, H., Hecker, M. & Helmann, J. D. Role of the Fur regulon in iron transport in *Bacillus subtilis*. *J. Bacteriol.* **188**, 3664–3673 (2006).

324. Krauss, U., Svensson, V., Wirtz, A., Knieps-Grünhagen, E. & Jaeger, K.-E. Cofactor trapping, a new method to produce flavin mononucleotide. *Appl. Environ. Microbiol.* **77**, 1097–100 (2011).

325. Palm, G. J. *et al.* Specific binding of divalent metal ions to tetracycline and to the Tet repressor/tetracycline complex. *JBIC J. Biol. Inorg. Chem.* **13**, 1097–1110 (2008).

326. Nielsen, J. L., Schramm, A., Bernhard, A. E., Van Den Engh, G. J. & Stahl, D. A. Flow cytometry-assisted cloning of specific sequence motifs from complex 16S rRNA gene libraries. *Appl. Environ. Microbiol.* **70**, 7550–7554 (2004).

327. Liang, W.-J. *et al.* The *gusBC* genes of *Escherichia coli* encode a glucuronide transport system. *J. Bacteriol.* **187**, 2377–85 (2005).

328. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

329. Grisewood, M. J. *et al.* OptZyme: computational enzyme redesign using transition state analogues. *PLoS One* **8**, e75358 (2013).

330. Smith, W. S., Hale, J. R. & Neylon, C. Applying neutral drift to the directed molecular evolution of a β-glucuronidase into a β-galactosidase: two different

evolutionary pathways lead to the same variant. *BMC Res. Notes* **4**, 138 (2011).

331. Broome, A.-M., Bhavsar, N., Ramamurthy, G., Newton, G. & Basilion, J. P. Expanding the utility of beta-galactosidase complementation: piece by piece. *Mol. Pharm.* **7**, 60–74 (2010).

332. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2012).

333. Kamens, J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.* **43**, D1152–D1157 (2015).

334. Quinn, J. Y. J. *et al.* SBOL Visual: a graphical language for genetic designs. *PLoS Biol.* **13**, e1002310 (2015).