Beyond restrictions
Towards biotechnological exploitation of thermophilic clostridia

Joyshree Ganguly
Propositions

1. Establishing a genetic system for non-model organisms is strenuous, but essential for full exploitation of their biotechnological potential.
   (this thesis)

2. CRISPRi is a quick method to assess the impact of metabolic genes and particularly suited for non-model organisms.
   (this thesis)

3. Cognitive science is the important link between robots and humans and will lead to mutual benefits.


5. Gender quota for executive boards is a flawed approach for empowering women in leadership roles.

6. Sustainable developments can only become affordable by simultaneous attainment of long-term environmental and social benefits with short-term economic wins.

Propositions belonging to the thesis, entitled

Beyond restrictions - Towards biotechnological exploitation of thermophilic clostridia

Joyshree Ganguly
Wageningen, 20th May 2020
Beyond restrictions

Towards biotechnological exploitation of thermophilic clostridia

Joyshree Ganguly
Beyond restrictions

Towards biotechnological exploitation of thermophilic clostridia

Joyshree Ganguly
Thesis committee

Promotor
Prof. Dr Richard van Kranenburg
Special Professor Bacterial Cell Factories
Wageningen University and Research
Head of Regulatory and Biotechnology Innovation
Corbion, Gorinchem

Other members
Prof. Dr E.J. Smid, Wageningen University and Research
Prof. Dr O.P. Kuipers, University of Groningen
Dr M. Wells-Bennik, Nizo, Ede
Dr P. Lamers, Corbion, Gorinchem

This research was conducted under the auspices of the Graduate School VLAG
(Advanced studies in Food Technology, Agrotechnology, Nutrition and Health Sciences)
Beyond restrictions
Towards biotechnological exploitation of thermophilic clostridia

Joyshree Ganguly

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 20th May 2020
at 4 p.m. in the Aula.
Joyshree Ganguly
Beyond restrictions - Towards biotechnological exploitation of thermophilic clostridia
178 pages.

PhD thesis, Wageningen University and Research, Wageningen, NL (2020)
With references, with summary in English

DOI: https://doi.org/10.18174/517117
Ma and Bapi, for you.
## Contents

**Chapter 1**  General introduction and thesis outline  

**Chapter 2**  Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: Advances and prospects  

**Chapter 3**  CRISPR interference (CRISPRi) as transcriptional repression tool for  

*Hungateiclostridium thermocellum* DSM1313  

**Chapter 4**  Breaking the restriction barriers and applying a gene silencing tool in  

*Pseudoclostridium thermosuccinogenes*  

**Chapter 5**  Heterologous expression of the *Pseudoclostridium thermosuccinogenes*  

xylose gene cluster in *Hungateiclostridium thermocellum*  

**Chapter 6**  Characterization of sporulation dynamics of *Pseudoclostridium thermosuccinogenes* using flow cytometry  

**Chapter 7**  Thesis summary and general discussion  

---  

Bibliography  

Acknowledgments  

Overview of completed training activities  

List of publications  

About the author
Chapter 1

General introduction and thesis outline
Chapter 1

GENERAL INTRODUCTION

Microbial biotechnology with relevance to thermophiles

Petrochemical refinery processes using fossil-based resources have been used to make a vast amount of chemicals and materials. Due to our increasing concerns on limited fossil resources and environmental impacts, there has been much interest in replacement of fossil-based technologies to bio-based production methods such as those using microbes and fermentation (Soetaert and Vandamme, 2006, Rittmann, 2008, Zhang, 2017). The sustainable production of desired products from renewable resources using microbes is termed as microbial biotechnology, that can aid the transition towards a bio-based economy (Donohue and Cogdell, 2006, Fairley, 2011).

Some aspects involved in microbial biotechnology are strain development, fermentation, product recovery and purification processes (Figure 1.1). Strain development includes host selection and metabolic pathway modulation, including molecular genetics, protein and genetic engineering purposes (Demain, 2000). For microbial conversion, the choice of the production organism is a crucial factor and thermophiles can be potential candidates. Thermostable microorganisms and enzymes have been topics of interest from the early 1960's by the pioneering work of Brock et.al. (Brock and Freeze, 1969). Microorganisms are classified into three groups based on their optimal growth temperatures: psychrophiles (< 20 °C), mesophiles (20-45 °C) and thermophiles (> 45 °C) (Brock, 2001, Turner, et al., 2007). Few years later, the thermophiles were further divided into hyperthermophiles with temperatures above 80 °C (Turner, et al., 2007). The therophilic bacteria can be explored and exploited as platform organisms for potential industrial applications. For instance, thermophiles such as Thermoanaerobacter species, Hungateiclostridium thermocellum and Caldicellulosiruptor bescii are relevant organisms for bioethanol production (Chung, et al., 2015, Jiang, et al., 2017). Parageobacillus thermoglucosidasius has been modified to produce isobutanol and ethanol (Lin, et al., 2014, Olson, et al., 2015). They offer several advantages over mesophilic organisms, such as faster reaction rates, increased substrate and product solubility as well as decreased downstream-processing and cooling costs (Turner, et al., 2007, Jiang, et al., 2017). In addition, simultaneous saccharification and fermentation (SSF) can be performed efficiently (Bhalla, et al., 2013). SSF is an effective process for bioconversion of lignocellulosic biomass where enzymatic hydrolysis and sugars fermentation are combined. The ability of therophilic organisms to use lignocellulosic biomass is critical to make the production process more...
sustainable (Kambam and Henson, 2010, Naik, et al., 2010). Specifically, anaerobic thermophiles such as clostridia are preferred over aerobes since oxygen limits productivities due to the low oxygen transfer in large-scale reactors. Moreover, aeration systems are an additional cost in industrial fermentations (Weusthuis, et al., 2011). Hence, thermophilic anaerobes are suited as potential platform organisms for bio-based industrial applications.

**Figure 1.1.** Schematic overview on microbial production of bio-based products from renewable resources in a biorefinery. The key aspects of microbial biotechnology are selection of suitable host strains, reconstruction of metabolic pathways and fermentation. Suitable host strains include model organisms, natural overproducers and diverse organisms with attractive properties (thermophilic bacteria). Reconstruction of metabolic pathways was attained by genome and protein engineering strategies. Fermentation together with strain development provides valuable feedback. These factors along with downstream processes aid the translation from benchtop research to applications. GRAS: generally recognized as safe. Adapted from Choi et al. (2019).

**Clostridia as potential candidate for industrial exploitation**

Clostridia are classified as Gram-positive, anaerobic and endospore-forming bacteria with low guanine and cytosine content (26-32%), that belongs to the phylum *Firmicutes* (Durre, 2014). The impressive heterogeneity within this class allows clostridia to thrive in widely diverse environments. Among the plethora of *Clostridium* species, few are associated with human and animal diseases (Joseph, et al., 2018). The most significant human-pathogenic species are
**General introduction**

*Clostridium tetani, Clostridium botulinum, Clostridium difficile* and *Clostridium perfringens* responsible for tetanus, botulism, diarrhoea and gas gangrene, respectively (Popoff and Bouvet, 2013, Joseph, et al., 2018). In addition, several non-pathogenic species with medical importance have gained research interest. Specifically, *Clostridium novyi*, whose spores are being used as probiotic and in antitumor therapy (Theys and Lambin, 2015). Other clostridial species are involved in industrial applications due to their diversity of substrate utilization and unique metabolic capabilities. In this regard, *Clostridium acetobutylicum* and *Clostridium beijerinckii* are well-known for solvent production, such as acetone and butanol (Durre, 2014). *Clostridium pasterianum* and *Clostridium propionicum* have been exploited to produce n-butanol and amino acids, respectively (Cardon and Barker, 1946, Leaver, et al., 1955, Schwarz, et al., 2017). The gas-fermenting acetogen *Clostridium autoethanogenum* has been used for ethanol production (Daniell, et al., 2012). Furthermore, the thermophilic clostridia *H. thermocellum* and *Pseudoclostridium thermoanaerobium* can produce ethanol and succinate as major products, correspondingly (McBee, 1954, Sridhar, et al., 2000, Koendjbiharie, et al., 2018). To sum up, the vast metabolic diversity within this class makes it quite interesting for biotechnological exploitation.

*H. thermocellum* was first isolated in 1926 (Viljoen, et al., 1926) and was not fully characterized until 30 years later. It can grow on cellulose and cellobiose (but not on hemicellulose), at temperatures ranging from 50-68 °C, with an optimal 55 °C and the pH ranges between 6.7 to 7.0 (McBee, 1954, Akinosho, et al., 2014). It has the potential to be used as a whole-cell biocatalyst and can break down lignocellulose to convert into lactic acid, acetic acid, formic acid, succinic acid, ethanol, hydrogen and carbon dioxide (Akinosho, et al., 2014). Lignocellulosic feedstocks are the most abundant renewable carbon sources in nature (Kumar and Sharma, 2017). They are comprised of cellulose (38-50 %), hemicellulose (17-32 %) and lignin (15-30 %) (Vismara, et al., 2009). Currently, the conversion of lignocellulose into fermentable sugars comprises several steps. First, different pre-treatment methods are required for degradation of hemi-cellulose and lignin layers, rendering cellulose available for further digestion. Subsequently, enzymatic hydrolysis is necessary. The released polysaccharide chains are hydrolysed into monomeric sugars, which are then transported to specific organisms and converted to ethanol or other useful products (Jónsson and Martín, 2016). The major bottleneck in the conversion of lignocellulose to ethanol is the expensive enzymatic hydrolysis process. To circumvent this hurdle, consolidated bioprocessing (CBP) is an attractive alternative as the enzyme production, hydrolysis and fermentation can be combined into a single step. The
presence of a cellulosome, a hydrolytic multi-enzyme system for the breakdown of cellulose, is the most distinctive feature that makes *H. thermocellum* a relevant CBP candidate (Akinosho, et al., 2014). This approach improves processing efficiencies, as no exogenous hydrolytic enzymes need to be added, and the sugar product inhibition of cellulases is no longer an issue due to the direct consumption by the organism (Lynd, et al., 2005, Xu, et al., 2009, Olson, et al., 2012). Besides, the fermentation efficiency is enhanced which results in a more cost-effective production process for bio-fuels and bio-chemicals (Olson, et al., 2010, Olson, et al., 2012).

Next to *H. thermocellum*, *P. thermosuccinogenes* is another biotechnologically-relevant thermophile, having the ability to ferment inulin to succinate (Sridhar, et al., 2000, Koendjbiharie, et al., 2018). Inulin is present in roots and tubers of plants like Jerusalem artichoke, chicory and dahlia (Sridhar, et al., 2000). The enzyme inulinase is responsible for the degradation of inulin into glucose and fructose. The inulin or glucose fermentation generates succinate and acetate as major products, whereas formate, lactate and ethanol as minor products, respectively (Sridhar, et al., 2000, Koendjbiharie, et al., 2018). *P. thermosuccinogenes* DSM5809 strain can be genetically engineered to enhance succinate production. However, the current genetic toolbox for both the thermophiles is limited. It has been shown previously that transformation efficiencies of several bacteria are adversely affected due to the presence of restriction-modification (RM) systems (Monk and Foster, 2012, Suzuki and Yoshida, 2012, Costa, et al., 2017, Riley, et al., 2019), correlated with improper DNA methylation as described in the next section. Thus, all the active RM systems must be circumvented to overcome foreign DNA cleavage and develop genetic tools to manipulate these organisms.

