Version: Accepted Version

Article:

Gyulev, Ivan, Willson, Benjamin James, Hennessy, Rosanna Catherine et al. (2018) Part by part:synthetic biology parts used in solventogenic Clostridia. ACS Synthetic Biology. pp. 311-327. ISSN: 2161-5063

https://doi.org/10.1021/acssynbio.7b00327

Part by part: synthetic biology parts used in solventogenic Clostridia

Authors:

Ivan S. Gyulev[1*], Benjamin J. Willson[1*], Rosanna C. Hennessy[2], Preben Krabben[3], Elizabeth R. Jenkinson[4], Gavin H. Thomas[1§]

Author affiliation:

1. Department of Biology, University of York, Wentworth Way, York, YO10 5DD, United Kingdom

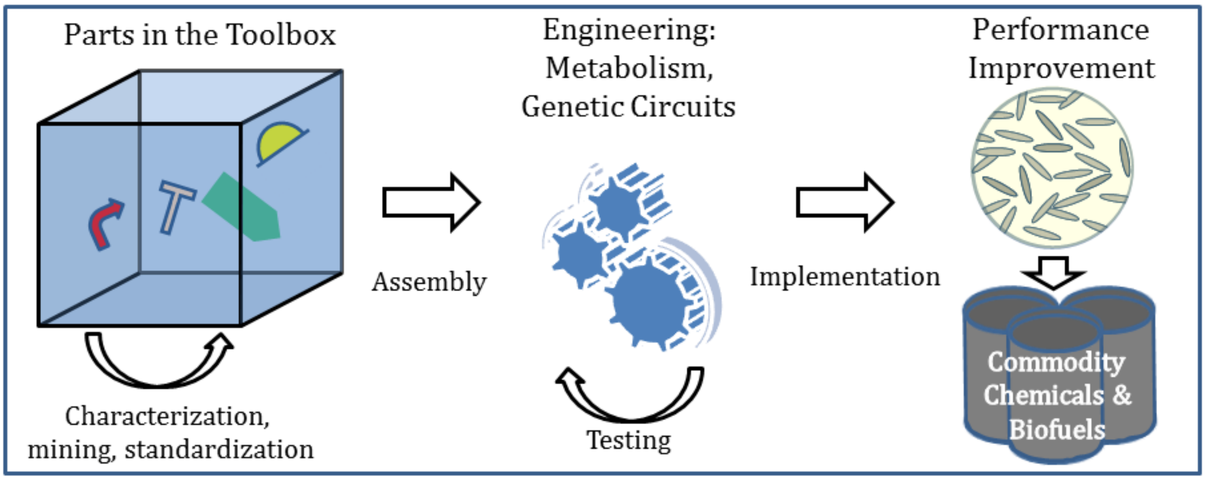
2. Microbial Ecology and Biotechnology, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

3. 9 Marsh Lane, Didcot, Oxfordshire, OX11 8FD, United Kingdom

4. Green Biologics Limited, Milton Park, Abingdon, Oxfordshire, OX14 4RU, United Kingdom

*** These authors contributed equally to this work.**

§ - Phone: 44-1904-328678. E-mail: gavin.thomas@york.ac.uk.



Graphical Table of Contents.

ABSTRACT: The solventogenic Clostridia are of interest to the chemical industry because of their natural ability to produce chemicals such as butanol, acetone and ethanol from diverse feedstocks. Their use as whole cell factories presents multiple metabolic engineering targets that could lead to improved sustainability and profitability of *Clostridium* industrial processes. However, engineering efforts have been held back by the scarcity of genetic and synthetic biology tools. Over the last decade, genetic tools to enable transformation and chromosomal modifications have been developed, but the lack of a broad palette of synthetic biology parts remains one of the last obstacles to the rapid engineered improvement of these species for bioproduction. We have systematically reviewed existing parts that have been used in the modification of solventogenic Clostridia, revealing a narrow range of empirically chosen and non-engineered parts that are in current use. The analysis uncovers elements, such as promoters, transcriptional terminators and ribosome binding sites where increased fundamental knowledge is needed for their reliable use in different applications. Together, the review provides the most comprehensive list of parts used and also presents areas where an improved toolbox is needed for full exploitation of these industrially important bacteria.

Keywords

Clostridium, ABE fermenation, solventogenesis, transcription termination, reporters, promoters

Abbreviations

ABE: acetone-butanol-ethanol

ACE: allele-coupled exchange

aSD: anti Shine-Dalgarno sequence, found at the 3’ end of 16S rRNA

CDS: coding sequence

FbFP: flavin-binding fluorescent protein

RBS: ribosome binding site

RNAP: RNA-polymerase

SD: Shine-Dalgarno sequence

UTR: untranslated region

Acknowledgements:

We acknowledge the BBSRC for support for Ivan Gyulev through an iCASE PhD studentship with Green Biologics Ltd (BB/M016331/1), Benjamin Willson from an IB Catalyst grant for Project DETOX (BB/N01040X/1) and Rosanna Hennessy for a BBSRC IUK award (BB/L011522/1). We thank Stephen Thorpe for writing a Python script that we used to compare the overlaps of the outputs of the terminator prediction algorithms.

One of the earliest industrial biotechnology processes used by man was the production of acetone by bacterial fermentation. The bacterium used was from the genus *Clostridium*, which are Gram-positive, spore forming, obligately anaerobic organisms[1]. In addition to acetone, these processes produced butanol and ethanol, leading to the fermentation being given the name 'ABE' for 'acetone-butanol-ethanol'[2]. 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', reassimilating the produced acids and producing solvents[3]. 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[4,5], namely *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharobutylicum*, and *Clostridium saccharoperbutylacetonicum*, although ABE fermentation has been observed in other clostridia[6–8]. 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[9]. While the ABE process had until recently fallen out of favour due to competition from the petrochemical industry, the necessity of identifying alternative fuels has renewed interest in the production of butanol as a potential biofuel candidate and sustainable commodity chemical[10].

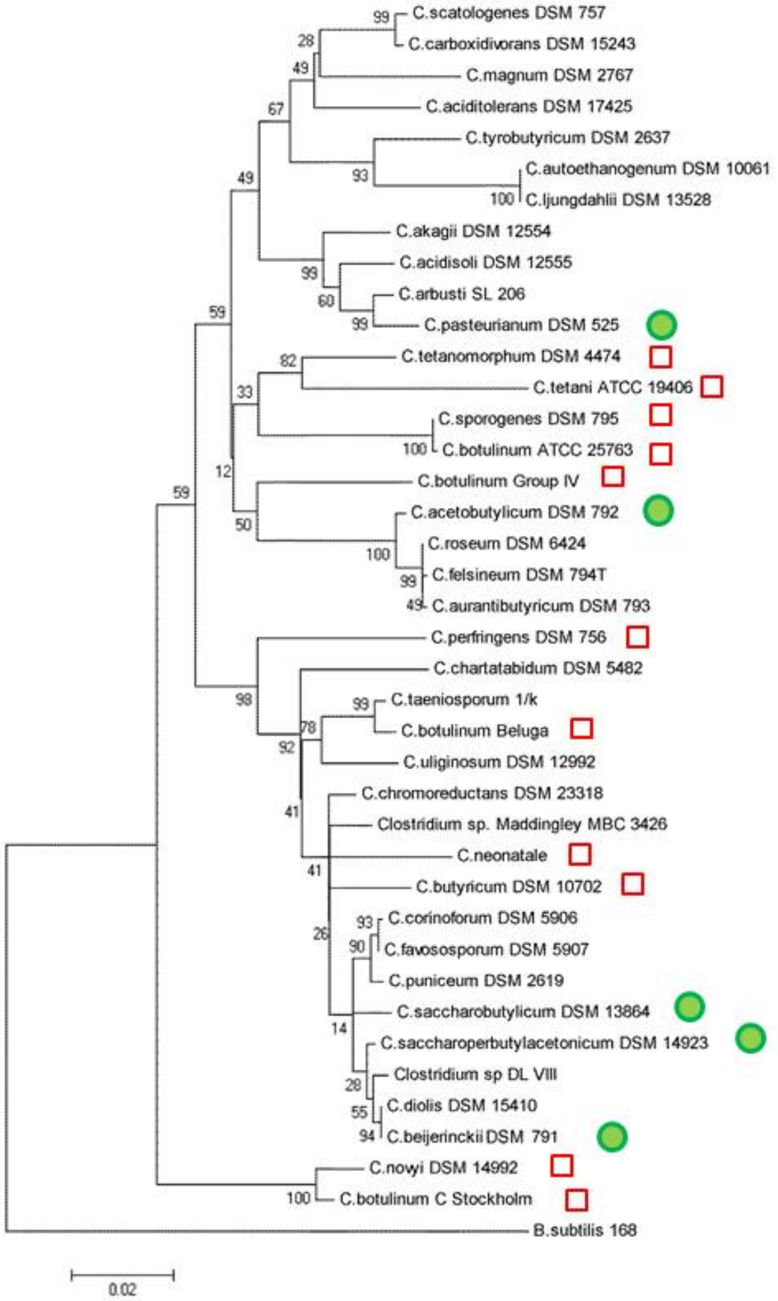


Figure 1. *Phylogenetic tree of 16S sequences from selected organisms from the genus Clostridium. The tree was built using Maximum Likelihood method based on the Tamura-Nei model*[11]*. The bootstrap consensus tree inferred from 1000 replicates*[12] *is taken to represent the evolutionary history of the taxa analyzed*[12]*. Evolutionary analyses were conducted in MEGA5*[13]*. Green circles denote butanol-producing species used in industrial biotechnology (IB), red squares mark risk group 2 species*[14] *(risk group 3 being highest risk).*

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*[15] (Figure 1). Interestingly, another comparative genomics paper which included more genus-wide species 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*[16]*.* 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[17]. Topics of engineering interest have included improved characteristics such as solvent production[18,19], sugar utilisation[20,21], growth on alternative feedstocks such as lignocellulose[22–25] and the production of alternative products[26–28].

A shortage of biological parts for the solventogenic Clostridia

To manipulate the solventogenic Clostridia using rational engineering, the development of reliable standard genetic parts for use in synthetic biology is essential. The assembly of these parts into synthetic gene cassettes and larger gene clusters means that they must work together in a consistent and predictable manner to become a useful tool for strain engineering. The basic parts of synthetic biology are the minimal sequence elements with biological function in gene expression (Figure 2), including promoters, ribosome binding sites, transcriptional terminators and other factors in the mRNA that affect stability. Also, synthetic biology tools rely on gene reporters to measure levels of gene expression, and these elements are often combined into plasmids which need their own origins 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.

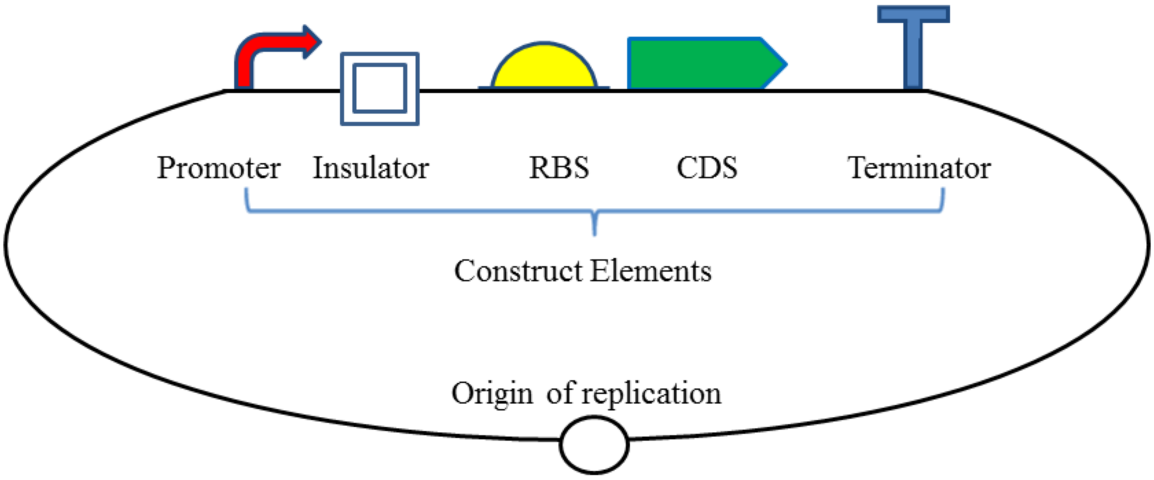


Figure 2. *Features of genetic constructs depicted using symbols from the SBOLv[29]. Examples of promoters, ribosomal binding sites, reporters, terminators and replicons are described in the main text.*

It is also important to note that, of the existing body of molecular biology research in this genus, the majority has been carried out in *C. acetobutylicum* ATCC 824. Work in this species has benefited from a relatively early development of a reliable electroporation protocol[30] and from the publication of the whole genome sequence in 2001, making it the first published clostridial genome[31]. However, *C. acetobutylicum* is only one of many industrially promising solventogenic clostridia. While the engineering of some of these species has historically been hindered by a lack of transformation protocols, these are now available for the transformation of the aforementioned five industrial species[30,32–35]. 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[30]. 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[33] or M.BepI[36] 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[34]. 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[15] 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[37]; conversely, *C. beijerinckii* NCIMB 8052 can be transformed with much greater efficiency even with DNA from a *dam*$^+$/*dcm*$^+$ host[38]. While an analysis of such developments in the entire genus *Clostridium* would be beyond the scope of this review, recent publications by Pyne et al.[39] and Minton et al.[40] provide a comprehensive review of the development of *Clostridium* strains for genetic engineering. 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 weaknesses in the 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. Here, we assess the currently available parts used in synthetic biology projects, starting first with reporters available to measure gene expression, then the promoters, terminators working at the level of transcription and then post-transcriptional factors including mRNA stability, ribosome binding sites, riboswitches and codon usage. Finally we review the replicons currently being used to assemble synthetic gene clusters and briefly provide an overview of methods to introduce these genetic elements onto the chromosome, which has been reviewed elsewhere[40]. Overall, we aim to summarize and systematize the existing biological parts used in the industrial solventogenic *Clostridium* species in order to highlight areas where more research is required and accelerate progress in creating a full range of synthetic biology tools for manipulating these industrially important organisms.

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. 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.

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.

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)[41]. The major limitation of the GFP-like proteins for anaerobes is the requirement of molecular oxygen for chromophore maturation and fluorescence[42]. 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*[43], where yellow fluorescent protein (YFP) was synthesized and fluorescence was developed after exposure of live cells to atmospheric oxygen[44].

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)[45] and mCherry (red)[46]. 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)[46]. So far, the technique has not been used to systematically quantify gene expression but rather to label proteins and track their intracellular localization[46,47].

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.

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[42]. This domain was engineered to emit fluorescence by mutagenesis of a reactive cysteine in the FMN-binding pocket[42].

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

The reporters’ functionality in *C. acetobutylicum* was 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*[48]. 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[49]. The functionality of the evoglow Bs2 variant has also been demonstrated in *C. pasteurianum*[50] and in *Clostridium ljungdahlii*[51].

Another FbFP example, developed from the *Arabidopsis thaliana* LOV2 domain of the blue-light receptor protein Phot2, is the improved LOV (iLOV) FbFP[52]. The iLOV FbFP has been further modified for enhanced properties, generating variants such as photostable iLOV 2.1 (phiLOV 2.1) and others[53,54]. 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 strong *xyl* promoter)[55]. 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[55]. While the FbFP reporters will 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.

One of the drawbacks of the FbFPs is that all of the available variants are of the same colour; additionally, unfolding caused by translocation has been reported to result in loss of fluorescence due to the loss of the bound flavin cofactor[54]. 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[56], experimental results showed that the mutation resulted in an overall brightness decrease and a blue shift of emission[57]. Research is underway to develop more fluorescent derivatives of the LOV domains and to explore the potential for diversifying their colour spectrum[56,57]. 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[53]; for applications where a longer exposure time is crucial, the dimmer, stable protein is preferable, but for an automated single measurement application such as flow cytometry, the brighter variant may be more suitable. Overall, all 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*.

Enzymatic Reporter Proteins

Enzymatic reporters catalyse a (preferably) unique reaction either *in vivo* or *in vitro* and the specific activity is calculated to estimate protein levels and thus gene expression. Enzymatic reporters 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, by definition, bulk population measurements.

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*). The system was first developed for use in *C. perfringens* (a medically relevant non-solventogenic bacterium)[58] and has since been used in *C. acetobutylicum*[59]. Chloramphenicol acetyltransferase is an enzyme that catalyzes the covalent attachment of an acetyl group from acetyl-CoA to chloramphenicol[60], 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[60, 61]. 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[62] (knockouts of which would be more laborious and would likely have growth and phenotypic effects) and the cost of the substrate acetyl-CoA.

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*)[62]. 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[63]. 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-glucoronidase, respectively), 4-MU emits light at 460 nm when excited by 365 nm light[64]. 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*[65] 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*$^{\mathrm{opt}}$) has been used successfully as a secreted reporter in *C. difficile* by addition of a zinc metalloprotease PPEP-1 signal sequence[66]. 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[67]; 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.

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[68]. Lipases are enzymes which hydrolyse the ester bonds found in long-chain acylglycerols, releasing fatty acids[69]. Activity can thus be assessed by an enzymatic assay measuring the cleavage of *p*-nitrophenyl-esters[69] such as *p*-nitrophenyl decanoate[68]. 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[70].

A colorimetric alkaline phosphatase assay was developed and used in *C. difficile* in 2015 based on the *phoZ* gene product from *Enterococcus faecalis*[71]. 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[71]. 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*).

Luciferase

The luciferase (*lucB*) reporter was also used in *C. acetobutylicum* successfully, allowing luminescence detection[72]. 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[66].

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 a choice of more than one reporter, enzymatic and fluorescent, and their systematic and comparative use within the genus. Such an approach could provide insight into the sources of gene expression variability between strains and provide a starting point for future engineering. To achieve this seamlessly, it is advisable to explore translational fusions between reporters; examples from *E. coli* include a FRET pair YFP-FbFP fusion[73] and

Gemini (lacZα-GFP) fusion[74]. If successfully applied to the *Clostridium* species, such bi-functional reporters have the potential to become a single standard reporter. Additionally, α-peptide complementation strategies using the *E. coli lacZ* gene in *B. subtilis*[75] and *Mycobacterium smegmatis*[76] 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[77] and is a dimer[78] (rather than a tetramer like *E. coli* LacZ).

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). 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.