**Restriction-modification systems**

RM systems act as defence system in bacteria and archaea by recognising exogenous DNA. Around 95% of genome-sequenced bacteria encode RM systems (Roberts, et al., 2010). They comprise two enzymes, a DNA methyltransferase and a restriction endonuclease, exceptional type IV RM system without any methylase (Table 1.1). The methyltransferase methylate specific nucleotides in a DNA sequence protecting self-DNA within the host’s genome from foreign DNA. The restriction endonuclease recognizes the same sequence as the methyltransferase and cleaves the foreign DNA that lacks proper methylation. Four different types of RM systems have been identified such as type I, II, III and IV. These types differ in enzyme composition, mode of action and cofactor requirements as described in Table 1.1. Type
I is a multi-subunit complex composed of two restriction (R) subunits, two modification (M) subunits and one specificity (S) subunit (R2M2S). The restriction activity is triggered when the protein complex composed by all subunits (R2M2S) recognizes non-palindromic, non-contiguous, unmethylated DNA sequences. It also requires ATP, S-adenosyl-L-methionine (SAM) and Mg2+ as cofactors. In contrast, the methylation is accomplished by the protein complex, composed of only the M and S subunits (M2S), using SAM as cofactor and methyl donor. Methylation by type I RM systems generates target sequences with N6-methyl adenine nucleotides (6mA). Type II RM systems consist of two enzymes; the restriction endonuclease and the methyltransferase. The restriction endonuclease needs Mg2+ as a cofactor to be able to cleave palindromic, contiguous and non-methylated, exogenous DNA sequences. Although, the methyltransferase uses SAM to transfer the methyl group to the C5 carbon or N4 amino group of cytosine (5mC and 4mC, respectively), or N6 amino group of adenines (6mA). Type III systems comprise two R subunits and two M subunits. Two non-palindromic and unmethylated sites inversely oriented are recognized by the restriction endonuclease machinery (R2M2), which required ATP, SAM and Mg2+. Type IV lacks the methyltransferase enzyme. The restriction endonuclease recognizes non-specific and variable target sites, which are methylated, hydroxy methylated or glucosyl-hydroxy methylated with a different pattern compared with the host-specific methylation pattern (López-Larrea, 2011, Suzuki, 2012, Vasu and Nagaraja, 2013). RM systems avoid not only phage invasion or gene transfer among bacteria, but also genetic manipulation as they digest foreign DNA that is not correctly methylated. Consequently, they have greatly hampered the study of the clostridial physiology as well as its potential exploitation for industrial purposes (Dong, et al., 2010).
Table 1.1. General characteristics of the different types of RM systems

<table>
<thead>
<tr>
<th>RM</th>
<th>Recognition Sequences</th>
<th>Restriction</th>
<th>Cleavage site</th>
<th>Methylation</th>
<th>Nucleobase</th>
<th>Cofactor</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Non-palindromic</td>
<td>R₂M₂S</td>
<td>Variable</td>
<td>M₂S</td>
<td>⁶m⁶A</td>
<td>ATP</td>
</tr>
<tr>
<td></td>
<td>Non-contiguous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAM</td>
</tr>
<tr>
<td></td>
<td>Unmethylated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mg²⁺</td>
</tr>
<tr>
<td>II</td>
<td>Palindromic</td>
<td>RE</td>
<td>Fixed</td>
<td>MT</td>
<td>⁶m⁶A</td>
<td>Mg²⁺</td>
</tr>
<tr>
<td></td>
<td>Contiguous</td>
<td></td>
<td></td>
<td></td>
<td>⁵m⁶C</td>
<td>SAM</td>
</tr>
<tr>
<td></td>
<td>Unmethylated</td>
<td></td>
<td></td>
<td></td>
<td>⁴m⁶C</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Non-palindromic</td>
<td>R₂M₂S</td>
<td>Variable</td>
<td>M₂</td>
<td>⁶m⁶A</td>
<td>ATP</td>
</tr>
<tr>
<td></td>
<td>Contiguous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAM</td>
</tr>
<tr>
<td></td>
<td>Unmethylated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mg²⁺</td>
</tr>
<tr>
<td>IV</td>
<td>Methylated</td>
<td>RE</td>
<td>Variable</td>
<td>-</td>
<td>-</td>
<td>Mg²⁺</td>
</tr>
<tr>
<td></td>
<td>Hydroxy methylated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucosyl-hydroxy methylated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|Genome engineering tools for thermophilic clostridia |

The development of genetic tools for engineering purposes is important for creation of efficient microbial cell factories. Unfortunately, genetic methodologies within clostridia are often species-specific and even strain-dependent, so genetic tool development is difficult and requires extensive experimentation. The current genetic toolbox for thermophilic clostridia, *H. thermocellum*, is very limited. The genetic tools developed previously for *H. thermocellum* include plasmid-based homologous recombination strategy based on different counter-selection systems such as 5-phosphate decarboxylase (*pyrF*), hypoxanthine phosphoribosyl transferase (*hpt*) and thymidine kinase (*tdk*), or the Targetron technology which have been adapted to disrupt different genes (Tripathi, et al., 2010, Argyros, et al., 2011, Mohr, et al., 2013). The plasmid-based homologous recombination strategy comprehends deletion of a screenable gene and later complementing the mutant host with the same gene as a counter selective marker. To illustrate, the deactivation of *pyrF* gene creates uracil auxotrophic mutants which require uracil supplementation for growth and also resistance to the antimetabolites 5-fluoroorotic acid (5-FOA) (Tripathi, et al., 2010). (Tripathi, et al., 2010). This technique has been used to knockout the phosphotransacetylase (*pta*) gene in thermophilic bacterium *H. thermocellum*. Growth defects with *pyrF* deleted strains does not make this a favourable method. Another counter
selection marker, the hpt gene is involved in re-assimilation of purines during DNA and RNA synthesis, but the presence of purine antimetabolites such as 8-azahypoxanthine (8-AZH) can lead to cellular toxicity. Both pyrF and hpt markers used as counter selection systems, pose a disadvantage with the creation of a strain with deleted chromosomal copy of the respective genes compared to the tdk marker which can be used in a wild type strain. The tdk gene has a role in DNA synthesis and confers sensitivity to 5-fluoro-2-deoxyuridine (FUDR). In the presence of FUDR the tdk marker is often used to select against the replicating plasmids used for genetic manipulation (Shao, et al., 2016). This method has been applied for counter-selection in both eukaryotes and prokaryotes (Gardiner and Howlett, 2004, Shao, et al., 2016). However, counter selection approaches for gene disruption is a multi-step, time-consuming process with low editing efficiencies (Tripathi, et al., 2010).

Other tools include a Targetron system derived from the thermophilic cyanobacterium Thermosynechococcus elongatus, that has been successfully developed as a genome engineering tool for gene disruption in thermophiles, specifically H. thermocellum (Mohr, et al., 2013). Mobile group II introns are retroelements with a ribozyme-based DNA integration machinery, called retro-homing. They are formed during RNA splicing and consist of an intron-encoded protein (IEP) and the excised intron lariat RNA (LtrA). The IEP is a multifunctional protein that combines reverse transcriptase, RNA splicing and DNA endonuclease activities. After translation, the IEP binds to the unspliced LtrA and forms the ribonucleoprotein (RNP) complex. The RNP complex recognizes the target site and unwinds the DNA helix. Afterwards, the intron RNA reverse splices into the DNA strand to which it is base paired, and the opposite DNA strands cleaved by IEP to synthesize the intron cDNA, which is integrated into the genome of the host by DNA repair mechanisms or recombination as depicted in Figure 7.2a, chapter 7 (Song, et al., 2015, Bruder, et al., 2016). Albeit this method denotes a significant development for gene knockouts in Clostridium, it has several limitations. The major drawbacks are the unfeasibility to disrupt small genes without IEP recognition sites, creating clean knockouts and the random insertion events (Song, et al., 2015, Bruder, et al., 2016, Joseph, et al., 2018).

**CRISPR-Cas-mediated genome editing**

The CRISPR-Cas9 technology has provided a precise and high-throughput technique for genome editing. The Clustered Regularly Interspaced Palindromic Repeats (CRISPR) and the CRISPR-associated (Cas) genes encode an adaptive immune system that provides rapid and
robust adaption against invasive genetic elements. It mainly confers resistance to prokaryotic, primary double-stranded (ds) DNA viruses and plasmids (Marraffini, 2015, Mohanraju, et al., 2016); although in certain cases, also exogenous RNA can be hampered (Mohanraju, et al., 2016). They have been classified into two classes based on comparative and phylogenetic genomics, and structural analysis of the Cas genes. Class I systems, including types I, III and IV, are present in bacteria and archaea, whereas class II systems comprise types II, V and VI and are almost restricted to bacteria. Type-II CRISPR-Cas9 and type V CRISPR-Cas12a (or CRISPR-Cpf1) have been widely implemented as efficient genetic tools for precise genome engineering of both eukaryotes and prokaryotes (Mohanraju, et al., 2016).

Type-II CRISPR-Cas9 is a RNA-guided nuclease which catalyzes strand-specific cleavage. Its structure is characterized by the two conserved HNH and RuvC nickase domains conferring the endonuclease activity. HNH domain base pairs with the RNA, cleaves the DNA strand whereas the RuvC domain cleaves the single-stranded (ss-) DNA. Also, Cas9 requires the crRNA and tracrRNA that have been fused together into a chimeric, single-guide RNA (sgRNA) to develop its potential for genome editing purposes (Selle and Barrangou, 2015). The sgRNA chimera has three different domains: (i) spacer - 20 nucleotide region, complementary to a specific target sequence in the genome (protospacer) (ii) 42 nucleotide hairpin Cas9 binding and (iii) 40 nucleotide transcription terminator (Qi, et al., 2013). The design of the spacer allows directing the Cas9 protein to the desired protospacer. The only prerequisite of the target region is the presence of a close protospacer adjacent motif (PAM) sequence (Selle and Barrangou, 2015). The PAM is a 3-8 nucleotides sequence, commonly found at the 3’ end of the protospacer, and it is highly variable among Cas proteins. Cas9 from Streptococcus pyogenes (spCas9) needs the PAM sequence 5’-NGG-3’. SpCas9 recognizes the target DNA via Watson-Crick base pairing by the complementary sgRNA and introduces double-stranded DNA breaks (DSBs), which are lethal for most bacteria since the non-homologous end joining (NHEJ) repair machinery is not functional. Subsequently, the presence of homologous recombination template can introduce the desired modifications in the genome, thus aiding in survival of the bacteria. SpCas9 is inactive at temperatures higher than 42 ºC and can thus be used as genome editing tool in mesophiles and facultative thermophiles (Mougiakos, et al., 2017). In obligate thermophiles, such as H. thermocellum and P. thermosuccinogenes, spCas9 cannot be used for any purpose. Recently, two thermophilic Cas9 variants, namely Geocas9 and ThermoCas9, have been characterized for genome engineering purposes (Harrington, et al., 2017, Mougiakos, et al., 2017). The GeoCas9 and ThermoCas9 have been isolated from Geobacillus stearothermophilus.
and *Geobacillus thermodenitrificans* T12, respectively (Harrington, et al., 2017, Daas, et al., 2018). Both the Cas9 proteins catalyze the RNA-guided DNA cleavage at elevated temperatures (70°C). The vital requirement for successful cleavage of the protoscaler is the presence of the PAM sequence. The most preferred PAM sequence turned out to be 5’-NNNNCRAA-3’ for GeoCas9 and 5’-NNNNCNAA-3’ for ThermoCas9. Gene deletions using the GeoCas9 together with thermophilic recombinases (expressing Exo/Beta homologs from *Acidithiobacillus caldus*) lead to efficient CRISPR/Cas genome editing in thermophilic clostridia *H. thermocellum* (Walker, et al., 2020). Likewise, the ThermoCas9 in previous studies was developed for gene editing purposes in *Bacillus smithii* ET138 and *Pseudomonas putida* KT2440 (Mougiakos, et al., 2017). In addition, transcriptional silencing has been successful in *B. smithii* at 55 °C. The catalytically dead variant of ThermoCas9 (ThermodCas9) has been developed by mutating the RuvC and HNH nuclease domains (D8A and H582A). The two silencing mutations knockout the endonuclease activity of the Cas protein. Therefore, the ThermodCas9 can be used for RNA-guided transcription regulation, named as CRISPR interference (CRISPRi, Figure 7.2b, chapter 7). By co-expressing the sgRNA chimera with the ThermodCas9, transcription initiation or elongation of the target gene is aborted which is explained in detail in chapter 2. In brief, CRISPRi tool enable tuning of gene expression particularly when essential genes or pathways, which require a basal expression, are targeted. It also serves as a method for fast assessment of the possible impact of the genetic modifications in the microbial metabolism.

**Sporulation**

Sporulation is a multifaceted cell differentiation process that occurs in Gram-positive bacteria such as bacilli and clostridia, which allows them to adapt to harsh environments by producing resistant spores. These bacteria have two stages in their life cycle; the metabolically active vegetative form and the dormant spores. To date, the most studied organism on sporulation is *Bacillus subtilis*. In contrast, knowledge on clostridial sporulation, related to its mode of action and regulation at molecular and cellular level is still limited. The early stage of clostridial sporulation differs from *Bacillus* with the absence of the phosphorelay system, considering the evolutionary relationship between them, the latter adapted to an aerobic lifestyle with improved sensing and regulating abilities, which probably resulted in the phosphorelay system (Paredes, et al., 2005, Stephenson and Lewis, 2005, Durre, 2016). The regulation of sporulation includes the activation or repression of specific genes in the sporulation cascade. In *Bacillus*, Spo0A is triggered by phosphorylation, mediated by orphan kinases that activate Spo0F and Spo0B.
proteins, and the other genes involved in the sporulation cascade (Piggot and Hilbert, 2004). In clostridia, the Spo0A phosphorylation machinery is in dispute (Al-Hinai, et al., 2015), but likely some orphan histidine kinases are involved in activation of the Spo0A (Steiner, et al., 2012, Mearls and Lynd, 2014). For example, thermophilic bacterium H. thermocellum has three histidine kinases, namely clo1313_0286, clo1313_2735 and clo1313_1942, which positively control the sporulation process and deletion of these kinases generated strains with an inability to form heat-resistant spores (Mearls and Lynd, 2014). The phosphorylated Spo0A together with σH activates the SpoIIA operon, including the sigma factor σF, which is blocked by anti-sigma factor SpoIIAB. For σF activation, membrane-bound SpoIIE factor is essential. The activated σF triggers and express SpoIIR in the fore-spore, that activates expression of the σE operon (King, et al., 1999, Higgins and Dworkin, 2012, Al-Hinai, et al., 2015). Active σE in the mother cell, controls expression of other genes during sporulation stage II. Consequently, sigma factor σG that is activated regulates stage III sporulation genes and in a way, stimulate the last sigma factor σK in the mother cell for maturation and release of an endospore as described in Figure 1.2 (Al-Hinai, et al., 2015). The Spo0A regulator of sporulation cascade influence other cellular functions, including solventogenesis in some clostridia. Solventogenic strains like C. beijerinckii, C. pasteurianum and C. acetobutylicum, are suitable for butanol, acetone, ethanol or hydrogen production (Durre, 2016). From the solvent production point of view in an industry, an asporogenous strain would be beneficial, as sporulation is an energy demanding process and part of the carbon source is invested for this process instead of synthesis of desired products. Also, mature spores are metabolically inactive. These features will impact productivity negatively in a sporulating culture. Previous studies exhibited non-sporulating cultures influence solvent production during fermentation (Ravagnani, et al., 2000). Despite sporulation playing an important role in solvent production, the interconnection between both remains elusive. Hence, the underlying molecular mechanisms and regulation of the spore formation process and its influence on metabolism are not that well understood.
Chapter 1