Transcription initiation

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[79]. 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)[80,81]. 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[82], an additional *E. coli* promoter feature is the 'extended -10' region[83] (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[84], a feature that was not explicable through higher AT-content alone. In Clostridia additional sequences with resemblances to UP elements (termed phased A-tracts) have been described[85] and a conserved extended -10 region that differs from its *E. coli analogue* has also been proposed[86]. 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[87]. The most commonly used constitutive promoter in *C. acetobutylicum* (Table 1) is that of the native thiolase gene (*thlA*)[88,89], which has been used for the expression of genes involved in production of solvents such as butanol[90] and isopropanol[26], the transcriptional regulator gene *tetR*[91] (used a minimal promoter variant- *miniPthl*), and, in a modified form, for the expression of cellulosomal scaffoldins[92] and glycoside hydrolases[25,93]. The *C. acetobutylicum* thiolase enzyme catalyzes the condensation of acetoacetyl-CoA from acetyl-CoA which are central metabolic intermediates[88]. 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[89]); however, in its normally chromosomal context it is also the subject of regulation by the redox-sensing transcriptional repressor Rex[94]. Rex inactivation was found to increase native *thlA* activity in *C. acetobutylicum* about 12-fold[95]. The Rex-binding site has been omitted from the core promoter in commonly used synthetic constructs[96]; yet this binding site is conserved in the promoters of orthologous genes from other solventogens (according to the RegPrecise database[97] 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[48] and Girbal 2003[63] studies).

Thiolase promoters have also been used for expression in other clostridia such as *C. beijerinckii*[98] and *C. pasteurianum*[99]. 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[100]. 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[63], and the *C. perfringens bgaL* promoter, which is repressed by BglR in the absence of lactose[101]. 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[24,25,93]. 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[59]. A TetR-repressed, anhydrotetracycline-inducible promoter has also been used in *C. acetobutylicum*, generated by the fusion of the chloramphenicol acetyltransferase promoter *pcm* with the tetracycline operator *tetO*[91]. This promoter could achieve up to 313-fold induction, although high levels of anhydrotetracycline were inhibitory to growth. However, operators are known to influence the basal activity of the promoter[102]. 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[103]. An alternative strategy - CRISPR-mediated repression of transcription - has been demonstrated in several solventogenic species, namely C. *acetobutylicum*[104], *C. beijerinckii*[98] and *C. pasteurianum*[50],

Orthogonal expression systems, i.e. promoters that require other non-native elements for activity (commonly an alternative sigma factor or phage polymerase), first found widespread use in the cloning of genes with toxic products. 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[105]. 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[59]. 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*[86,106]. 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[107]. 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. *Promoters used in engineering of C. acetobutylicum*

Promoter	Comments
Constitutive	
***thlA* (*C. acetobutylicum*)**	Widely used for constitutive gene expression in *C. acetobutylicum* for example expression on *C. beijerinckii ald*, *C. ljungdahlii bdh*[90], *C. acetobutylicum adc*, *ctfA*, *ctfB*,[26] & *tetR* (*miniPthl* variant)[91]. Activity analyzed using GusA[63] and several FbFP[48]reporters
***thl* (*C. perfringens*)**	Used for expression of *srtA* genes from *C. acetobutylicum*, *L. monocytogenes*, and *B. cereus*[25].
***fdx* (*C. sporogenes*)**	Activity analyzed using CatP reporter[59]. Used in the ClosTron system for expression of the Ll.LtrB intron[108]. Used for expression of *spoA* integrated into the chromosome for complementation of a *spoA* mutant[109].
Controlled	
***thl* (*thl* with double *lac* operator)**	*lac*-repressed version of *thlA*. Used to express a miniscaffoldin *cipc1*[92] & weakened version for expression of mannanase *man5K*[93].
***thlOid* (*thl* with single *lac* operator)**	A *lac*-repressible version, used to drive chromosomal expression of the *C. cellulolyticum* glycoside hydrolase *xyn10A*[25].
***fac* (single *lac* operator, derived from *C. pasteurianum*)**	Activity analyzed using CatP reporter[59,110] and Pp2 FbFP[48]. Formerly used in the ClosTron system for expression of the Ll.LtrB intron[110]. Used for expression of *codA* in the *C. acetobutylicum* knockout vector pMTL-SC7515[109].
***fdxOid* (*fdx* with single *lac* operator)**	Used for chromosomal expression of *C. cellulolyticum* glycoside hydrolase *cel9G*[25]. Has RBS from *C. acetobutylicum thlA* promoter.
***facOid* (*fac* with single *lac* operator)**	Used for chromosomal expression of *C. cellulolyticum* glycoside hydrolase *cel48F*[25]. Has RBS from *C. acetobutylicum thlA* promoter.
***tcdB* (*C. difficile*)**	Developed for an orthogonal expression system, requires the exogenous sigma factor TcdR for function[59].
***xylA* (*S. xylosus*)**	Repressed by XylR and inducible by D-xylose; assessed with β-glucuronidase reporter giving 17-fold induction with D-xylose[63].
***pcm* (*tetO1*-containing variants)**	Repressed by TetR; inducible in the presence of anhydrotetracycline[91].
***bgaL* (*C. perfringens*)**	Repressed by BgaR and inducible by lactose[101].
***xyl/tetO* (*B. subtilis*)**	Repressed by TetR[111]; activity in *C. acetobutylicum* shown by phiLOV 2.1 Opt FbFP expression[55]
Fermentation phase-specific	
***adc* (*C. acetobutylicum*)**	Activity assessed using β-glucuronidase[63], β-galactosidase[72] and Pp2 FbFP[48]. Primarily active after onset of solventogenesis.
***ptb* (*C. acetobutylicum*)**	Activity assessed using β-glucuronidase[63], luciferase[72] and Pp2 FbFP[48]. Active during acidogenesis (not solventogenesis)
***ptb* (*C. beijerinckii*)**	Activity assessed through expression of *lacI*[59].
***sol* (*C. acetobutylicum*)**	Activity assessed using β-galactosidase & luciferase reporters[72]. Weak expression, primarily active during late exponential phase.
***bdhA* (*C. acetobutylicum*)**	Analyzed using β-galactosidase reporter[72]. Comparatively weak expression, primarily active in early exponential phase.
***bdhB* (*C. acetobutylicum*)**	Analyzed using β-galactosidase reporter[72]. Primarily active until onset of solventogenesis.
***hydA* (*C. acetobutylicum*)**	Activity analyzed using β-glucuronidase[63] and Pp2 FbFP[48]. High activity during acidogenesis, decreases to low after phase shift
***adhe2* (*C. acetobutylicum*)**	Activity analyzed using Pp2 FbFP[48]. Strong expression during solventogenesis[48,112]

Transcription termination

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[113]. The specific DNA recognition sites, *rut* sites, have been used before in genetic circuits[114], but not extensively, perhaps due to the relatively poor mechanistic understanding of the process[115]. The second mechanism, which also has been reported to be more widespread[116], is referred to as Rho-independent or intrinsic termination and relies on the folding of a short GC rich hairpin followed by a poly-U transcribed sequence. The hairpin folding and subsequent transcription of the U-tract causes the polymerase to pause and leads to transcript and RNAP dissociation[117]. This is also facilitated by the fact that ribo-uracil-deoxyribo-adenine is the weakest nucleotide base pair[118]. 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[119].

While a number of terminators have been used in the construction of clostridial expression constructs, often derived from clostridial genes such as *adc*[63,120], *fdx*[109], CD0164[109] and *thlA*[90,121], there have been few published analyses of terminator strengths in clostridia. 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[59], and screening of a selection of terminators in *C. acetobutylicum*[25] 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[122]. However, this effect can be quantified in an appropriately designed assay[123]. 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*[124]. 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[125]. A recent study demonstrated the gradual increase in termination efficiency as distance between the stop codon and hairpin increased[126]. 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[126]. Expanding the range of characterized clostridial terminators is clearly necessary for the development of a clostridial synthetic biology toolbox.

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[116,127–129]. These bioinformatics tools were developed to aid genome annotations by identifying the 3' ends of operons and may not then detect all the features required by a functional terminator. In order to evaluate the relative difference in sequences retrieved by different programs, we compared the output of three searches performed on the *C. acetobutylicum* ATCC 824 genome (Figure 3A). The three search algorithms that were chosen are as follows: TransTermHP (TTHP)[128], RNIE[116] and WebGeSTer (WG)[130]. TTHP is a widely used tool, while RNIE's authors state that their tool eliminates false positives in comparison to TTHP[116]. Finally, WebGeSTer was selected because the authors classify 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[129].