**Figure 1.2.** Sporulation signaling cascade with Spo0A the master regulator and the key sporulation sigma factors in both a) *Bacillus* and b) *Clostridium* models. The *Bacillus* model, shows Spo0A is phosphorylated (Spo0A-P) by a phosphorelay system initiated by orphan histidine kinases (HKs). Spo0A-P initiates the sporulation sigma factor cascade involving σF, σE, σG, and σK. In the *Clostridium* model, phosphorelay system was absent and orphan HKs phosphorylate Spo0A directly. Green genes designate confirmed functional roles in *B. subtilis*. Red genes signify functional roles verified by gene inactivation in *Clostridium*. Blue genes indicate hypothetical functionality in *Clostridium* based on homology to *B. subtilis*, genetic orchestration and phenotype. Gray ovals represent assumed protein interactions. A single red question mark means suspected transcriptional activity. The different stages of sporulation are depicted on the right of the *Clostridium* model. Adapted from (Al-Hinai, et al., 2015).
THESIS OUTLINE

Chapter 1 introduces the concept of microbial biotechnology with a focus on thermophiles for sustainable production of bio-based products. This thesis describes the development of thermophilic clostridia as a model on genetic accessibility with restriction modification systems and genetic tools like CRISPRi to modulate metabolic genes to impact fermentation profile. In addition, physiological features such as sporulation is being studied.

Chapter 2 provides a review of the metabolic engineering strategies using CRISPR-Cas-mediated technologies for the development of improved bacterial cell factories. In addition, we discuss about screening diverse environments in search of new CRISPR-Cas variants to expand its applications.

In chapter 3, a novel CRISPR-Cas variant, the ThermodCas9 was applied in *H. thermocellum* to enable gene repression. The CRISPRi system was implemented to knockdown metabolic genes such as lactate dehydrogenase (*ldh*) and phosphotransacetylase (*pta*). In this chapter the silencing effect on *ldh* and *pta* genes will be discussed showing a decrease in lactate and acetate production, respectively.

The establishment of a genetic system for non-model thermophile *P. thermosuccinogenes* is described in chapter 4. In this study, we developed a transformation protocol for *P. thermosuccinogenes* DSM5809. For improving transformation efficiency, we are mimicking DNA methylation patterns of *P. thermosuccinogenes* in an engineered *E. coli* HST04 strain. Plasmids were prepared from this host increased transformation efficiency by two orders of magnitude. Moreover, we developed CRISPRi silencing tool for downregulating the malic enzyme in *P. thermosuccinogenes*, that had an impact on the product formation.

Chapter 5 describes the utility of *P. thermosuccinogenes* xylose gene cluster when introduced into *H. thermocellum* towards consolidated bioprocessing for production biofuels and biochemicals. The enzymes involved in xylose metabolism are xylose isomerase and GTP-dependent xylulokinase, present in a gene cluster together with a transcriptional regulator. Preliminary qRT-PCR data shows upregulation of the xylose metabolizing genes in *H. thermocellum*. In addition, product profiles of the modified strain were characterized based on cellobiose and xylose consumption in anaerobic bottles and fermenters and the differences are discussed in this chapter.
In **chapter 6** we report the development of flow cytometry method for investigation of sporulation populations exemplified by *P. thermosuccinogenes*. Flow cytometer parameters such as side scatter and carboxy fluorescein diacetate dye distinguished sporulation morphologies and viability. Microscopy and heat inactivation confirmed the identities of mature spores, dark and bright phase endospores, forespores and vegetative cells populations. This described approach can be applied to monitor sporulation dynamics during fermentations.

**Chapter 7** provides the main conclusions of this thesis and places the results in a broader perspective. It elaborates on the vital factors and strategies to be considered for genetic accessibility in relation to genetic tool development. In addition, the genetic manipulation confers the influence on metabolism with metabolic or physiological aspects in thermophilic non-model organisms.
Chapter 2

Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: Advances and prospects

Ioannis Mougiakos1*, Elleke F. Bosma2*, Joyshree Ganguly3*, John van der Oost1, Richard van Kranenburg1, 3

1Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands
2The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet B220, 2800 Kgs. Lyngby, Denmark
3Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
*contributed equally

Chapter has been published as:

ABSTRACT

High engineering efficiencies are required for industrial strain development. Due to its user-friendliness and its stringency, CRISPR-Cas-based technologies have strongly increased genome engineering efficiencies in bacteria. This has enabled more rapid metabolic engineering of both the model host *Escherichia coli* and non-model organisms like clostridia, bacilli, streptomycetes and cyanobacteria, opening new possibilities to use these organisms as improved cell factories. The discovery of novel Cas9-like systems from diverse microbial environments will extend the repertoire of applications and broaden the range of organisms in which it can be used to create novel production hosts. This review analyses the current status of prokaryotic metabolic engineering towards the production of biotechnologically relevant products, based on the exploitation of different CRISPR-related DNA/RNA endonuclease variants.
INTRODUCTION

The transition towards a bio-based economy demands the development of fermentation-based processes economically competitive with the currently employed unsustainable production processes (Hillmyer, 2017). Unfortunately, only very few natural organisms are suitable for their direct application in an industrial process. Therefore, efficient metabolic engineering via targeted genome engineering is required and the development and use of simple and high-throughput genome engineering tools generally applicable to many model and non-model organisms is of great importance (Donohoue, et al., 2018). Most prokaryotes possess homology directed repair (HDR) systems, which have since long been exploited in a great variety of microorganisms for targeted chromosomal integrations of desired modifications (Cromie, et al., 2001, Court, et al., 2002, Bowater and Doherty, 2006). In the HDR-based systems, plasmid-borne homologous recombination templates, which often harbour selection markers for screening purposes, are introduced into the genome through double or sequential single crossover events. The HDR-based approach is usually combined with additional systems, such as the site-specific resolvase based Cre-lox or FLP-FRT systems, for excision of the markers from the genomes for recycling purposes. However, these systems leave genomic scars that could be the cause of unwanted chromosomal rearrangements (Olorunniyi, et al., 2016, Ronda, et al., 2016). Alternatively, for a small number of bacteria, marker-less genomic modifications are possible via recombineering systems (Pines, et al., 2015). These systems are based on bacteriophage recombinases and ssDNA, dsDNA or plasmid-borne DNA fragments with sequence homology to the genomic target. However, due to the absence of marker-based selection, these systems mostly result in low mutation efficiencies (Pines, et al., 2015). The construction and screening of mutants in all these HDR-based approaches is time consuming, rendering these tools suboptimal for extensive metabolic engineering, particularly in non-model organisms with low transformability and recombination rates.

A breakthrough moment in the molecular microbiology field was the discovery of bacterial adaptive immune systems that are based on genomic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)—the memory of the systems—and CRISPR-associated (Cas) proteins (Barrangou, et al., 2007, Brouns, et al., 2008, Garneau, et al., 2010). The repurposing of the RNA-guided DNA endonuclease from the type IIa CRISPR-Cas system of Streptococcus pyogenes (spCas9) and of other Cas9 orthologues as genome editing tools brought an unprecedented revolution to the life sciences field (Barrangou, 2014, Mougiakos, et al., 2016,
The basis of the Cas9 engineering tools is the simple way in which Cas9 nucleases can be guided to the desired DNA target, denoted as protospacer, by a CRISPR–RNA: trans-activating CRISPR RNA (crRNA: tracrRNA) hybrid complex. For this purpose, the 5′-end of the crRNA module, denoted as spacer, has to be complementary to the selected protospacer (Deltcheva, et al., 2011) and a specific short DNA motif, denoted as protospacer adjacent motif (PAM), has to be present at the 3′-end of the selected protospacer (Deveau, et al., 2008, Mojica, et al., 2009). For further simplification of the engineering processes, the crRNA: tracrRNA complex can be combined into a chimeric single guide RNA (sgRNA) (Jinek, et al., 2012).

CRISPR-Cas-based genome editing is now broadly used in a variety of organisms, including human cells, zebrafish, plants, yeast and bacteria (Barrangou and Horvath, 2017, Komor, et al., 2017). The successful application of Cas9-based genome editing in eukaryotic cells is based on the error-prone correction of Cas9-induced double-strand DNA breaks (DSBs) by the efficient eukaryotic Non-Homologous End Joining (NHEJ) mechanism. Contrary to eukaryotes, most prokaryotes do not have an active NHEJ system (Bowater and Doherty, 2006) and Cas9-induced DSBs cannot be repaired, resulting in cell death. Recent studies have shown that engineering efficiencies in prokaryotes were strongly increased (often up to 100%) upon combining the existing homologous recombination and recombineering systems with Cas9-targeting; the Cas9-induced DNA breaks served, simultaneously, as recombination inducers and counter-selection tools (Jiang, et al., 2013, Oh and van Pijkeren, 2014, Mougiakos, et al., 2016). Moreover, catalytically inactive variants of Cas9 orthologues and variants fused with transcriptional activating factors have been developed and used for transcriptional regulation. Altogether, CRISPR-Cas9 orthologues and catalytically inactive mutants have accelerated the construction and screening of in silico designed strains, facilitating metabolic engineering of a wide range of bacterial species for industrial cell factory development (Figure 2.1).

In this review, we summarize recent bacterial metabolic engineering studies that focused on the construction and improvement of microbial cell factories, making use of CRISPR-Cas-based technologies. Additionally, we explore newly developed CRISPR-Cas tools and argue on how their application could improve the currently available technologies. Finally, we discuss how screening diverse environments can lead to discovery of new Cas-related variants to extend the repertoire of applications and create novel production hosts.
Chapter 2  Donohoue, et al., 2018) . The basis of the Cas9 engineering tools is the simple way in which repertoire of applications and create novel production hosts. screening diverse environments can lead to discovery o their application could improve the currently available technologies. Finally, we discuss how In this review, we summarize recent bacterial metabolic engineering studies that focused on the wide range of bacterial species for industrial cell factory development (Figure 2.1).

In silico designed strains, facilitating metabolic engineering of a block chemicals. Most of these compounds result from multi-step metabolic pathways and are often tightly regulated in their natural genomic context (Mewalal, et al., 2017, Delmulle, et al., 2018). Model organisms, like E. coli, have well-studied metabolisms, extensive and high-throughput molecular toolboxes and detailed in silico metabolic models. Hence, the use of such organisms minimizes the number of required engineering steps and maximizes engineering efficiencies. Nevertheless, the demands for engineering work remain high before efficient production strains are constructed due to the complexity of the metabolic pathways for many commercially interesting products (Donohoue, et al., 2018). Moreover, model organisms are often suboptimal as production hosts and the use of alternative organisms could benefit the production of many valuable chemicals. Cas-based genome engineering and silencing tools

![Figure 2.1. CRISPR-Cas9-based metabolic engineering of bacterial cell factories. Using Cas9 as counter-selection tool for traditional genome editing processes such as recombineering or plasmid-based homologous recombination has led to efficient metabolic engineering towards a wide range of products. Additionally, dCas9 can be used for CRISPRi to rapidly screen in silico predicted phenotypes prior to Cas9-based engineering.](image-url)
have enabled and accelerated complex metabolic engineering and systems-level understanding of metabolic pathways in a wide range of organisms (Barrangou, 2014, Mougiakos, et al., 2016, Komor, et al., 2017, Donohoue, et al., 2018) (Figure 2.1).

Chapter 2

CRISPR-Cas editing

Metabolic engineering strategies include plasmid-based expression or, preferably, chromosomal integration of heterologous metabolic pathways, and/or targeted genome editing and adaptive evolution for flux redistribution through native metabolic pathways (Table 2.1). In their pioneering work, Li et al. (Li, et al., 2015) combined Cas9-induced targeted DSBs with ss or ds λ-RED recombineering for the introduction of a heterologous β-carotene biosynthetic pathway into the E. coli genome. They further substituted the promoters and ribosome binding sites (RBSs) from the native MEP pathway genes to achieve different levels of overexpression of the corresponding enzymes. Further engineering steps, including numerous deletions and promoter/RBS substitutions of central carbon metabolism genes, improved pyruvate and glyceraldehyde-3P supply and lead to the construction of a highly improved β-carotene producing strain (Li, et al., 2015). This extensive study was possible only due to the development of the Cas9-based tools, revealing their great potential for efficient and diverse manipulation of genomic DNA.

The Cas9-recombineering method was further exploited with the development of the CRISPR-enabled trackable genome engineering (CREATE) tool (Garst, et al., 2017). Application of this tool in E. coli cells allowed their simultaneous transformation with multiple libraries of plasmid-borne recombination templates, each designed to introduce easily trackable mutations at different genomic loci (Garst, et al., 2017). The CREATE tool was employed to introduce multiple RBS variations for each of the genes in a genomically integrated isopropanol production pathway in E. coli, leading to the time-efficient construction and testing of ∼1000 strains. The isopropanol titer of the best strain was 1.5-fold higher compared to the initial integration strain, but still lower compared to the plasmid-based overexpression approach (Liang, et al., 2017). Cas9-based downregulation or deletion of competing pathways in strains already overexpressing heterologous pathways towards the desired product could further improve titers. This has been proven successful for many compounds, including n-butanol in E. coli (Heo, et al., 2017), and Clostridium saccharoperbutylicum (Wang, et al., 2017), isopropanol-butanol-ethanol in Clostridium acetobutylicum (Wasels, et al., 2017), succinic acid in Synechococcus elongatus (Li, et al., 2016), γ-amino-butyric acid (GABA) in
Corynebacterium glutamicum (Cho, et al., 2017), and 5-aminolevulinic acid in E. coli (Ding, et al., 2017) (Table 2.1). However, the use of natural producers can substantially reduce the complexity of engineering steps towards production and tolerance build-up. The tolerance of Streptomyces species to antibiotics has been exploited for the production of antibiotic and antitumor compounds simply by Cas9-facilitated genomic integration of multiple biosynthetic gene cluster (BGC) copies (Li, et al., 2017, Meng, et al., 2017). In C. glutamicum, the natural proline production was enhanced 6.5-fold through a codon saturation mutagenesis approach to relieve product inhibition (Jiang, et al., 2017). Noteworthy, this work was performed using the Cas12a (formerly Cpf1) RNA-guided endonuclease, making it the first application of a non-Cas9-based CRISPR-Cas/recombineering genome editing tool in bacteria for metabolic engineering purposes (Table 2.1).

Finally, Cas9-based editing tools were successfully employed for membrane engineering purposes in E. coli. The β-carotene storage capacity of E. coli cell membranes was increased by chromosomal integration of heterologous membrane-bending protein genes using plasmid-borne homologous recombination and Cas9-targeting (Wu, et al., 2017). Furthermore, the Cas9-recombineering tool proved efficient for the enhancement of the E. coli lipid content by simultaneous chromosomal integration of a heterologous fatty acid regulatory transcription factor gene together with a delta9 desaturase and an acetyl-CoA carboxylase gene (Xia, et al., 2016) (Table 2.1).
<table>
<thead>
<tr>
<th>Product</th>
<th>Species</th>
<th>Strategy</th>
<th>Chromosomal modifications made using CRISPR-Cas-editing</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terpenoids</td>
<td></td>
<td></td>
<td>Knock-in <em>almgs</em> under control of various regulatory parts (in a β-carotene production strain)</td>
<td>(Wu, et al., 2017)</td>
</tr>
<tr>
<td>Isopropanol</td>
<td><em>Escherichia coli</em></td>
<td>Heterologous, chromosomal</td>
<td>Knock-in and RBS replacement of <em>thl, atoDA, adc, adh</em></td>
<td>(Liang, et al., 2017)</td>
</tr>
<tr>
<td>n-Butanol</td>
<td><em>Escherichia coli</em></td>
<td>Heterologous, plasmid</td>
<td>Modification of <em>gltA</em> 5'-UTR for expression reduction</td>
<td>(Heo, et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>*Clostridium saccharoper-</td>
<td>Competing pathways</td>
<td>Knock-out <em>pta, buk</em></td>
<td>(Wang, et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>butylacetonicum*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isopropanol</td>
<td><em>Clostridium acetobutylicum</em></td>
<td>Heterologous and native,</td>
<td>Knock-in <em>ctfAB, adc, adh</em></td>
<td>(Wasels, et al., 2017)</td>
</tr>
<tr>
<td>butanol-</td>
<td></td>
<td>chromosomal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethanol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acids</td>
<td>Org.-acids</td>
<td>Antibiotics/anti-tumor</td>
<td>Lipid content</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>L-proline</td>
<td>Succinic acid</td>
<td>Pristinamycin I (PI)</td>
<td>Fatty acids</td>
<td></td>
</tr>
<tr>
<td>γ-amino-</td>
<td>Synechococcus elongatus</td>
<td>Pristinamycin II (PI)</td>
<td>Escherichia coli</td>
<td></td>
</tr>
<tr>
<td>butyric acid (GABA)</td>
<td></td>
<td>Chloramphenicol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-Amino-</td>
<td>Streptomyces pristinaespiralis</td>
<td>Anti-tumor compound YM-216391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>levulinic acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escherichia coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product strategies</th>
<th>Codons (to relieve product inhibition)</th>
<th>Codon saturation mutagenesis γ-glutamyl kinase by CRISPR-Cpf1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomal</td>
<td>Knock-out Ncgl1221, gabT, gabP and various combinations thereof</td>
<td>(Jiang, et al., 2017)</td>
</tr>
<tr>
<td>Knock-in cctfAB, adc, adh</td>
<td>coaA point mutation (R106A), serA promoter replacement and C-terminal residues deletion, knockout sucCD, hemB translational downregulation by start codon substitution</td>
<td>(Ding, et al., 2017)</td>
</tr>
<tr>
<td>Knock-out glgc and knock-in glutA-ppc (under nitrogen starvation conditions)</td>
<td>Knock-in of artificial attachment/integration (attB) sites in which the biosynthetic pathway is subsequently inserted</td>
<td>(Meng, et al., 2017)</td>
</tr>
<tr>
<td>Knock-in fudR, delta9 and acc (deletions made previously)</td>
<td></td>
<td>(Li, et al., 2017)</td>
</tr>
</tbody>
</table>
CRISPR interference

Next to the integration or deletion of genes and pathways, an important metabolic engineering strategy is the fine-tuning of gene expression. Whereas in eukaryotic systems siRNA-techniques have since long enabled transcriptional control, for prokaryotes such silencing tools have only recently become available with the CRISPR interference (CRISPRi) tool, which is based on dCas9: the catalytically inactive variant of the Cas9 endonuclease (Qi, et al., 2013). This tool can be used for complete or partial repression; repression strength can be tuned by altering the position of the selected protospacer within the targeted gene (Figure 2.2), or by employing an inducible promoter for the expression of the dCas9 or the sgRNA module and altering the amount of the corresponding inducer. This is crucial when targeting essential genes, competing pathways (which can also be biomass-producing) or regulators for which a basal expression level is required (Figure 2.2a–d). It can also be used as a quick alternative to the often laborious RBS/promoter-engineering to tune production pathway activity to either modulate amounts and properties of the target product (Figure 2.2e) (Li, et al., 2017, Tao, et al., 2017) or prevent accumulation of toxic intermediates (Kim, et al., 2016) (Figure 2.2f). Although CRISPRi does not lead to the construction of stably genetically modified strains, it is a powerful method for quick evaluation of the possible effects of genetic modifications to the metabolism of a microorganism, allowing to design genome editing approaches and to gain insights into microbial metabolism (Figure 2.1, Table 2.2).
Next to the integration or deletion of genes and pathways, an important metabolic engineering strategy is the fine-tuning of gene expression. Whereas in eukaryotic systems siRNA techniques have since long enabled transcriptional control, for prokaryotes such silencing tools have only recently become available with the CRISPR interference (CRISPRi) tool, which is based on dCas9: the catalytically inactive variant of the Cas9 endonuclease (Qi, et al., 2013).

This tool can be used for complete or partial repression; repression strength can be tuned by altering the position of the selected protospacer within the targeted gene (Figure 2.2), or by employing an inducible promoter for the expression of the dCas9 or the sgRNA module and altering the amount of the corresponding inducer. This is crucial when targeting essential genes, competing pathways (which can also be biomass-producing) or regulators for which a basal expression level is required (Figure 2.2a–d). It can also be used as a quick alternative to the often laborious RBS/promoter-engineering to tune production pathway activity to either modulate amounts and properties of the target product (Figure 2.2e) (Li, et al., 2017, Tao, et al., 2017) or prevent accumulation of toxic intermediates (Kim, et al., 2016) (Figure 2.2f).

Although CRISPRi does not lead to the construction of stably genetically modified strains, it is a powerful method for quick evaluation of the possible effects of genetic modifications to the metabolism of a microorganism, allowing to design genome editing approaches and to gain insights into microbial metabolism (Figure 2.1, Table 2.2).
Figure 2.2. Overview of CRISPRi-based metabolic engineering strategies to increase production of the target product (P). Abbreviations: C: carbon source; I: intermediate metabolite; B: by-product or biomass; P: target product; TI: toxic intermediate; R: repressor; M: morphology. See legend of Figure 2.1 for graphic legend. Arrows represent intracellular carbon flows and their thickness corresponds to the flow rate. Arrows with dashed outlines represent merged pathways. **a)** Repression of competing pathway that leads to by-product or biomass formation with (i) indicating dCas9 targeting the promoter region, resulting in stronger repression than in (ii), where the coding region is targeted. **b)** Repression of competing pathway that leads to product consumption with (i) indicating dCas9 targeting the promoter region, resulting in stronger repression than in (ii), where the coding region is targeted. **c)** Repression of repressor of the target product pathway. **d)** Repression of cell shape/morphology genes to increase cell size and storage capacity for the target product. **e)** Repression of the product pathway to change product composition or properties. **f)** Repression of the product pathway to prevent accumulation of toxic intermediates.

Similar to CRISPR-Cas-based editing, model organisms such as *E. coli* and *C. glutamicum* were the first organisms for which CRISPRi-based metabolic engineering was applied (Table 2.2). A heterologous polyhydroxyalkanoate (PHA) biosynthesis pathway was introduced into *E. coli* for the production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) [P(3HB-co-4HB)] (Lv, et al., 2015). The 4HB fraction of the polymer was enhanced via CRISPRi-based downregulation of multiple TCA cycle genes aiming to increase the supply of the 4HB-precursor succinate semialdehyde (Lv, et al., 2015) (Figure 2.2e). Subsequent works in *E. coli* and in the natural PHA-producer *Halomonas* focused on silencing cell morphology genes to increase the storage capacity of the cells for PHAs (Elhadi, et al., 2016, Tao, et al., 2017) (Figure 2.2d), as well as biosynthetic pathway genes to control PHA content and chemical properties such as molecular weight and polydispersity (Li, et al., 2017, Tao, et al., 2017) (Figure 2.2d). Cell morphology engineering, through CRISPRi-based repression of *E. coli* cell division and shape genes, combined with expression of a heterologous production pathway, was also used for the production of the phytochemical zeaxanthin (Shen, et al., 2016) (Table 2.2). Several other studies used CRISPRi-based repression of competing pathway or repressor genes for the enrichment of precursor pools (Figure 2.2a,c), aiming at the enhancement of the natural amino acid production by *C. glutamicum* (Cleto, et al., 2016) and the *E. coli*-based heterologous production of various phytochemicals including naringenin (Cress, et al., 2015, Wu, et al., 2015), resveratrol (Wu, et al., 2017), pinosylvin (Wu, et al., 2017), anthocyanin (Cress, et al., 2017), as well as medium chain fatty acids (MCFAs) (Wu, et al., 2017) and n-butanol (Kim, et al., 2017) (Table 2.2). Additionally, CRISPRi-based repression of essential genes was used to minimize carbon loss towards biomass formation by decoupling growth and production (Li, et al., 2016) or flux balancing (Kim, et al., 2016) (Table 2.2). Notably, many of these studies
employed multiplex silencing using an sgRNA approach, whereas Cress et. al. used a dual RNA (crRNA/tracrRNA) approach, developing a rapid CRISPR-array assembly method denoted as CRISPPathBrick (Cress, et al., 2015). This tool could facilitate multiplex CRISPRi-based silencing in non-model organisms with limited genetic toolboxes.