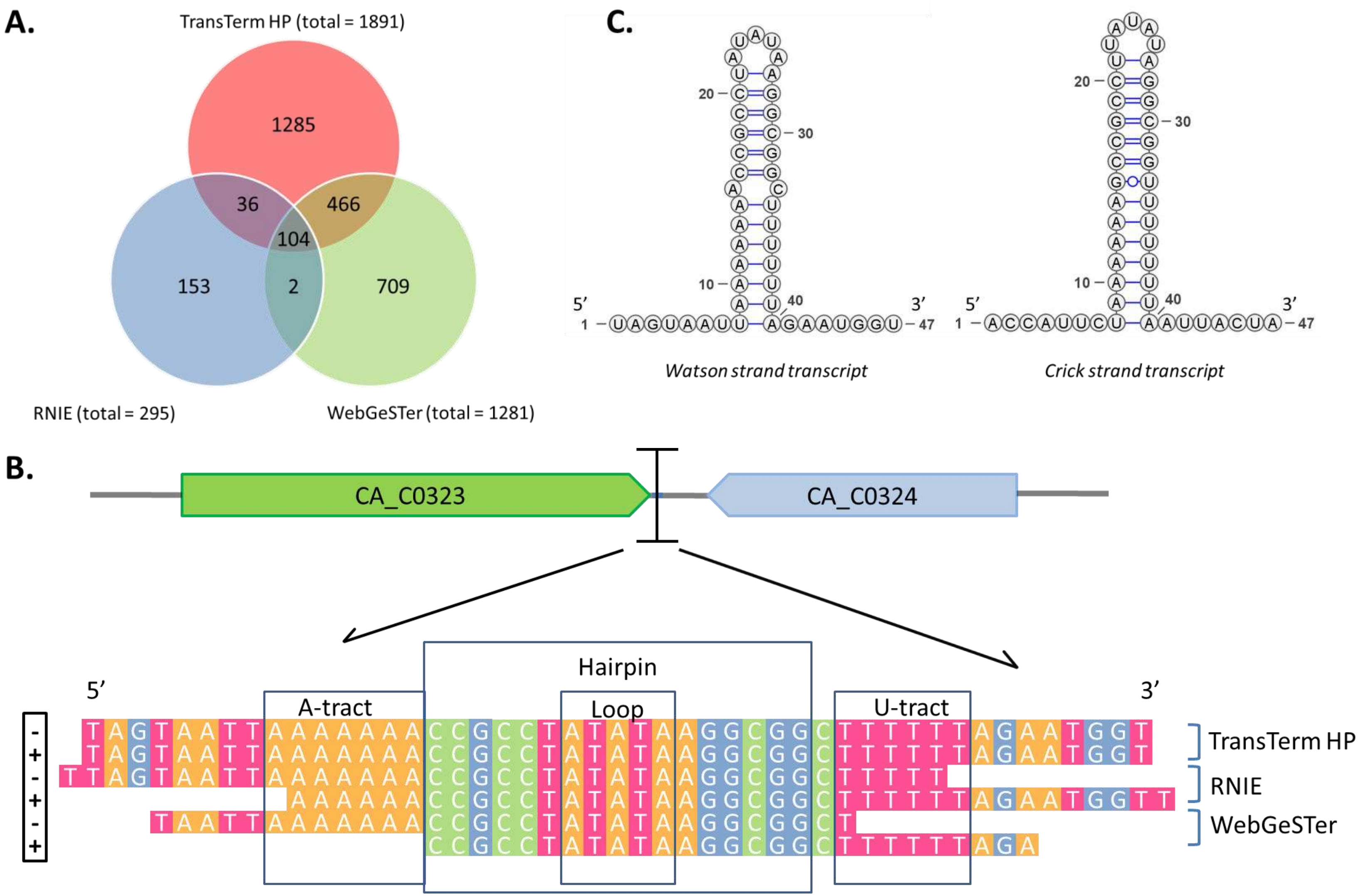


Figure 3. *Application of terminator prediction software to C. acetobutylicum.* ***A****. Outputs from predicted terminator sets derived from the C. acetobutylicum genome by application of TransTermHP (TTHP), RNIE and WedGeSTer were compared.* ***B****. An alignment of predictions of a putative bi-directional terminator (marked by T sign) by all three algorithms (note all three produce separate predictions for plus (+)and minus (-))****C****. RNA structures predicted by RNAfold[215,216] for the TransTermHP sequence, drawn with VARNA[217]; it is worth noting that the base-pairing between the A-tract and U-tract may not form until after termination[218]. Terminator lengths: WG: maximum length (max)-48bp, minimum (min) - 28bp, average (avg) - 35.8bp; RNIE: max-45bp, min- 34bp, avg-41.7bp,TTHP: max-89bp,min-41bp,avg-57.2bp.*

In our analysis, each algorithm was used to produce sets of putative terminator predictions, which were examined to identify overlapping predictions and are presented in Fig. 3A. Using the default settings of the different algorithms, the total number of predictions per program varied considerably, with TTHP and WG each predicting over 1000 terminators while RNIE predicted fewer than 300 (Fig. 3A). Despite this much smaller set, only about half of the predictions were shared with either TTHP and WG sets or both, producing a small set of 104 terminators shared by all three methods, an example of which is illustrated in Fig. 3B, an intergenic region separating two converging coding sequences, CA_C0323 (a sensory transduction histidine kinase) and CA_C0324 (a tetratricopeptide-repeats-containing protein). This also illustrates a class of terminators with particular utility in synthetic biology, namely bi-directional terminators, which are capable of terminating transcription from both the plus and minus strands. Our analysis finds that 37 bi-directional terminators are supported by all three algorithms, supporting the concepts that there are plentiful targets in clostridial genomes for 'part mining' and future experimental characterisation of these likely strong terminators. An interesting feature of the bi-directional terminators is their hypersymmetry[131] (can be seen in Fig. 3B). This leads to the possibility to form an extended hairpin between the so-called A-tract (can act as U-tract in the reverse transcriptional orientation) and the U-tract (as shown in Fig. 3C). However, as pausing and termination occur at the U-tract[117], the formation of an extended hairpin may not be relevant and does not appear to contribute to strength in all terminators[132]. Also, note that in this example the terminator on the minus strand (also referred to as Crick strand) is predicted to have a more stable hairpin (rG and rU form a basepair while rA and rC do not), whereas terminators from both strands are capable of forming an extended hairpin if fully transcribed as a result of base pairing between the A- and U-tracts. Combined with having a longer U-tract (Fig. 3C), the Crick strand terminator is possibly the stronger one, interestingly WebGeSTer predicts a smaller terminator downstream of the one depicted in Fig. 3A (not shown). A previous analysis of the *C. acetobutylicum* genome for terminators using TransTerm[127], a predecessor of TTHP, revealed the presence of 66 putative bi-directional terminators[133], supporting the rationale of combining outputs from multiple algorithms to create a small part list for experimental characterisation.

Translation

Translation initiation

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[134]. 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[25] that generated several new hybrid promoters. Others have experimented with the length of the spacer[135,136]. 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[137]) which acts as a guide for the ribosome complex enabling mRNA recognition by the translation machinery and translation initiation. 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[138,139]. 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[106]. Recently, a systematic analysis of SD-aSD pairings in *B. subtilis* and in *E. coli* confirmed this trend[140]. 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[137,141,142] as revealed by sequence alignment) is of particular importance for translation initiation[137]. 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*[143] (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*[144], whereas the spacers in *Bifidobacterium longum*[145] 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*[135,136]. 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[135].

Yang *et al.* (2016) demonstrated the potential benefits of utilising modified spacers in *C. acetobutylicum*[135]. 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. It is conceivable that not only the length but also the sequence of the spacer might influence translation initiation, for example the introduction of a secondary SD within the primary SD's spacer region.

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[146,147]. There are reports that *de novo* design produces more accurate results than translation initiation rate prediction of natural sequences for Gram positives[148].

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[149,150], has been shown to strongly influence heterologous protein expression levels[151,152].

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*[153] and there are also bioinformatically observable variations within the *Clostridium* genus itself[154] 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*)[63], *T. thermosulfurigenes* (*lacZ*)[72] and the firefly *Photinus pyralis* (*lucB*)[72] (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* [55], as well as bespoke *C. acetobutylicum* codon-optimized ones[24,48]. Researchers have also codon-optimized several GFP-like proteins for use in Firmicutes[155] (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)[150] 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[156]. 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[157,158]. 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[159,160]. In addition, condition-specific usage tables have also been reported[161]. 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[158,162] 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[163]. 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.

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.

Controlling 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[122]. The presence of secondary structures at the 5' end of the mRNA has been observed to provide an increase in mRNA stability[122]. 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)[164]. 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[165]. Nevertheless, *C. acetobutylicum* also has a homolog of RNase Y[166], which fulfils the role of RNase E in *B. subtilis*[167]. 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*[168]. 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[122]. 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[168]. Additionally, a terminator with activity in the reverse orientation prevents the formation of antisense transcripts which are known to reduce protein expression levels[169].

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[170]. 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.

Antisense RNAs and Riboswitches

Bacterial antisense RNAs (asRNAs) are transcripts with complementarity to another RNA. Several studies have used asRNAs were used to reduce native gene expression in *C. acetobutylicum*[169,171] and in *C. pasteurianum*[172], 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[173]. They have defined sensor and effector domains and have been rationally designed to repress[174] and activate gene expression[175] 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[176]. 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[177]. 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 yet others have been shown to control access to recognition sequences such as RNase sites[178]. 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[179]. Riboswitches are found in all taxa and a number have been characterized in the solventogenic clostridia[180–182]. 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.

Replication

Plasmid origins of replication

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*)[96]. The pIM13 replication origin is thought to replicate via rolling-circle replication[183] while there is evidence that pCD6 replicates in similar fashion to pIP404[184,185] (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[186]; the pUB110 replicon was found to be somewhat more stable than pIM13, whereas the pAMβ1 replicon was highly unstable. The *C. beijerinckii* filamentous phage CAK1's origin of replication has been used in *C. beijerinckii* strains[187]. Additionally, the development of a replicon specific for *C. saccharoperbutylacetonicum* N1-4 was reported in 2007[188], this replicon is identical to the origin of the endogenous plasmid from *C. saccharoperbutylacetonicum* N1-504[15] . A thermosensitive origin pWV01ts derived from *Lactococcus lactis cremoris*[189] has been shown to work in both *C. ljungdahlii* and *C. acetobutylicum*[51]. Segregation and transformation frequencies are available; however, more work is needed to determine copy number and compatibility groups.

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[190–192]. 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. There is contradictory evidence as to the effects of chromosome location and levels of expression[193],[194,195]. The cause of some of the apparent transcriptional insulation of chromosomal genes (and thus promoter independence) was found to be due to the gene silencing activity of the HU-protein in *E. coli*[194], a protein that is also present in

Firmicutes[196]. Genomic engineering in *E. coli* has been carried out extensively; lambda-Red recombineering[197] is well-established, and the utility of the newly developed CRISPR technique has been demonstrated in this species[198,199]. However, in the solventogenic clostridia, until recently published examples of genomic integration had been still somewhat limited. An early enabling technology was ClosTron which adapted the Ll.LtrB intron for use in *Clostridium*[110]. Another method for genomic integration is Allele-Coupled Exchange (ACE), as demonstrated in *C. acetobutylicum*[24,25,190]. 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; currently, this involves either 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* [190]. 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*[101,109,200,201] and *C. beijerinckii*[202]; while most of these studies have focused on the generation of in-frame deletions and subsequent complementation, Al-Hinai et al.[101] demonstrate the integration of a heterologous gene through gene replacement. Furthermore, the generation of point mutations through recombineering has been demonstrated in *C. acetobutylicum*[203], 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*[121,204], *C. acetobutylicum*[104], *C. pasteurianum*[205], and *C. saccharoperbutylacetonicum*[206]. Future developments are certain to expand the range of genomic modifications that can be made in these organisms.