CRISPRi-based repression has already been used for metabolic engineering purposes in non-model organisms such as in Clostridium cellulovorans and Klebsiella pneumoniae for alcohol production (Wang, et al., 2017, Wen, et al., 2017), Synechococcus elongatus for succinate production (Huang, et al., 2016) (Table 2.2), as well as in Clostridium acetobutylicum to relieve carbon catabolite repression for sugar co-utilization (Bruder, et al., 2016). The CRISPRi tool is particularly useful in cyanobacteria, in which genome editing is complicated and time consuming due to slow growth and multiple chromosome copies (Huang, et al., 2016, Yao, et al., 2016). Furthermore, the ability to fine-tune expression levels using CRISPRi was exploited in Synechococcus sp., where repression of nitrogen assimilation gene glnA was shown to increase the pool of α-ketoglutarate (Gordon, et al., 2016). A moderate increase of this metabolite enhanced glycolytic flux and lactate production, whereas a too large increase resulted in a decrease in protein production (Gordon, et al., 2016). CRISPRi-based multiplex gene repression was established in Synechocystis sp. (Yao, et al., 2016) and subsequently used to determine optimal gene repression combinations for fatty alcohol production (Kaczmarzyk, et al., 2018) (Table 2.2). As in all studies using CRISPRi in cyanobacteria, dCas9 and sgRNAs were chromosomally integrated into the genome. This resulted in stable repression strains in the absence of selective pressure for single sgRNAs, but the use of repetitive promoter elements resulted in undesired recombination events when multiplexing was attempted (Kaczmarzyk, et al., 2018), highlighting the potential advantage of using a dual RNA approach. Finally, it was observed that a targeted gene with a very distant transcription start site (TSS) from the start codon could be efficiently repressed by CRISPRi only when employing multiple sgRNAs targeting within the gene or the preceding operon genes (Yao, et al., 2016). Hence, this study revealed that the efficient application of the CRISPRi tool is strongly connected with the precise identification of TSS for the targeted genes.
<table>
<thead>
<tr>
<th>Product pathway overexpression method</th>
<th>Genes targeted by CRISPRi</th>
<th>M</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competing pathways</td>
<td>sad/gadB, sucCD, sdhAB</td>
<td>+</td>
<td>(Lv, et al., 2015)</td>
</tr>
<tr>
<td>Cell morphology</td>
<td>ftsZ, mreB</td>
<td>-</td>
<td>(Elhadi, et al., 2016)</td>
</tr>
<tr>
<td>Cell morphology</td>
<td>ftsZ, prpC, gltA</td>
<td>-</td>
<td>(Tao, et al., 2017)</td>
</tr>
<tr>
<td>Cell morphology</td>
<td>ftsZ, mreB, pbp, rodZ</td>
<td>-</td>
<td>(Shen, et al., 2016)</td>
</tr>
<tr>
<td>Competing pathways</td>
<td>eno, adhE, mdh, fabB, fabF, sucC, fumC</td>
<td>+</td>
<td>(Wu, et al., 2015)</td>
</tr>
<tr>
<td>Competing pathways</td>
<td>fadR, fumC, sucABCD, scpC</td>
<td>+</td>
<td>(Cress, et al., 2015)</td>
</tr>
<tr>
<td>Competing pathways</td>
<td>fabD, fabH, fabB, fabF, fabI</td>
<td>-</td>
<td>(Wu, et al., 2017)</td>
</tr>
<tr>
<td>Repressor of product pathway</td>
<td>metI</td>
<td>-</td>
<td>(Cress, et al., 2017)</td>
</tr>
<tr>
<td>Terpenoids</td>
<td>(-)-α-bisabolol, isoprene, lycopene</td>
<td>Escherichia coli</td>
<td>Heterologous, plasmid</td>
</tr>
<tr>
<td>Terpenoids</td>
<td>Mevalonate</td>
<td>Heterologous, plasmid</td>
<td>Competing pathway: biomass</td>
</tr>
<tr>
<td>Amino acids</td>
<td>L-lysine, L-glutamate</td>
<td>Corynebacterium glutamicum</td>
<td>-</td>
</tr>
<tr>
<td>Alcohols</td>
<td>Acetone-butanol-ethanol</td>
<td>Clostridium cellulovorans and Clostridium beijerinckii</td>
<td>Native, plasmid</td>
</tr>
<tr>
<td>Alcohols</td>
<td>n-Butanol</td>
<td>Klebsiella pneumoniae</td>
<td>Heterologous, plasmid</td>
</tr>
<tr>
<td>Alcohols</td>
<td>Fatty alcohols</td>
<td>Synechocystis sp.</td>
<td>Heterologous, plasmid</td>
</tr>
<tr>
<td>Organic acids</td>
<td>Succinate</td>
<td>Synechococcus elongatus</td>
<td>-</td>
</tr>
<tr>
<td>Organic acids</td>
<td>Lactate</td>
<td>Synechococcus sp.</td>
<td>-</td>
</tr>
</tbody>
</table>

1See Figure 2.2 for a visualization of these strategies. Modifications other than CRISPRi are shown in brackets.
2M: multiplexing.
FUTURE PERSPECTIVES

In prokaryotes, CRISPR-Cas based genome editing has strongly increased engineering efficiencies by adding a powerful counter-selection method to existing engineering systems or by enhancing recombination efficiencies through induction of cellular DNA repair mechanisms (Jiang, et al., 2013, Mougiakos, et al., 2016). In model organism *E. coli* high-throughput tools such as crMAGE (Ronda, et al., 2016) and CREATE (Garst, et al., 2017) enabled multiplex engineering by combining the Cas9 and recombineering tools. It is expected that these high-throughput tools will be further developed into automated pipelines for rapid industrial strain development, but the thorough exploitation of their potential requires the additional development of rapid and easy screening and read-out systems. CRISPR-Cas-based counter-selection tools have increased editing efficiencies in many non-model organisms. However, further improvement of these tools still strongly depends on the development of basics, such as well-characterized inducible promoters (Yan and Fong, 2017). Tight control of the Cas9-expression would allow for efficient integration of an employed homologous recombination template prior to the counter-selection step. The use of intrinsic Cas9-properties, such as temperature-sensitivity, can substitute the requirement for inducible promoters in organisms that can grow under conditions outside the Cas9 activity range (Mougiakos, et al., 2017).

The use of alternative Class 2 CRISPR systems will further extend the Cas-based engineering toolbox. Recently, the Cpf1 (Cas12) RNA-guided DNA endonucleases from the type V CRISPR-Cas systems of *Francisella novicida* (Jiang, et al., 2017) and *Acidaminococcus* sp. (Zhang, et al., 2017) have been repurposed for bacterial genome editing and silencing. Cpf1 does not require a tracrRNA and can process its own precursor crRNA via its intrinsic RNase activity. Hence, the use of Cpf1 for the development of a multiplex engineering tool can prevent the issues encountered when using Cas9 and multiple sgRNAs for multiplex engineering (Zhang, et al., 2017). The use of a Cpf1-recombineering tool in *C. glutamicum*, which tolerates only very low levels of Cas9 expression (Cho, et al., 2017, Jiang, et al., 2017, Peng, et al., 2017), resulted in screenable editing efficiencies (Jiang, et al., 2017), while a DNase-dead Cpf1 (ddCpf1) variant was recently employed for multiplex silencing in *E. coli* (Zhang, et al., 2017). The newly discovered Class 2 Type VI system Cas13 (C2c2) RNA-guided RNA-nuclease can be used for silencing via the degradation of transcripts, or for tracking of transcripts using fluorescent-coupled catalytically inactive variants (Abudayyeh, et al., 2016, Nakade, et al., 2017). The RNA-guided RNA endonuclease from the type VI CRISPR-Cas systems of
Leptotrichia shahii and Leptotrichia wadei have already been successfully repurposed for RNA interference in E. coli (Abudayeh, et al., 2016, Abudayeh, et al., 2017, Nakade, et al., 2017). Furthermore, the repurposing of native CRISPR-Cas systems for genome editing (Li, et al., 2015), has been proved efficient and holds promise for organisms with low transformation efficiencies (Stout, et al., 2017). It is anticipated that the development of easy and rapid characterization techniques (Cong, et al., 2013, Esvelt, et al., 2013, Jenkinson and Krabben, 2015, Mitsunobu, et al., 2017) will accelerate the exploitation of novel CRISPR-Cas systems for the development of prokaryotic engineering tools. These tools will further expand the number of target sites, the range of easy-to-engineer organisms and they will increase the engineering speed by simultaneous usage of different Cas systems for genome editing and plasmid curing (Nakade, et al., 2017, Yan, et al., 2017), similar to the recently developed EXIT-circuit approach that combines Cas9-based editing and I-SceI-based plasmid curing (Tang, et al., 2017).

Screening natural resources for novel CRISPR-Cas systems will further expand the applications and range of organisms and environments in which CRISPR-Cas-based editing can be applied (Karvelis, et al., 2017, Koonin, et al., 2017, Murugan, et al., 2017). A recent example of this approach is the identification of a thermostable Cas9-orthologue in the genome of a Geobacillus thermodenitrificans strain isolated from a compost sample, which was further characterized and employed to establish the first Cas9-based engineering tool for thermophilic bacteria (Mougiakos, et al., 2017). The robustness of thermostable Cas9-based tools can be further exploited for applications in extreme environments, as was recently shown for another thermostable Cas9-orthologue with prolonged life time in blood plasma (Harrington, et al., 2017). The discovery of novel Cas nucleases with different properties, such as tolerance to alkaline or acidic pH and high saline concentrations, would be possible by screening selected environmental samples and metagenomic libraries. The characterization of these nucleases could lead to the development of engineering tools with wide applicability to biotechnologically relevant but currently unexploited extremophilic organisms.

CONCLUSIONS

Altogether, the developments in CRISPR-Cas-based bacterial genome engineering increase insight into metabolism on a systems level and enable more rapid strain engineering, which is crucial for the development of a bio-based economy using microbial cell factories. Rapid current developments and future applications, which will further expand the range of organisms
and applications of CRISPR-Cas-based editing for metabolic engineering, consist of fine-tuning of the tools, their adaptation to different hosts, their extension into combinations with other active components such as proteases, markers or activators, as well as the discovery and development of novel Cas-like systems.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGMENTS**

This work was supported by Dutch Technology Foundation STW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs (Grant Number 13938); Marine Biotechnology ERA-NET ‘ThermoFactories’ (Grant Number 5178-00003B) and the Novo Nordisk Foundation, Denmark; and Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068].
Chapter 3

CRISPR interference (CRISPRi) as transcriptional repression tool for *Hungateiclostridium thermocellum* DSM1313

Joyshree Ganguly\(^1\), Maria Martin-Pascual\(^2\), Richard van Kranenburg\(^{1,2}\)

\(^1\)Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
\(^2\)Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

Chapter has been published as:

ABSTRACT

*Hungateiclostridium thermocellum* DSM1313 has biotechnological potential as a whole-cell biocatalyst for ethanol production using lignocellulosic renewable sources. The full exploitation of *H. thermocellum* has been hampered due to the lack of simple and high-throughput genome engineering tools. Recently in our research group, a thermophilic bacterial CRISPR-Cas9-based system has been developed as a transcriptional suppression tool for regulation of gene expression. We applied ThermoCas9-based CRISPR interference (CRISPRi) to repress the *H. thermocellum* central metabolic lactate dehydrogenase (*ldh*) and phosphotransacetylase (*pta*) genes. The effects of repression on target genes were studied based on transcriptional expression and product formation. Single-guide RNA (sgRNA) under the control of native intergenic 16S/23S rRNA promoter from *H. thermocellum* directing the ThermodCas9 to the promoter region of both *pta* and *ldh* silencing transformants reduced expression up to 67% and 62%, respectively. This resulted in 24% and 17% decrease in lactate and acetate production, correspondingly. Hence, the CRISPRi approach for *H. thermocellum* to downregulate metabolic genes can be used for remodelling of metabolic pathways without the requisite for genome engineering. These data established for the first time the feasibility of employing CRISPRi-mediated gene repression of metabolic genes in *H. thermocellum* DSM1313.
INTRODUCTION

Microbial conversion of lignocellulosic renewable feedstock to fuels, chemicals and other bio-based products significantly reduces the dependence on non-sustainable fossil energy-based processes. This contributes to the transition towards a bio-based economy. Non-model production microbes with native ability of solubilizing lignocellulose have been improved for robust industrial production processes (Akinosho, et al., 2014, Lynd, 2016, Kumar and Sharma, 2017). The anaerobic thermophilic bacterium Hungateiclostridium thermocellum is proficient at solubilizing cellulose as it possesses a hydrolytic multienzyme complex called the cellulosome. This enzyme complex solubilizes cellulose to celldextrins that are further metabolized to produce ethanol, acetate, L-lactate, formate, hydrogen gas and carbon dioxide (McBee, 1954). Cellulose solubilization has the potential to improve product formation but is not cost-effective for the conversion of lignocellulose to ethanol. Consolidated bioprocessing (CBP) combines enzyme production, cellulose hydrolysis and fermentation into a single step, offering an economical advantage to the current multistep process (Jiang, et al., 2017). In CBP no exogenous hydrolytic enzymes are added, thereby reducing the costs. Moreover, pretreatment, hydrolysis and fermentation steps are combined in one vessel, reducing the number of unit operations (Bhalla, et al., 2013). Besides, fermentation efficiency is improved as it also lowers the cost of the production of biofuels and biochemicals. To this end, the efficient genetic engineering of H. thermocellum would render it a highly valuable ‘microbial cell factory’ for biotechnological exploitation.

Metabolic engineering of H. thermocellum is necessary to improve the yield of ethanol or alternative preferred products from mixed-acid fermentation. However, the current genetic toolbox for H. thermocellum is still very limited. A plasmid-based homologous recombination gene disruption method, including hpt and pyrF-based counter selection, has been developed (Tripathi, et al., 2010). Using hpt and pyrF as a counter-selection marker few genes such as pta, ldh, adhE, cipA, cel48S, spo0A and recA were deleted to interrogate the genetic base for specific phenotypes and to create modified strains for different functions (Olson, et al., 2010, Argyros, et al., 2011, Brown, et al., 2011, Olson, et al., 2013, van der Veen, et al., 2013, Groom, et al., 2018, Lo, et al., 2019). This tool has not been used extensively, as it is a laborious and time-consuming process. Subsequently, a Targetron system derived from the thermophilic cyanobacterium Synechococcus elongatus has been developed into a genome engineering tool for gene disruption in thermophiles, specifically H. thermocellum (Mohr, et al., 2013).
Targetron is a genome editing technology derived from mobile group II introns, which efficiently creates site-specific integrations for gene knockout and knock-in clostridia. The main limitations for this system include the inability to generate clean knockouts, reduced target specificity due to lack of intron-encoded protein (IEP) recognition sites for small genes, the variation in the intron-integration efficiency and the presence of ectopic intron insertion events (Song, et al., 2015, Bruder, et al., 2016, Joseph, et al., 2018).

The development of simple, and preferably high-throughput, genome engineering tools is crucial for efficient metabolic engineering, as well as for a full exploitation of this thermophile. The CRISPR-Cas9 technology has provided a precise technique for genome editing, paving the way for a new era in molecular biology. In addition, CRISPRi has commenced as an effective technique for gene downregulation and can be used to modulate gene expression. This approach uses a nuclease-deficient Cas9 (dCas9) in conjunction with spacer region of a single-guide RNA to repress specific genes targeting at the promoter and start of the gene. CRISPRi systems have been established in mesophilic clostridial species, including Clostridium acetobutylicum and Clostridium beijerinckii (Li, et al., 2016, Wang, et al., 2016, Wen, et al., 2017). Recently, a thermophilic Cas9, ThermoCas9, has been isolated from the thermophilic bacterium Geobacillus thermodenitrificans T12 (Daas, et al., 2018). It is a type-IIIC Cas9 that is active over a broad temperature range. A dCas9 variant has been generated and used for CRISPRi silencing of the lactate dehydrogenase activity in the thermophilic Bacillus smithii (Mougiakos, et al., 2017).

High repression of gene expression has been reported when the dCas9-sgRNA ribonucleoprotein complex targets the non-coding DNA strand (5’–3’) of a gene, specifically close to the ATG start codon, whereas the effect is less pronounced if the coding strand (3’–5’) is targeted. Furthermore, effective gene silencing is observed when targeting the promoter region, with the strongest effect at the −35 sequence, being independent of the DNA strand (Qi, et al., 2013). It has been hypothesized that the inhibition of the transcription is due to a physical collision between the RNA polymerase and the dCas9-sgRNA complex. During elongation, the RNA polymerase encounters the dCas9-sgRNA, which is bound to the non-template DNA, pausing the transcription elongation. Alternatively, if dCas9-sgRNA complex is bound to the template strand, slight repressive effect can be observed as the RNA polymerase could still read through the complex. The sgRNA faces the RNA polymerase, which could unzip the complex by helicase activity upon targeting the template strand (Qi, et al., 2013). Moreover, gene
knockdowns are less likely to be lethal and frequently permit analysing surviving mutants where drastic knockout techniques fail (Qi, et al., 2013, Pyne, et al., 2014, Peters, et al., 2016). In this study, ThermodCas9 was adapted for use in *H. thermocellum* as a tool to enable silencing of gene expression. The CRISPRi system was implemented to effectively knockdown metabolic genes such as lactate dehydrogenase (*ldh*) and phosphotransacetylase (*pta*) via appropriate sgRNA design. Selective repression of the *ldh* and *pta* genes decreased lactate and acetate production respectively. Hence, demonstrating the feasibility of employing CRISPRi for the metabolic engineering of *H. thermocellum* and production of bio-derived chemicals.

**RESULTS**

**Selection of promoters**

To establish the ThermodCas9-based CRISPRi tool in *H. thermocellum*, functional promoters are required for efficient expression of the ThermodCas9 and the sgRNA. The pThermoCas9i plasmid (Mougiakos, et al., 2017) encodes the catalytically inactive variant of ThermoCas9 (ThermodCas9) and the non-targeting sgRNA-expressing module under the control of the constitutive *xylL* promoter from *B. smithii* and *pta* promoter from *Bacillus coagulans* respectively. This plasmid was introduced in *H. thermocellum* DSM1313 via electroporation as described by Olson *et. al.* (Olson and Lynd, 2012). To validate expression of both genes, RNA was isolated from *H. thermocellum* transformants containing the non-targeting pThermoCas9i plasmid that were grown in CP medium. RT-PCR on cDNA showed that the ThermodCas9 gene was expressed, but the sgRNA was not expressed (Figure 3.1). From this, we concluded that the *xylL* promoter from *B. smithii* is functional in *H. thermocellum*, but the *pta* promoter from *B. coagulans* is not.

**Promoter screening for the establishment of CRISPRi system in *H. thermocellum* DSM1313**

Based on the previous results, it was necessary to replace the *B. coagulans pta* promoter with an alternative one to efficiently express the sgRNA. To develop an efficient system, we decided to assess the use of native promoter of small ribosomal intergenic 16S/23S rRNA from *H. thermocellum*. Due to the limited number of characterized promoters from this anaerobic thermophile, we selected approximately 250 bp long upstream sequence of the 16S and 23S rRNA genes anticipating the presence of promoters (Supplementary data S1). Ribosomal promoters are preferred for efficient sgRNA expression (Li, et al., 2013, Xu, et al., 2014). We
replaced the *pta* promoter of the initial constructed plasmids with the predicted intergenic 16S/23S rRNA promoters and tested sgRNA expression by RT-PCR (Figure 3.2). The intergenic 16S/23S rRNA promoter drove sgRNA gene expression. Therefore, we chose the *xyll* promoter to drive ThermodCas9 gene expression and the 16S/23S rRNA promoter to drive sgRNA expression.

Figure 3.1. RT-PCR of viable *H. thermocellum* transformants, containing the pThermoCas9i plasmid. a) Schematic illustration of plasmid pThermoCas9i_NT (Mougiakos, et al., 2017). Thermodcas9 gene under control of the *B. smithii* xyll promoter; sgRNA-expressing module under control of the *B. coagulans* pta promoter; and pNW33n backbone. b, c and d) Non-targeting 1 and 2, independent pThermoCas9i_NT (Mougiakos, et al., 2017) transformants of *H. thermocellum* (lanes 2 and 3) showing the RT-PCR products of 124 bp (b) and 282 bp (c) from thermodcas9 cDNA and the absence of the RT-PCR product of 169 bp (d) from the sgRNA; WT, RT-PCR with *H. thermocellum* DSM1313 wild-type cDNA; *B. smithii*, RT-PCR with *B. smithii* harbouring pThermoCas9i_NT (Mougiakos, et al., 2017) cDNA. e) MQ, RT-PCR with Milli-Q water (lane 1); WT, RT-PCR with *H. thermocellum* DSM1313 wild-type RNA (lane 3); Non-targeting 1 and 2, RT-PCR with *H. thermocellum* transformants 1 and 2 RNA, respectively, (lane 4 and 5); and *B. smithii*, RT-PCR with *B. smithii* harbouring pThermoCas9i_NT (Mougiakos, et al., 2017) cDNA (lane 6).

### Transcriptional repression on lactate dehydrogenase (*ldh*) and phosphotransacetylase (*pta*) genes

To investigate the utility of CRISPRi for gene repression in *H. thermocellum*, we aimed to target the non-template strand of both lactate dehydrogenase (*ldh*) and phosphotransacetylase (*pta*) genes. We constructed pThermoCas9i vectors with the *xyll* promoter expressing the ThermodCas9 and the intergenic 16S/23S rRNA promoter controlling the sgRNA together with a non-targeting spacer (pThermoCas9i_NT) or a spacer targeting the non-template DNA strand overlapping either the promoter region or the start of the gene for both *ldh* (*ldh_P; ldh_S*) and *pta* (*pta_P; pta_S*) genes respectively. Targeting and non-targeting plasmids were
electroporated to *H. thermocellum* DSM1313. A drop of two orders of magnitude in transformation efficiency was observed when using the targeting plasmid (10² CFU µg DNA⁻¹) in comparison with the controls (10⁴ CFU µg DNA⁻¹). All the tested colonies for the targeting plasmid showed a positive band for the expression of the ThermodCas9 and sgRNA. These clones were grown in 50 mL CP medium containing 6 µg/mL thiamphenicol (TmR) and cultured to OD₆₀₀ ~ 0.6 and > 1.0. Five millilitres of cells were sampled for RNA extraction and qRT-PCR analysis to determine the downregulation of the silencing transformants. qRT-PCR was performed using *gyrA* and *recA* as house-keeping genes. Eventually, *recA* was used as a reference gene as it appeared more consistent in expression than *gyrA* (data not shown).

The qRT-PCR analysis (Figure 3.3) on RNA isolated from exponentially growing cells harvested at an OD₆₀₀ ~ 0.6 showed 62% and 58% reduction in *ldh* gene expression, using the expression in control transformant cells with the pThermoCas9i_NT (NT control) as the reference, when targeting the promoter (*ldh_P*) and start of the gene region (*ldh_S*) respectively. When harvested at the early stationary phase at an OD₆₀₀ > 1.0, the repression effect was still
Chapter 3

retained 43% for the *ldh*<sub>P</sub> gene, but the *ldh*<sub>S</sub> transformant no longer showed repression and had the same expression level as the NT control. Similarly, at an OD<sub>600</sub> ~ 0.6 for the *pta*-targeting transformants the silencing effect was 67% and 62% when targeting the promoter (*pta*<sub>P</sub>) and start of the gene (*pta*<sub>S</sub>) respectively. When the OD<sub>600</sub> > 1.0, the expression was 45% and 39% reduced for the *pta*<sub>P</sub> and *pta*<sub>S</sub> transformants respectively. These data indicate that CRISPRi suppressed the gene expression, and the silencing efficacy was high in comparison to the control.

![Metabolic pathway representation of H. thermocellum DSM1313 for conversion of cellobiose to ethanol and major organic acids production.](image_url)

**Figure 3.3.** Effective suppression of *ldh* and *pta* genes by CRISPRi. a) Metabolic pathway representation of *H. thermocellum* DSM1313 for conversion of cellobiose to ethanol and major organic acids production. (ACK, acetate kinase; ADH, alcohol dehydrogenase; AldH, aldehyde dehydrogenase; LDH, lactate dehydrogenase; PFL, pyruvate-formate lyase; and PTA, phosphotransacetylase). b) Relative *ldh* and *pta* gene expression was evaluated using qRT-PCR in comparison with control non-targeting (NT) at an OD<sub>600</sub> ~ 0.6 and 1.0 respectively; positive silencing transformants for *ldh* gene referred as *ldh*<sub>P</sub> – targeting the promoter, *ldh*<sub>S</sub> – targeting the start of the gene; positive silencing transformants for *pta* gene referred as *pta*<sub>P</sub> – targeting the promoter, *pta*<sub>S</sub> – targeting the start of the gene. *H. thermocellum* DSM1313 cells were transformed with both non-targeting pThermoCas9i_NT and targeting pThermoCas9i_ldh/pta plasmids. The Tm<sup>R</sup> colonies of both the non-targeting and targeting plasmids were transferred to 50 mL CP (5 g/L yeast extract) medium containing 6 µg/mL Tm and cultured to OD<sub>600</sub> ~ 0.6 and > 1.0, and 5 mL cells were sampled for qRT-PCR analysis with technical triplicates. The expression levels were normalized to those in the NT control. Data represent the mean values of three biological replicates and the standard deviation. The level of significance of the differences when results were compared was estimated by means of analysis of variance (ANOVA), with α = 0.05.

In addition, the impact of the silencing on the other respective genes involved in product formation was evaluated. At OD<sub>600</sub> ~ 0.6, the *ldh* gene for *pta* silencing transformants targeting the promoter (*ldh*<sub>E1</sub>) and start of the gene (*ldh*<sub>E2</sub>) showed enhanced expression respectively compared with the NT control. For the early stationary phase cultures, the *ldh*<sub>E1</sub> expression was same as NT control but increased for the *ldh*<sub>E2</sub>. Contrarily, the *pta* gene had reduced expression compared with the NT control at early exponential phase, for both the *ldh* silencing
transformants (pta_E1, targeting the promoter and pta_E2, targeting the start of the gene). Hence, the non-targeted metabolic pta and ldh genes responded differently to the silencing of ldh and pta transformants respectively (data not shown).

**Impact of gene silencing on product formation using HPLC**

To evaluate the effect of repression of ldh and pta gene on organic acids production, silencing transformants were cultured to an OD$_{600}$ ~ 0.6 and > 1.0, and 5 mL cells were sampled for qRT-PCR analysis with technical triplicates. The same expression level as the NT control. Similarly, at an OD$_{600}$ ~ 0.6 for the ldh gene had reduced for the pta–, but the ldh gene referred as ldh_P and ldh_S transformants respectively. These data indicate that CRISPRi targeting the promoter, ldh_P transformant. Contrarily, both the pta– targeting the start of the gene referred as pta_S transformants respectively (data not shown). The Tm R colonies of both the non-targeting plasmids. The Tm R colonies of both the non-targeting plasmids. The Tm R colonies of both the non-targeting plasmids. The Tm R colonies of both the non-targeting plasmids.