Insulators

A biological part's adjacent sequences can have a profound effect on its behaviour compared to the sequence context in which it was characterized[207]. This poses a fundamental challenge to the synthetic biology principle of part creation and characterization[208]. To counteract this issue in reproducibility genetic engineers have started utilizing a new class of parts called insulators[209]. 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[123] 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)[208,210].

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.[211] Accordingly, the addition of an upstream and downstream insulating sequence has increased reproducibility in different genetic contexts[212]. 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[91].

Work in *E. coli* has demonstrated the utility of 'bicistronic 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[212]. 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 with 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[126]. Such strategies are yet to be implemented in *Clostridium* engineering projects.

Discussion

The development of biorenewables as an alternative to petroleum-derived commodity chemicals and fuels has resulted in the emergence of new markets[213]. Increasing productivity, broadening the range of feedstocks, improving tolerance to solvents and by-products are all existing challenges to achieving higher sustainability and ensuring the economic viability of *Clostridium*-derived biorenewables[214]. *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 research. 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 analyzing the behaviour of pathways and genetic circuits. We have shown in this review that there are multiple areas where improved knowledge of parts would provide a better toolkit for synthetic biologists using these organisms and consequently improve the rate at which industrial biotechnology and bioenergy process development can improve.

References

(1) Breed, R. S., Murray, E. G. D., and Smith, N. R. (1957) Bergey's manual of determinative bacteriology 7th ed. Williams & Wilkins Co., Baltimore.

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

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

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

(5) Keis, S., Shaheen, R., and Jones, D. T. (2001) 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.

(6) Hellinger, E. (1947) *Clostridium aurantibutyricum* (n.sp.): A Pink Butyric Acid *Clostridium*. *J. Gen. Microbiol. 1*, 203–210.

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

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

(9) Sarchami, T., Munch, G., Johnson, E., Kießlich, S., and Rehmann, L. (2016) A Review of Process-Design Challenges for Industrial Fermentation of Butanol from Crude Glycerol by Non-Biphasic *Clostridium pasteurianum*. *Fermentation 2*, 13.

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

(11) Tamura, K., and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol. 10*, 512–26.

(12) Felsenstein, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution (N. Y). 39*, 783–791.

(13) Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol. 28*, 2731–9.

(14) ABAS. (2015) Classification of prokaryotes (bacteria and archaea) into risk groups (TRBA 466). *Bundesarbeitsblatt. http://www.baua.de* 19.

(15) Poehlein, A., Solano, J. D. M., Flitsch, S. K., Krabben, P., Winzer, K., Reid, S. J., Jones, D. T., Green, E., Minton, N. P., Daniel, R., and Dürre, P. (2017) Microbial solvent formation revisited by comparative genome analysis. *Biotechnol. Biofuels 10*, 58.

(16) Udaondo, Z., Duque, E., and Ramos, J. L. (2017) The pangenome of the genus *Clostridium*. *Environ. Microbiol. 19*, 2588–2603.

(17) Shaheen, R., Shirley, M., and Jones, D. T. (2000) Comparative fermentation studies of industrial strains belonging to four species of solvent-producing clostridia. *J. Mol. Microbiol. Biotechnol. 2*, 115–124.

(18) Jang, Y.-S., Lee, J. Y., Lee, J., Park, J. H., Im, J. A., Eom, M.-H., Lee, J., Lee, S.-H., Song, H., Cho, J.-H., Seung, D. Y., and Lee, S. Y. (2012) Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *MBio 3*, e00314-12.

(19) Pyne, M. E., Sokolenko, S., Liu, X., Srirangan, K., Bruder, M. R., Aucoin, M. G., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016) Disruption of the reductive 1,3-propanediol pathway triggers production of 1,2-propanediol for sustained glycerol fermentation by *Clostridium pasteurianum*. *Appl. Environ. Microbiol. 82*, 5375–5388.

(20) Xiao, H., Li, Z., Jiang, Y., Yang, Y., Jiang, W., Gu, Y., and Yang, S. (2012) Metabolic engineering of D-xylose pathway in *Clostridium beijerinckii* to optimize solvent production from xylose mother liquid. *Metab. Eng. 14*, 569–578.

(21) Xiao, H., Gu, Y., Ning, Y., Yang, Y., Mitchell, W. J., Jiang, W., and Yang, S. (2011) 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.

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

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

(24) Kovács, K., Willson, B. J., Schwarz, K., Heap, J. T., Jackson, A., Bolam, D. N. N., Winzer, K., and Minton, N. P. (2013) Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels 6*, 117.

(25) Willson, B. J., Kovács, K., Wilding-Steele, T., Markus, R., Winzer, K., and Minton, N. P. (2016) Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels 9*, 109.

(26) Collas, F., Kuit, W., Clément, B., Marchal, R., López-Contreras, A. M., and Monot, F. (2012) Simultaneous production of isopropanol, butanol, ethanol and 2,3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains. *AMB Express 2*, 45.

(27) González-Pajuelo, M., Meynial-Salles, I., Mendes, F., Andrade, J. C., Vasconcelos, I., and Soucaille, P. (2005) Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1,3-propanediol from glycerol. *Metab. Eng. 7*, 329–336.

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

(29) Quinn, J. Y., Cox, R. S., Adler, A., Beal, J., Bhatia, S., Cai, Y., Chen, J., Clancy, K., Galdzicki, M., Hillson, N. J., Le Novère, N., Maheshwari, A. J., McLaughlin, J. A., Myers, C. J., Umesh, P., Pocock, M., Rodriguez, C., Soldatova, L., Stan, G. B. V., Swainston, N., Wipat, A., and Sauro, H. M. (2015) SBOL Visual: A Graphical Language for Genetic Designs. *PLoS Biol. 13*, e1002310.

(30) Mermelstein, L. D., and Papoutsakis, E. T. (1993) *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage ϕ3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol. 59*, 1077–1081.

(31) Nölling, J., Breton, G., Omelchenko, M. V, Makarova, K. S., Zeng, Q., Gibson, G., Hong Mei Lee, Dubois, J., Qiu, D., Hitti, J., Aldredge, T., Ayers, M., Bashirzadeh, R., Bochner, H., Boivin, M., Bross, S., Bush, D., Butler, C., Caron, A., Caruso, A., Cook, R., Daggett, P., Deloughery, C., Egan, J., Ellston, D., Engelstein, M., Ezedi, J., Gilbert, K., Goyal, A., Guerin, J., Ho, T., Holtham, K., Joseph, P., Keagle, P., Kozlovsky, J., LaPlante, M., LeBlanc, G., Lumm, W., Majeski, A., McDougall, S., Mank, P., Mao, J. I., Nocco, D., Patwell, D., Phillips, J., Pothier, B., Prabhakar, S., Richterich, P., Rice, P., Rosetti, D., Rossetti, M., Rubenfield, M., Sachdeva, M., Snell, P., Spadafora, R., Spitzer, L., Shimer, G., Thomann, H. U., Vicaire, R., Wall, K.,

Wang, Y., Weinstock, K., Lai Peng Wong, Wonsey, A., Xu, Q., Zhang, L., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V, and Smith, D. R. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol. 183*, 4823–4838.

(32) Oultram, J. D., Loughlin, M., Swinfield, T.-J. J., Brehm, J. K., Thompson, D. E., and Minton, N. P. (1988) Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiol. Lett. 56*, 83–88.

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

(34) Lesiak, J. M., Liebl, W., and Ehrenreich, A. (2014) 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.

(35) Herman, N. A., Li, J., Bedi, R., Turchi, B., Liu, X., Miller, M. J., and Zhang, W. (2017) 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.

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

(37) Kolek, J., Sedlar, K., Provaznik, I., and Patakova, P. (2016) 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.

(38) Oh, Y. H., Eom, G. T., Kang, K. H., Choi, J. W., Song, B. K., Lee, S. H., and Park, S. J. (2015) Optimized Transformation of Newly Constructed *Escherichia coli*-Clostridia Shuttle Vectors into *Clostridium beijerinckii*. *Appl. Biochem. Biotechnol. 177*, 226–236.

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

(40) Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., Liew, F., Kelly, M. L., Sheng, L., Schwarz, K., and Zhang, Y. (2016) A roadmap for gene system development in *Clostridium*. *Anaerobe 41*, 104–112.

(41) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods 2*, 905–909.

(42) Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauß, U., Guterl, J.-K., Wendorff, M., Losi, A., Gärtner, W., and Jaeger, K.-E. (2007) Reporter proteins for *in vivo* fluorescence without oxygen. *Nat. Biotechnol. 25*, 443–445.

(43) Jean, D., Briolat, V., and Reysset, G. (2004) Oxidative stress response in *Clostridium perfringens*. *Microbiology 150*, 1649–1659.

(44) Hartman, A. H., Liu, H., and Melville, S. B. (2011) Construction and Characterization of a Lactose-Inducible Promoter System for Controlled Gene Expression in *Clostridium perfringens*. *Appl. Environ. Microbiol. 77*, 471–478.

(45) Ransom, E. M., Williams, K. B., Weiss, D. S., and Ellermeier, C. D. (2014) Identification and characterization of a gene cluster required for proper rod shape, cell division, and pathogenesis in *Clostridium difficile*. *J. Bacteriol. 196*, 2290–2300.

(46) Ransom, E. M., Ellermeier, C. D., and Weiss, D. S. (2015) Use of mCherry red fluorescent protein for studies of protein localization and gene expression in *Clostridium difficile*. *Appl. Environ. Microbiol. 81*, 1652–1660.

(47) Ransom, E. M., Weiss, D. S., and Ellermeier, C. D. (2016) Use of mCherryOpt fluorescent protein in *Clostridium difficile*. *Methods Mol. Biol. 1476*, 53–67.

(48) Schulz, F. (2013) Fluoreszenzproteine in *Clostridium acetobutylicum* - Ein neues *in vivo* Reportersystem. Universität Rostock.

(49) Li, Y., Xu, T., Tschaplinski, T. J., Engle, N. L., Yang, Y., Graham, D. E., He, Z., and Zhou, J. (2014) Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation abolishment and carbon alleviation. *Biotechnol. Biofuels 7*, 25.

(50) Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016) Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus *Clostridium*. *Appl. Environ. Microbiol. 82*, 6109–6119.

(51) Molitor, B., Kirchner, K., Henrich, A. W., Schmitz, S., and Rosenbaum, M. A. (2016) Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Sci. Rep. 6*, 31518.

(52) Chapman, S., Faulkner, C., Kaiserli, E., Garcia-Mata, C., Savenkov, E. I., Roberts, A. G., Oparka, K. J., and Christie, J. M. (2008) The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci. 105*, 20038–20043.

(53) Christie, J. M., Hitomi, K., Arvai, A. S., Hartfield, K. a., Mettlen, M., Pratt, A. J., Tainer, J. a., and Getzoff, E. D. (2012) Structural tuning of the fluorescent protein iLOV for improved photostability. *J. Biol. Chem. 287*, 22295–22304.

(54) Gawthorne, J. A., Reddick, L. E., Akpunarlieva, S. N., Beckham, K. S. H., Christie, J. M., Alto, N. M., Gabrielsen, M., and Roe, A. J. (2012) Express Your LOV: An Engineered Flavoprotein as a Reporter for Protein Expression and Purification. *PLoS One 7*, e52962.

(55) Buckley, A. M., Jukes, C., Candlish, D., Irvine, J. J., Spencer, J., Fagan, R. P., Roe, A. J., Christie, J. M., Fairweather, N. F., and Douce, G. R. (2016) Lighting Up *Clostridium difficile*: Reporting Gene Expression Using Fluorescent Lov Domains. *Sci. Rep. 6*, 23463.

(56) Khrenova, M. G., Nemukhin, A. V., and Domratcheva, T. (2015) Theoretical characterization of the flavin-based fluorescent protein iLOV and its Q489K mutant. *J. Phys. Chem. B 119*, 5176–5183.

(57) Davari, M. D., Kopka, B., Wingen, M., Bocola, M., Drepper, T., Jaeger, K. E., Schwaneberg, U., and Krauss, U. (2016) Photophysics of the LOV-Based Fluorescent Protein Variant iLOV-Q489K Determined by Simulation and Experiment. *J. Phys. Chem. B 120*, 3344–3352.

(58) Bullifent, H. L., Moir, A., and Titball, R. W. (1995) The construction of a reporter system and use for the investigation of *Clostridium perfringens* gene expression. *FEMS Microbiol. Lett. 131*, 99–105.

(59) Zhang, Y., Grosse-Honebrink, A., and Minton, N. P. (2015) A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One 10*, e0122411.

(60) Shaw, W. V. (1975) Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods Enzymol. 43*, 737–755.

(61) Roberts, M. C., Swenson, C. D., Owens, L. M., and Smith, A. L. (1980) Characterization of chloramphenicol-resistant *Haemophilus influenzae*. *Antimicrob. Agents Chemother. 18*, 610–615.

(62) Tummala, S. B., Welker, N. E., and Papoutsakis, E. T. (1999) Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol. 65*, 3793–3799.

(63) Girbal, L., Mortier-Barrière, I., Raynaud, F., Rouanet, C., Croux, C., and Soucaille, P. (2003) Development of a sensitive gene expression reporter system and an inducible promoter-repressor system for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol. 69*,

4985–4988.

(64) Moberg, L. J. (1985) Fluorogenic assay for rapid detection of *Escherichia coli* in food. *Appl. Environ. Microbiol. 50*, 1383–1387.

(65) Sabathé, F., Croux, C., Cornillot, E., and Soucaille, P. (2002) *amyP*, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol. Lett. 210*, 93–98.

(66) Oliveira Paiva, A. M., Friggen, A. H., Hossein-Javaheri, S., and Smits, W. K. (2016) 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.

(67) Quixley, K. W., and Reid, S. J. (2000) Construction of a reporter gene vector for *Clostridium beijerinckii* using a *Clostridium* endoglucanase gene. *J. Mol. Microbiol. Biotechnol. 2*, 53–7.

(68) Oh, Y. H., Eom, G. T., Kang, K. H., Joo, J. C., Jang, Y. A., Choi, J. W., Song, B. K., Lee, S. H., and Park, S. J. (2016) Construction of heterologous gene expression cassettes for the development of recombinant *Clostridium beijerinckii*. *Bioprocess Biosyst. Eng. 39*, 555–563.

(69) Gilham, D., and Lehner, R. (2005) Techniques to measure lipase and esterase activity *in vitro*. *Methods 36*, 139–147.

(70) Hassan, E. A., Abd-Alla, M. H., Bagy, M. M. K., and Morsy, F. M. (2015) In situ hydrogen, acetone, butanol, ethanol and microdiesel production by *Clostridium acetobutylicum* ATCC 824 from oleaginous fungal biomass. *Anaerobe 34*, 125–131.

(71) Edwards, A. N., Pascual, R. A., Childress, K. O., Nawrocki, K. L., Woods, E. C., and McBride, S. M. (2015) An alkaline phosphatase reporter for use in *Clostridium difficile*. *Anaerobe 32*, 98–104.

(72) Feustel, L., Nakotte, S., and Dürre, P. (2004) Characterization and development of two reporter gene systems for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol. 70*, 798–803.

(73) Potzkei, J., Kunze, M., Drepper, T., Gensch, T., Jaeger, K.-E., and Buechs, J. (2012) Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor. *BMC Biol. 10*, 28.

(74) Martin, L., Che, A., and Endy, D. (2009) Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS One 4*.

(75) Haima, P., van Sinderen, D., Bron, S., and Venema, G. (1990) An improved β-galactosidase α-complementation system for molecular cloning in *Bacillus subtilis*. *Gene 93*, 41–47.

(76) Knipfer, N., Nooruddin, L., and Shrader, T. E. (1998) Development of an alpha-complementation system for mycobacterial promoter analysis. *Gene 217*, 69–75.

(77) Volkov, I. Y., Lunina, N. A., Berezina, O. V., Velikodvorskaya, G. A., and Zverlov, V. V. (2005) *Thermoanaerobacter ethanolicus* gene cluster containing the α- and β-galactosidase genes *melA* and *lacA* and properties of recombinant LacA. *Mol. Biol. 39*, 799–805.

(78) Huber, R. E., Roth, N. J., and Bahl, H. (1996) Quaternary structure, Mg2+ interactions, and some kinetic properties of the β-galactosidase from *Thermoanaerobacterium thermosulfurigenes* EM1. *J. Protein Chem. 15*, 621–629.

(79) Browning, D. F., and Busby, S. J. W. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol. 2*, 57–65.

(80) Helmann, J. D., and Chamberlin, M. J. (1988) Structure and function of bacterial sigma factors. *Annu. Rev. Biochem. 57*, 839–872.

(81) Sauer, U., Santangelo, J. D., Treuner, A., Buchholz, M., and Dürre, P. (1995) Sigma factor and sporulation genes in *Clostridium*. *FEMS Microbiol. Rev. 17*, 331–340.

(82) Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. (1998) Identification of an UP element consensus sequence for bacterial promoters. *Biochemistry 95*, 9761–9766.

(83) Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A., and Hayward, R. S. (1993) 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.

(84) Sinoquet, C., Demey, S., and Braun, F. (2008) Large-scale computational and statistical analyses of high transcription potentialities in 32 prokaryotic genomes. *Nucleic Acids Res. 36*, 3332–3340.

(85) Katayama, S., Matsushita, O., Jung, C. M., Minami, J., and Okabe, A. (1999) 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.

(86) Graves, M. C., and Rabinowitz, J. C. (1986) *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.

(87) Pyne, M. E. (2014, April) Development of genetic tools for metabolic engineering of

Clostridium pasteurianum. University of Waterloo.

(88) Wiesenborn, D. P., Rudolph, F. B., and Papoutsakis, E. T. (1988) Thiolase from *Clostridium acetobutylicum* ATCC 824 and Its Role in the Synthesis of Acids and Solvents. *Appl Env. Microbiol 54*, 2717–2722.

(89) Stim-Herndon, K. P., Petersen, D. J., and Bennett, G. N. (1995) Characterization of an acetyl-CoA C-acetyltransferase (thiolase) gene from *Clostridium acetobutylicum* ATCC 824. *Gene*, pp 81–85.

(90) Bormann, S., Baer, Z. C., Sreekumar, S., Kuchenreuther, J. M., Dean Toste, F., Blanch, H. W., and Clark, D. S. (2014) Engineering *Clostridium acetobutylicum* for production of kerosene and diesel blendstock precursors. *Metab. Eng. 25*, 124–130.