Impact of gene silencing on product formation using HPLC

To evaluate the effect of repression of ldh and pta gene on organic acids production, silencing transformants were cultured to an OD$_{600}$ ~ 0.5-1.5, in anaerobic bottles containing 50 mL CP medium (1 g/L yeast extract) for 2 days. Samples were taken at two-time points for HPLC analysis for the two transformants in comparison with the NT control. Moreover, the fermentation profile was observed to see the effect of knockdown of ldh and pta genes on other organic acids and ethanol production.

It was clearly noted that repression of ldh and pta genes at the promoter region led to a 24% and 17% reduction on lactate and acetate production in comparison with NT cells respectively (Figure 3.4). At time point T1 (OD$_{600}$ ~ 0.5-0.7), the acetate production in the pta silencing transformants (pta_P and pta_S) was suppressed. In contrast, the lactate production for these transformants was increased around 25% to the NT control. For the ldh silencing transformants (ldh_P and ldh_S), lactate was reduced drastically with slender decrease in acetate as well. Other products for both ldh and pta silencing transformants such as formate had slightly better yield in all silencing transformants whereas ethanol did not show any significant differences from the control. When the OD$_{600}$ > 1.0 (T2), the yield (mol of lactate/mol of cellobiose consumed) for ldh_P was 0.36 and ldh_S was 0.55 whereas that of NT was 0.48. Thus, lactate production was still low for the ldh_S transformant. Contrarily, both the pta silencing transformants had similar levels of acetate, formate and ethanol production compared with the NT control (Supplementary data, Figure S1).

In summary, HPLC showed decrease in acetate and lactate production at an exponential phase but in the late stationary phase the fermentation products come to same yields as the NT cells. This shows that despite that ThermodCas9 can silence specific metabolic genes, *H. thermocellum* can escape the silencing upon prolonged incubation.
The first steps towards the adaptation of ThermodCas9-based CRISPRi tool showed the functionality of the CRISPRi tool in thermophilic clostridia, we targeted two ThermodCas9 expressing strains. Positive transformants referred as *ldh*P, *ldh*S, *pta*_P, *pta*_S and control *ldh*_non-targeting (NT) were analyzed in HPLC at OD$_{600}$ = 0.5-0.7 (T1). The silencing transformants for *ldh* gene referred as *ldh*_P – targeting the promoter, *ldh*_S – targeting the start of the gene; positive silencing transformants for *pta* gene referred as *pta*_P – targeting the promoter, *pta*_S – targeting the start of the gene. *H. thermocellum* DSM1313 cells were transformed with both non-targeting pThermoCas9i_NT and targeting pThermoCas9i_ *ldh* and *pta* plasmids. The Tm$^R$ colonies of both the non-targeting and targeting plasmids were transferred to 50 mL CP medium (1 g/L yeast extract) containing 6 µg/mL Tm and cultured to OD$_{600}$ ~ 0.5-0.7 (T1) and > 1.0 (T2), and 1 mL cells were sampled for 2 days for HPLC analysis. Data represent the mean values of three biological replicates and the standard deviation. The level of significance of the differences when results were compared was estimated by means of analysis of variance (ANOVA), with $\alpha = 0.05$.

**DISCUSSION**

CRISPRi offers reduction in gene expression and can be used to interrogate gene function and modulate cellular activities. Recent studies have explored CRISPRi to control the metabolic pathways in *E. coli* and other non-model organisms for improved production of numerous biotechnological products (Mougiakos, et al., 2018). In this study, we applied the CRISPR-ThermoCas9-based transcriptional regulation tool for the first time an effective system in *H. thermocellum* DSM1313. The existing methods for genome editing and transcriptional regulation purposes are laborious, regularly inefficient and technically complex in the anaerobic bacteria, which are known to be challenging for genetic manipulation. The recent discovery of a thermostable Cas9 coupled to its transformation into a powerful genetic tool has opened new possibilities for genome manipulation of the thermophilic bacteria (Mougiakos, et al., 2017). The dCas9-based CRISPRi allows the transcriptional regulation of the gene of interest without...
completely disrupting its function, resulting in less pleiotropic effects than gene knockouts (Dominguez, et al., 2016, Peters, et al., 2016).

To evaluate the functionality of the CRISPRi tool in thermophilic clostridia, we targeted two genes of the central metabolism; the lactate dehydrogenase (ldh) and phosphotransacetylase (pta) in H. thermocellum DSM1313. It has been proven that the deletions of these genes are not detrimental for its survival and allows the increase in ethanol production (Argyros, et al., 2011, Biswas, et al., 2014, Papanek, et al., 2015). We constructed two ThermodCas9 expressing vectors, with sgRNA genes targeting the promoter and start of the gene plus a control plasmid with a non-targeting sgRNA. Based on previous studies, the target sites were in the non-template strand, near the promoter region and start of both ldh and pta genes. According to literature, the highest transcriptional downregulation effect was expected to occur where ThermodCas9 targets the promoter or beginning of the coding sequence, the lowest repression in the transformants where the middle of the gene was targeted (Larson, et al., 2013, Qi, et al., 2013, Yao, et al., 2016).

The first steps towards the adaptation of ThermodCas9-based CRISPRi tool showed the xyll promoter from Bacillus smithii allows the expression of ThermodCas9 protein in H. thermocellum. However, the pta promoter from Bacillus coagulans failed to transcribe the gene encoding the sgRNA. Bacilli and clostridia are closely related bacteria that belong to the phylum Firmicutes (Talukdar, et al., 2015). The close-relatedness among these two classes might explain the functionality of the xyll promoter from Bacillus smithii in H. thermocellum, specifically because both are thermophilic strains. Bacillus coagulans is also a facultative anaerobic thermophilic bacterium, but the reason for the non-functioning of the pta promoter in H. thermocellum remains unclear.

To develop an efficient CRISPRi tool, we successfully managed to get the activity of putative native intergenic 16S/23S rRNA promoter for the expression of the chimeric sgRNA. The promoter region within 16S/23S intergenic spacer sequence has homology to the prokaryotic consensus promoter structure. In E. coli, it has been shown that this promoter has a role in the transcription of ribosomal RNAs (Mankin, et al., 1987, Zacharias and Wagner, 1989). Now, we show that this is also likely to be the case for H. thermocellum.

This study shows the first attempt to exploit the CRISPRi tool in thermophilic clostridia to modulate gene expression using the central metabolism phosphotransacetylase (pta) and lactate dehydrogenase (ldh) genes as examples. The results showed downregulation levels of 67% and 62%, respectively, when the promoter and the start of pta gene were targeted. Consequently,
the organic acids production changed in comparison with NT control: acetate production decreased between 17% and 27%; formate production increased between 5% and 12%; and ethanol had no significant change. As studied before, the \textit{pta} knockout mutant generated using the Targetron system had 13.5% and 81.5% decreased acetate and lactate production, respectively, and 42% improvement in ethanol production (Mohr, et al., 2013). Contrarily, we observed 30% of increased lactate production and this correlated well with the upsurge in \textit{ldh} gene expression for \textit{pta} silencing transformants. Argyros \textit{et. al.} (Argyros, et al., 2011) have shown similar impact of improved lactate titre with \textit{pta} knockout. In addition, no significant increase in ethanol production was noted in comparison with the NT control. The plausible explanation for this could be despite of the silencing, the residual expression still resulted in sufficient enzyme activity for the cell to still produce acetate and ATP instead of ethanol and NAD(P). Besides, further genetic engineering of pyruvate metabolism is required to increase ethanol production as discussed in previous studies (Dash, et al., 2017, Tian, et al., 2017).

The \textit{ldh} silencing transformants \textit{ldh\_P} and \textit{ldh\_S} similarly had 62% and 58% downregulation levels with decrease in lactate production in comparison with NT control. Mohr \textit{et. al.} (Mohr, et al., 2013) have shown with the \textit{ldh} knockout mutant a 4% reduction in lactate production in comparison with wild type. In contrast, we showed 24% decrease in lactate production and 15% increase in formate; however, no increase in acetate or ethanol was detected. This was also noted with minor reduction in \textit{pta} gene expression at exponential phase for both \textit{ldh\_P} and \textit{ldh\_S}. In another study, the \textit{ldh} mutants of \textit{H. thermocellum} were characterized, acetate titre was also not altered. However, no prominent difference was observed in gene expression and acetate production in the stationary phase. Thus, signifying lactate dehydrogenase deletion does not affect the flux through the pyruvate node if ethanol is produced (van der Veen, et al., 2013). Moreover, for ethanol production, there were no significant differences observed when compared to the NT control. This result was comparable to earlier reports of \textit{H. thermocellum} \textit{pta} mutant (Tripathi, et al., 2010). Future work in developing a \textit{H. thermocellum} strain could involve \textit{pta} and \textit{ldh} multiplex silencing or knockout using dead or active thermophilic Cas9, respectively, to study if there is any impact on product formation. Nevertheless, the genetic system developed here demonstrates a step forward towards manipulating \textit{H. thermocellum} strain with slight effects on other genes in the metabolic pathways.

The main challenge for applying thermophilic CRISPR-Cas system seems to be its robustness. In previous studies, this system has been successfully applied in \textit{B. smithii} to reduce the highly expressed \textit{ldhL} gene expression for about 90% resulting in 40% reduction in lactate production.
(Mougiakos, et al., 2017). Similarly, we have repressive effect for both pta and ldh genes with reduction in acetate and lactate, respectively. However, at stationary phase the cells overcome the repressive effect with less impact on product formation. This CRISPRi loss-of-function effect has been shown in several growth studies where essential genes were targeted (Zhao, et al., 2016, Liu, et al., 2017), correlating it with the presence of suppressor mutations in the dead Cas9 coding sequence at stationary phase. Even a single mutation of the first seven nucleotides intensely decreased repression, suggesting the importance of the ‘seed region’ sequence for binding of both the type I and type II CRISPR systems (Jinek, et al., 2012, Qi, et al., 2013). Therefore, it is likely that the importance of these metabolic genes poses a strong selection pressure for the cells that have overcome this burden by suppressor mutations for survival affecting the repression at stationary phase. Further studies are required to demonstrate that this is actually the case.

To date, *H. thermocellum* strains have been genetically engineered, improving the production of ethanol using cellulose as a carbohydrate source. Two main strategies have been followed. On one hand, competing pathways for carbon flux are disrupted (Argyros, et al., 2011, Biswas, et al., 2014, Papanek, et al., 2015, Kannuchamy, et al., 2016). In another approach, competing pathways for electron flux are eliminated (Biswas, et al., 2015, Rydzak, et al., 2015, Lo, et al., 2017). This is the first time a silencing tool has been applied to *H. thermocellum* DSM1313 to transcriptionally regulate the expression of genes involved in central metabolism. In this case, qRT-PCR has shown clear effect on the repression of both the genes in the exponential growth phase which leads to efficient redistribution of the organic acids and ethanol in *H. thermocellum* DSM1313.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains and growth conditions**

All bacterial strains used in this study are listed in Table S1, Supplementary data. High efficiency NEB (New England Biolabs, CFU ~ 10⁹ μg/ pUC19 plasmid DNA) chemically competent *E. coli* DH5α was used for cloning purposes. *E. coli* Acella (Acella™ Chemically Competent cells, Edge Bio) was used for plasmid propagation and isolation when *H. thermocellum* DSM1313 was transformed to ensure the transfer of Dam-methylated plasmids. *E. coli* strains were cultured in Lysogeny-Broth (LB) medium (1% tryptone, 1%
NaCl, 0.5% yeast extract) at 37 °C and 200 rpm. When appropriate the medium was supplemented with 20 μg/mL chloramphenicol. 