(91) Dong, H., Tao, W., Zhang, Y., and Li, Y. (2012) Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab. Eng. 14*, 59–67.

(92) Perret, S., Casalot, L., Fierobe, H. P., Tardif, C., Sabathe, F., Belaich, J. P., and Belaich, A. (2004) Production of Heterologous and Chimeric Scaffoldins by *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol. 186*, 253–257.

(93) Mingardon, F., Perret, S., Bélaïch, A., Tardif, C., Bélaïch, J. P., and Fierobe, H. P. (2005) Heterologous production, assembly, and secretion of a minicellulosome by *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol. 71*, 1215–1222.

(94) Wietzke, M., and Bahl, H. (2012) The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol. 96*, 749–761.

(95) Zhang, L., Nie, X., Ravcheev, D. A., Rodionov, D. A., Sheng, J., Gu, Y., Yang, S., Jiang, W., and Yang, C. (2014) Redox-responsive repressor rex modulates alcohol production and oxidative stress tolerance in *Clostridium acetobutylicum*. *J. Bacteriol. 196*, 3949–3963.

(96) Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009) A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods 78*, 79–85.

(97) Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., Kazanov, M. D., Riehl, W., Arkin, A. P., Dubchak, I., and Rodionov, D. A. (2013) RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics 14*, 745.

(98) Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., and Blaschek, H. P. (2016) Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. *Biotechnol.*

Bioeng. 113, 2739–2743.

(99) Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2014) Expansion of the genetic toolkit for metabolic engineering of *Clostridium pasteurianum*: chromosomal gene disruption of the endogenous CpaAI restriction enzyme. *Biotechnol. Biofuels 7*, 163.

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

(101) Al-Hinai, M. A., Fast, A. G., and Papoutsakis, E. T. (2012) 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.

(102) Garcia, H. G., Sanchez, A., Boedicker, J. Q., Osborne, M., Gelles, J., Kondev, J., and Phillips, R. (2012) Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep. 2*, 150–161.

(103) Zong, Y., Zhang, H. M., Lyu, C., Ji, X., Hou, J., Guo, X., Ouyang, Q., and Lou, C. (2017) Insulated transcriptional elements enable precise design of genetic circuits. *Nat. Commun. 8*, 52.

(104) Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., and Yang, S. (2016) CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol. J. 11*, 961–972.

(105) Tabor, S., and Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes (T7 DNA polymerase/T7 gene 5 protein/proteolysis/β-lactamase/rifampicin). *Biochemistry 82*, 1074–1078.

(106) McLaughlin, J. R., Murray, C. L., and Rabinowitz, J. C. (1981) Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* beta-lactamase gene. *J. Biol. Chem. 256*, 11283–91.

(107) Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., Kapanidis, A. N., Wade, J. T., and Grainger, D. C. (2017) Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol. 2*, 16249.

(108) Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C., and

Minton, N. P. (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods 80*, 49–55.

(109) Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S. T., Heap, J. T., Winzer, K., and Minton, N. P. (2016) Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels 9*, 4.

(110) Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., and Minton, N. P. (2007) The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods 70*, 452–464.

(111) Geissendörfer, M., and Hillen, W. (1990) Regulated expression of heterologous genes in *Bacillus subtilis* using the Tn10 encoded *tet* regulatory elements. *Appl. Microbiol. Biotechnol. 33*, 657–63.

(112) Fontaine, L., Meynial-Salles, I., Girbal, L., Yang, X., Croux, C., and Soucaille, P. (2002) 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*, 821–830.

(113) Ciampi, M. S. (2006) Rho-dependent terminators and transcription termination. *Microbiology 152*, 2515–2528.

(114) Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., Mutalik, V. K., and Arkin, A. P. (2012) An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods 9*, 1088–1094.

(115) Banerjee, S., Chalissery, J., Bandey, I., and Sen, R. (2006) Rho-dependent transcription termination: more questions than answers. *J. Microbiol. 44*, 11–22.

(116) Gardner, P. P., Barquist, L., Bateman, A., Nawrocki, E. P., and Weinberg, Z. (2011) RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res. 39*, 5845–5852.

(117) Gusarov, I., and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol. Cell 3*, 495–504.

(118) Sugimoto, N., Nakano, S. ichi, Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995) Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry 34*, 11211–11216.

(119) Naville, M., and Gautheret, D. (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol. 11*, R97.

(120) Ravagnani, A., Jennert, K. C. B., Steiner, E., Grünberg, R., Jefferies, J. R., Wilkinson, S. R., Young, D. I., Tidswell, E. C., Brown, D. P., Youngman, P., Gareth Morris, J., and Young, M. (2000) Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Mol. Microbiol. 37*, 1172–1185.

(121) Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., and Blaschek, H. P. (2016) 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.

(122) Carrier, T. A., and Keasling, J. D. (1997) Controlling messenger RNA stability in bacteria: Strategies for engineering gene expression. *Biotechnol. Prog. 13*, 699–708.

(123) Chen, Y.-J., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., and Voigt, C. A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods 10*, 659–664.

(124) Lin, M. T., Wang, C. Y., Xie, H. J., Cheung, C. H. Y., Hsieh, C. H., Juan, H. F., Chen, B. Sen, and Lin, C. (2016) Novel Utilization of Terminators in the Design of Biologically Adjustable Synthetic Filters. *ACS Synth. Biol. 5*, 365–374.

(125) Abe, H., Abo, T., and Aiba, H. (1999) Regulation of intrinsic terminator by translation in *Escherichia coli*: transcription termination at a distance downstream. *Genes Cells 4*, 87–97.

(126) Li, R., Zhang, Q., Li, J., and Shi, H. (2015) Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. *Nucleic Acids Res. 44*, 2554–2563.

(127) Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A., and Ecker, D. J. (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res. 29*, 3583–94.

(128) Kingsford, C. L., Ayanbule, K., and Salzberg, S. L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol. 8*, R22.

(129) Mitra, A., Angamuthu, K., Jayashree, H. V., and Nagaraja, V. (2009) Occurrence, divergence and evolution of intrinsic terminators across Eubacteria. *Genomics 94*, 110–116.

(130) Mitra, A., Kesarwani, A. K., Pal, D., and Nagaraja, V. (2011) WebGeSTer DB-A transcription terminator database. *Nucleic Acids Res. 39*, D129-35.

(131) Wright, J. J., Kumar, A., and Hayward, R. S. (1992) Hypersymmetry in a transcriptional terminator of *Escherichia coli* confers increased efficiency as well as bidirectionality. *EMBO J. 11*, 1957–64.

(132) Yang, M. Te, Scott, H. B., and Gardner, J. F. (1995) Transcription termination at the *thr* attenuator: Evidence that the adenine residues upstream of the stem and loop structure are not required for termination. *J. Biol. Chem. 270*, 23330–23336.

(133) Paredes, C. J., Rigoutsos, I., and Papoutsakis, T. (2004) Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Res. 32*, 1973–1981.

(134) Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001) Anatomy of *Escherichia coli* ribosome binding sites. *J. Mol. Biol. 313*, 215–228.

(135) Yang, Y., Lang, N., Yang, G., Yang, S., Jiang, W., and Gu, Y. (2016) Improving the performance of solventogenic clostridia by reinforcing the biotin synthetic pathway. *Metab. Eng. 35*, 121–128.

(136) Yang, G., Jia, D., Jin, L., Jiang, Y., Wang, Y., Jiang, W., and Gu, Y. (2017) Rapid Generation of Universal Synthetic Promoters for Controlled Gene Expression in Both Gas-Fermenting and Saccharolytic *Clostridium* Species. *ACS Synth. Biol. 6*, 1672–1678.

(137) Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) 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.

(138) Hager, P. W., and Rabinowitz, J. C. (1985) Translational specificity in *Bacillus subtilis*. *Mol. Biol. bacilli 2*, 1–32.

(139) Band, L., and Henner, D. J. (1984) *Bacillus subtilis* requires a “stringent” Shine-Dalgarno region for gene expression. *DNA 3*, 17–21.

(140) Abolbaghaei, A., Silke, J. R., and Xia, X. (2017) How Changes in Anti-SD Sequences Would Affect SD Sequences in *Escherichia coli* and *Bacillus subtilis*. *G3 (Bethesda). 7*, 1607–1615.

(141) Dunn, J. J., and Studier, F. W. (1981) Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol. 148*, 303–330.

(142) Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol. 6*, 1219–1229.

(143) Garnier, T., Canard, B., and Cole, S. T. (1991) Cloning, mapping, and molecular

characterization of the rRNA operons of *Clostridium perfringens*. *J. Bacteriol. 173*, 5431–5438.

(144) Ma, J., Campbell, A., and Karlin, S. (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol. 184*, 5733–5745.

(145) He, J., Sakaguchi, K., and Suzuki, T. (2012) 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.

(146) Salis, H. M. (2011) The ribosome binding site calculator. *Methods Enzymol. 498*, 19–42.

(147) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946–950.

(148) Reeve, B., Martinez-Klimova, E., De Jonghe, J., Leak, D. J., and Ellis, T. (2016) The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synth. Biol. 5*, 1342–1347.

(149) Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol. 2*, 13–34.

(150) Sharp, P. M., and Li, W. H. (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res. 15*, 1281–1295.

(151) Baca, A. M., and Hol, W. G. J. (2000) 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.

(152) Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations 12*, 4–6.

(153) Mondal, S. K., Kundu, S., Das, R., and Roy, S. (2016) 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. 34*, 1649–1666.

(154) Musto, H., Romero, H., and Zavala, A. (2003) Translational selection is operative for synonymous codon usage in *Clostridium perfringens* and *Clostridium acetobutylicum*. *Microbiology 149*, 855–863.

(155) Sastalla, I., Chim, K., Cheung, G. Y. C., Pomerantsev, A. P., and Leppla, S. H. (2009) Codon-optimized fluorescent proteins designed for expression in low-GC gram-positive

bacteria. *Appl. Environ. Microbiol. 75*, 2099–2110.

(156) Wright, F. (1990) The "effective number of codons" used in a gene. *Gene 87*, 23–29.

(157) Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science (80-.). 324*, 255–258.

(158) Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One 4*, e7002.

(159) Xia, X. (2007) An improved implementation of codon adaptation index. *Evol. Bioinforma. 3*, 53–58.

(160) Xia, X. (2015) A major controversy in codon-Anticodon adaptation resolved by a new codon usage index. *Genetics 199*, 573–579.

(161) Lanza, A. M., Curran, K. A., Rey, L. G., and Alper, H. S. (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. *BMC Syst. Biol. 8*, 33.

(162) Welch, M., Villalobos, A., Gustafsson, C., and Minshull, J. (2009) You're one in a googol: optimizing genes for protein expression. *J. R. Soc. Interface 6*, S467–S476.

(163) Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet. 12*, 32–42.

(164) Carrier, T. A., and Keasling, J. D. (1999) Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog. 15*, 58–64.

(165) Condon, C., and Putzer, H. (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res. 30*, 5339–5346.

(166) Durand, S., Tomasini, A., Braun, F., Condon, C., and Romby, P. (2015) sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol. Rev. 39*, 316–330.

(167) Lehnik-Habrink, M., Newman, J., Rothe, F. M., Solovyova, A. S., Rodrigues, C., Herzberg, C., Commichau, F. M., Lewis, R. J., and Stülke, J. (2011) RNase Y in *Bacillus subtilis*: A natively disordered protein that is the functional equivalent of RNase E from *Escherichia coli*. *J. Bacteriol. 193*, 5431–5441.

(168) Lee, J., Jang, Y. S., Papoutsakis, E. T., and Lee, S. Y. (2016) Stable and enhanced gene expression in *Clostridium acetobutylicum* using synthetic untranslated regions with a stem-loop. *J. Biotechnol. 230*, 40–43.

(169) Desai, R. P., and Papoutsakis, E. T. (1999) Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. *Appl. Environ. Microbiol. 65*, 936–945.

(170) Song, J.-W., Woo, J.-M., Jung, G. Y., Bornscheuer, U. T., and Park, J.-B. (2016) 3′-UTR engineering to improve soluble expression and fine-tuning of activity of cascade enzymes in *Escherichia coli*. *Sci. Rep. 6*, 29406.

(171) Tummala, S. B., Junne, S. G., and Papoutsakis, E. T. (2003) 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.

(172) Pyne, M., Moo-Young, M., Chung, D., and Chou, C. (2015) Antisense-RNA-Mediated Gene Downregulation in *Clostridium pasteurianum*. *Fermentation 1*, 113–126.

(173) Storz, G., Altuvia, S., and Wassarman, K. M. (2005) AN ABUNDANCE OF RNA REGULATORS. *Annu. Rev. Biochem. 74*, 199–217.

(174) Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol. 8*, 447–454.

(175) Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol. 22*, 841–847.

(176) Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: De-novo-designed regulators of gene expression. *Cell 159*, 925–939.

(177) Garst, A. D., Edwards, A. L., and Batey, R. T. (2011) Riboswitches: Structures and mechanisms. *Cold Spring Harb. Perspect. Biol. 3*, 1–13.

(178) Mellin, J. R., and Cossart, P. (2015) Unexpected versatility in bacterial riboswitches. *Trends Genet. 31*, 150–156.

(179) Wittmann, A., and Suess, B. (2012) Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett. 586*, 2076–2083.

(180) Saad, N. Y., Schiel, B., Brayé, M., Heap, J. T., Minton, N. P., Dürre, P., and Becker, H. D. (2012) Riboswitch (T-box)-mediated control of tRNA-dependent amidation in *Clostridium acetobutylicum* rationalizes gene and pathway redundancy for asparagine and asparaginyl-tRNAAsn synthesis. *J. Biol. Chem. 287*, 20382–20394.

(181) André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., Martin-Verstraete, I., and Soutourina, O. (2008) S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res. 36*, 5955–5969.

(182) Kim, P. B., Nelson, J. W., and Breaker, R. R. (2015) An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell 57*, 317–328.

(183) Projan, S. J., Monod, M., Narayanan, C. S., and Dubnau, D. (1987) 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*, 5131–5139.

(184) Garnier, T., and Cole, S. T. (1988) Identification and Molecular Genetic Analysis of Replication Functions of the Bacteriocinogenic Plasmid plP404 from *Clostridium perfringens*. *Plasmid 19*, 151–160.

(185) Purdy, D., O'Keeffe, T. A. T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N. P. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol. 46*, 439–452.

(186) Lee, S. H., Kwon, M. A., Choi, S., Kim, S., Kim, J., Shin, Y. A., and Kim, K. H. (2015) A new shuttle plasmid that stably replicates in *Clostridium acetobutylicum*. *J. Microbiol. Biotechnol. 25*, 1702–1708.

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

(188) Nakayama, S., Irie, R., Kosaka, T., Matsuura, K., Yoshino, S., and Furukawa, K. (2007) New host-vector system in solvent-producing *Clostridium saccharoperbutylacetonicum* strain N1-4. *J. Gen. Appl. Microbiol. 53*, 53–6.

(189) Maguin, E., Duwat, P., Hege, T., Ehrlich, D., and Gruss, A. (1992) New thermosensitive plasmid for gram-positive bacteria. *J. Bacteriol. 174*, 5633–5638.

(190) Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K., and Minton, N. P. (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res. 40*, e59.

(191) Gu, P., Yang, F., Su, T., Wang, Q., Liang, Q., and Qi, Q. (2015) A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies. *Sci. Rep. 5*, 9684.

(192) Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res. 38*, e92.

(193) Bryant, J. A., Sellars, L. E., Busby, S. J. W., and Lee, D. J. (2014) Chromosome

position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res. 42*, 11383–11392.

(194) Berger, M., Gerganova, V., Berger, P., Rapiteanu, R., and Lisicovas, V. (2016) Genes on a Wire : The Nucleoid- Associated Protein HU Insulates Transcription Units in *Escherichia coli*. *Nat. Publ. Gr. 6*, 1–12.

(195) Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., Van Rij, T., and Hamoen, L. W. (2016) Effect of Genome Position on Heterologous Gene Expression in *Bacillus subtilis*: An Unbiased Analysis. *ACS Synth. Biol. 5*, 942–947.

(196) Anuchin, A. M., Goncharenko, A. V, Demidenok, O. I., and Kaprelyants, A. S. (2011) Histone-like proteins of bacteria (review). *Appl. Biochem. Microbiol. 47*, 580–585.

(197) Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc. 4*, 206–223.

(198) Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. jie, Chen, T., and Zhao, X. (2015) Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng. 31*, 13–21.

(199) Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol. 81*, 2506–2514.

(200) Croux, C., Nguyen, N.-P.-T., Lee, J., Raynaud, C., Saint-Prix, F., Gonzalez-Pajuelo, M., Meynial-Salles, I., and Soucaille, P. (2016) 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.

(201) Yoo, M., Croux, C., Meynial-Salles, I., and Soucaille, P. (2016) 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.

(202) Zhang, N., Shao, L., Jiang, Y., Gu, Y., Li, Q., Liu, J., Jiang, W., and Yang, S. (2015) I-SceI-mediated scarless gene modification via allelic exchange in *Clostridium*. *J. Microbiol. Methods 108*, 49–60.

(203) Dong, H., Tao, W., Gong, F., Li, Y., and Zhang, Y. (2014) A functional *recT* gene for recombineering of *Clostridium*. *J. Biotechnol. 173*, 65–67.

(204) Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., and Blaschek, H. P. (2015)

Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J. Biotechnol. 200*, 1–5.

(205) Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016) Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep. 6*, 25666.

(206) Wang, S., Dong, S., Wang, P., Tao, Y., and Wang, Y. (2017) Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Appl. Environ. Microbiol. 83*.

(207) Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D., and Church, G. M. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. 110*, 14024–14029.

(208) Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol. 30*, 1137–1142.

(209) Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res. 39*, 1131–1141.

(210) Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A., and Arkin, A. P. (2012) RNA processing enables predictable programming of gene expression. *Nat. Biotechnol. 30*, 1002–1006.

(211) Nielsen, A. A. K., Segall-Shapiro, T. H., and Voigt, C. A. (2013) Advances in genetic circuit design: Novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol. 17*, 878–892.

(212) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354–360.

(213) Bardhan, S. K., Gupta, S., Gorman, M. E., and Haider, M. A. (2015) Biorenewable chemicals: Feedstocks, technologies and the conflict with food production. *Renew. Sustain. Energy Rev. 51*, 506–520.

(214) Papoutsakis, E. T. (2008) Engineering solventogenic clostridia. *Curr. Opin. Biotechnol. 19*, 420–9.

(215) Lorenz, R., Bernhart, S. S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C.,

Stadler, P. F. P., and Hofacker, I. I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol. 6*, 26.

(216) Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The Vienna RNA websuite. *Nucleic Acids Res. 36*, W70-4.

(217) Darty, K., Denise, A., and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics 25*, 1974–1975.

(218) Wilson, K. S., and Hippel, P. H. Von. (1995) Transcription termination at intrinsic terminators: The role of the RNA hairpin (*Escherichia coli*/RNA polymerase/rho-independent termination). *Biochemistry 92*, 8793–8797.