*H. thermocellum* DSM1313 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). *H. thermocellum* DSM1313 wild type and transformants were grown anaerobically in CTFUD adapted from Olson and Lynd (Olson and Lynd, 2012) (0.13% ammonium sulphate, 0.15% potassium phosphate monobasic, 0.013% calcium chloride dehydrated, 0.05% L-cysteine hydrochloride, 1.15% MOPS sodium salt, 0.0001% ferrous sulphate heptahydrate, 0.26% magnesium chloride hexahydrate, 0.45% BD BBL™ Yeast Extract, 0.5% Na-resazurin solution 0.1% w/v and 0.3% sodium citrate tribasic dihydrate) broth and agar media at 50-60 °C range (Olson and Lynd, 2012). For plasmid selection after electroporation, 6 µg/mL thiamphenicol was added. For RNA isolation and HPLC analysis, the silencing transformants were grown in CP medium adapted from Plugge (0.408 g/L potassium dihydrogen phosphate, 0.534 g/L sodium phosphate dibasic dihydrate, 0.3 g/L ammonium chloride, 0.3 g/L sodium chloride, 0.1 g/L magnesium chloride hexahydrate, 0.11 g/L calcium chloride dihydrate, 4.0 g/L sodium bicarbonate, 1.0 g/L L-cysteine and 5.0 g/L (RNA isolation) or 1.0 g/L (HPLC) yeast extract (BD Bacto), 0.5 mg resazurin, 1 mL vitamin solution, 1 mL trace elements solution I and 1 mL trace elements solution II. The final volume of 50 mL medium was dispensed in serum bottles under 80:20 N₂/CO₂ atmosphere with ~70 kPa overpressure and then autoclaved. A stock solution comprising sodium bicarbonate and L-cysteine was autoclaved separately and added later. Likewise, vitamin stock solution was added to calcium chloride dihydrate solution after it was autoclaved. Cellobiose as a carbon source was also autoclaved separately and added later to a final concentration of 5.0 g/L (Plugge, 2005). The vitamin solution, 1000× concentrated, contained per litre 20 mg biotin, 20 mg folic acid, 100 mg pyridoxine-HCl, 50 mg thiamine-HCl, 50 mg riboflavin, 50 mg nicotinic acid, 50 mg Ca-d-pantothenate, 1 mg vitamin B12, 50 mg 4-aminobenzoid acid and 50 mg lipoic acid. Trace elements solution I, 1000× concentrated, contained per litre 50 mM HCl, 61.8 mg H₃BO₄, 99.0 mg MnCl₂·4H₂O, 1.49 g FeCl₂·4H₂O, 119 mg CoCl₂·6H₂O, 23.8 mg NiCl₂·6H₂O, 68.2 mg ZnCl₂ and 17.0 mg CuCl₂·2H₂O. Trace elements solution II, which was 1000× concentrated, contained per litre 10 mM NaOH, 17.3 mg Na₂SeO₃, 33.0 mg Na₂WO₄·2H₂O and 24.2 mg Na₂MoO₄·2H₂O.

**CRISPRi plasmids construction**

Both the plasmids and primers used in this study are listed in Table S2, Supplementary data. Plasmids were built using fragments with the appropriate overhangs via NEBuilder® HiFi DNA
assembly cloning kit according to the manufacturer’s protocol. All PCR reactions for amplification of fragments were performed with the NEB Q5® High-Fidelity DNA polymerase (M0491). PCR fragments were subjected to 1% w/v agarose gel electrophoresis and isolated using Zymoclean™ Gel DNA Recovery kit. Plasmid pThermoCas9i (Mougiakos, et al., 2017) was used as a template for the construction of all the plasmids. Plasmids were designed for the knockdown of the lactate dehydrogenase (ldh: Clo1313_1160) and phosphotransacetylase (pta; Clo1313_1185) in the genome of *H. thermocellum* DSM1313. Two different protospacers were selected for each targeted gene and a non-targeting with random sequences was used as a control (Table S3, Supplementary data). They differ on the position within the targeted gene. The target sites are the promoter, start of the gene and the non-targeting spacer with random sequences. To block transcription, the non-template strand of the gene was targeted. Plasmids with alternative promoters to drive the sgRNA expression were based on the previously built ThermodCas9 plasmids by exchanging the *B. coagulans* pta promoter of the sgRNA expression module with 250 bp of the *H. thermocellum* intergenic region of the 16S and 23S rRNA genes (Table S1). All plasmids were introduced by heat shock into chemically competent *E. coli* DH5α cells (Sambrook, et al., 1989). Plasmids were isolated from selected single transformants by using the GeneJET Plasmid Miniprep Kit® (Thermo Scientific). Plasmid sequences were confirmed by standard sequencing from Macrogen (MACROGEN Inc. DNA Sequencing Service; Amsterdam, The Netherlands) using the primers present in Table S4, Supplementary data. After sequence confirmation, plasmids were electroporated into *E. coli* Acella (Sambrook, et al., 1989) before being introduced in *H. thermocellum* by electroporation as described by Olson and Lynd (Olson and Lynd, 2012).

**RNA isolation**

RNA isolation of *H. thermocellum* mutants, harbouring the pThermoCas9i plasmids, was performed using 5 mL of overnight cultures in the mid-log phase (OD$_{600}$ nm reached 0.6). In order to harvest the cells, the culture was brought inside the anaerobic chamber where 5 mL was transferred to a 50 mL Greiner tube. Cells were centrifuged aerobically at 4800 g and 4 °C for 15 minutes. The supernatant was carefully decanted, and the pellet was resuspended in 0.5 mL of ice cold TE buffer (pH 8.0). The pellet was placed on ice during the next steps of the protocol, unless otherwise specified. The cell suspension was divided in two screw-capped tubes containing 0.5 g of zirconium beads, 30 μL of 10% w/v SDS, 30 μL of 3 M sodium acetate (pH 5.2) and 500 μL of Roti® aqua Phenol (Carl Roth GmbH, Karlsruhe, Germany). Cells were disrupted in the Fastprep apparatus (MP Biomedicals, Solon, OH, USA) at speed 6 for 40 s,
centrifuged at 10,000 rpm and 4 °C, for 5 minutes. 300 µL of aqueous phase was transferred to a sterile 1.5 mL Eppendorf tube, and 300 µL of chloroform-isoamyl alcohol (Carl Roth GmbH, Karlsruhe, Germany) was added. The mixture was centrifuged at 14,000 rpm and 4 °C, for 3 minutes. 250 µL of water phase was transferred to a sterile 1.5 mL Eppendorf tube and mixed with 250 µL of the Lysis/Binding buffer of the Maxwell® 16 LEV Total RNA Cells Kit. Consequently, the rest of the steps were followed according to the RNA isolation Kit. The purified RNA was measured in the NanoDrop spectrophotometer to determine the quality and the concentration of the elution and stored at -80 °C.

**First-strand cDNA synthesis and RT-PCR**

The first-strand cDNA synthesis was performed in a 20 µL reaction, containing 2500 ng of RNA, 125 ng of random primers, 10 mM dNTP mix, 5× First-Strand Buffer, 0.1 M DTT, the SuperScript™ III RT and RNase-free water, following the manufacturer’s instruction of the SuperScript™ III Reverse Transcriptase protocol (Invitrogen (Life Technologies Europe BV), The Netherlands). The cDNA as well as RNA was used in RT-PCR to analyse the expression of both, the ThermodCas9 protein and the sgRNA. The RNA was used as a negative control. The primers BG11642 and BG11643 were used to amplify 169 bp of the sgRNA using the NEB Q5® High- Fidelity DNA polymerase. The primers BG11636 and BG11637 were used to amplify 282 bp of the ThermodCas9 using the NEB Q5® High-Fidelity DNA polymerase.

**Quantitative real time PCR**

To assess the functionality of ThermodCas9 based CRISPRi as a silencing tool, a quantitative real-time PCR (qRT-PCR) was performed using the cDNA synthesized from the RNA of *H. thermocellum* DSM1313, expressing the ThermodCas9 and the sgRNA targeting the lactate dehydrogenase (*ldh*) and phosphate acetyltransferase (*pta*) genes. As controls, the cDNA was also synthesized from the RNA of the wild-type *H. thermocellum* DSM1313 strain and the transformant strains with ThermodCas9 and the non-targeting sgRNA. It was used to measure the relative expression level of the *ldh* and *pta* genes of the transformant *H. thermocellum*, comparing it to the transformant *H. thermocellum* with the ThermodCas9 and the non-targeting sgRNA.

The qRT-PCR was performed by using the iQ™ SYBR® Green Supermix from Bio-Rad. The final volume of the reaction was set to 20 µL; thus, all the components were scaled accordingly relevant to the manufacturer’s protocol. The cDNA samples were diluted in sterile Milli-Q water. The amount of cDNA used as a template was relative to 50 ng of RNA. The house-
keeping gene used to measure the relative expression was the recombinase A (recA) of *H. thermocellum* DSM1313. The primers used to amplify the *ldh* and *pta* genes of *H. thermocellum* DSM1313 were BG14575, BG14580 and BG15853, BG15854 respectively (Table S6, Supplementary data). The qRT-PCR was run in a Bio-Rad C1000 Thermal Cycler.

**High-pressure liquid chromatography**

A high-pressure liquid chromatography (HPLC) system ICS-5000 was used for the organic acids and ethanol quantification. The system has Aminex HPX 87H column from Bio-Rad Laboratories and is equipped with a UV1000 detector operating on 210 nm and a RI-150 40 °C refractive index detector. The mobile phase consisted of 0.16 N H$_2$SO$_4$ and the column was operated at 0.8 mL/minute. All samples were diluted 4:1 with 10 mM DMSO in 0.01 N H$_2$SO$_4$. The *H. thermocellum* NT control, *ldh* and *pta* silencing transformant strains were grown in CP medium (1 g/L yeast extract) for 2 days and samples were taken at different time points from OD$_{600}$ ~ 0.5-1.5 to analyse using HPLC. Sugars (cellobiose, glucose, ethanol and glycerol) and organic acids (acetic acid, lactic acid, succinic acid and formic acid) were used as standards with a concentration range between 1.25 and 25 mM.

**CONFLICT OF INTEREST**

None declared.

**ACKNOWLEDGMENTS**

The authors thank Ioannis Mougiakos for providing the ThermodCas9 plasmids and technical assistance with plasmid design. We are grateful to Daniel Olson for his advice on the transformation protocol of *H. thermocellum* DSM1313.

Funding: This work was supported by Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068].
SUPPLEMENTARY DATA

S1 Promoter intergenic 16S/23S rRNA of *H. thermocellum* DSM1313

TGAGG GCCTATGTCGGCTTTGAGAAGAATTGCAAATATAAATTGCAACCTTGAAAATATTATAATGCTGTA AAGCAAGATGAAAGGTAACAGCAGCATGACCAGAAGCTTAAAGAGTGACCAGGATGCAAGT GACTCTTAGGATCTCATTAGATGGTGAAAAGACATCACTTTAAGGAGGACCAAAGGAAAGGACC TGAGG

Figure S1. Effects of CRISPRi-mediated suppression on lactate and acetate production and impact on the fermentation profile. For lactate and acetate production, positive transformants referred as *ldh* _P, ldh_ _S, pta_ _P, pta_ _S and control ldh non-targeting (NT) were analyzed in HPLC at OD<sub>600</sub> > 1.0 (T2). The silencing transformants for *ldh* gene referred as *ldh* _P - targeting the promoter, *ldh* _S - targeting the start of the gene; positive silencing transformants for *pta* gene referred as *pta* _P - targeting the promoter, *pta* _S - targeting the start of the gene. *H. thermocellum* DSM1313 cells were transformed with both non-targeting pThermoCas9i_NT and targeting pThermoCas9i _ldh and _pta plasmids. The Tm<sup>R</sup> colonies of both the non-targeting and targeting plasmids were transferred to 50 mL CP medium (1g/L yeast extract) containing 6 µg/mL Tm<sup>R</sup> and cultured to OD<sub>600</sub> ~ 0.5-0.7 (T1) and > 1.0 (T2), and 1 mL cells were sampled for 2 days for HPLC analysis. Data represent the mean values of three biological replicates and the standard deviation. The level of significance of the differences when results were compared was estimated by means of analysis of variance (ANOVA), with α=0.05.
Table S1 Bacterial strains used in the present study

<table>
<thead>
<tr>
<th>Strains</th>
<th>Organism</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>fluA2 (argF-lacZ) U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 thi-1 hsdR17</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>F’ ompT hsdSB (rB− mB+) gal dcm (DE3) ΔendA ΔrecA</td>
<td>Edge Bio</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>DSM1313</td>
<td>DSMZ</td>
</tr>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-NT (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a non-targeting spacer)</td>
<td>This study</td>
</tr>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-PTA-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-PTA-S (16S/23S native rRNA promoter driving the expression of the sgRNA intergenic promoter a spacer targeting the start region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-LDH-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of ldh gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DH5α</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-LDH-S (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the start region of ldh gene)</td>
<td>This study</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid</td>
<td>This study</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-NT (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a non-targeting spacer)</td>
<td>This study</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-PTA-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>Species</td>
<td>Strain</td>
<td>Description</td>
<td>Transcriptional Repression Tool</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-P TA-S (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the start region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-LDH-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of ldh gene)</td>
<td>This study</td>
</tr>
<tr>
<td>Acella</td>
<td>Escherichia coli</td>
<td>E. coli with pThermoCas9i plasmid-16S/23S rRNA-LDH-S (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the start region of ldh gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>H. thermocellum with pThermoCas9i plasmid-16S/23S rRNA-N NT (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a non-targeting spacer)</td>
<td>This study</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>H. thermocellum with pThermoCas9i plasmid-16S/23S rRNA-PTA-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>H. thermocellum with pThermoCas9i plasmid-16S/23S rRNA-PTA-S (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the start region of pta gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>H. thermocellum with pThermoCas9i plasmid-16S/23S rRNA-LDH-P (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the promoter region of ldh gene)</td>
<td>This study</td>
</tr>
<tr>
<td>DSM1313</td>
<td>Hungateiclostridium thermocellum</td>
<td>H. thermocellum with pThermoCas9i plasmid-16S/23S rRNA-LDH-S (16S/23S native rRNA intergenic promoter driving the expression of the sgRNA with a spacer targeting the start region of ldh gene)</td>
<td>This study</td>
</tr>
</tbody>
</table>
This study used CRISPRi as transcriptional repression tool for *H. thermocellum*.

### Table S2 Plasmids used in the present study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Primers (Fw &amp; Rv)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_1313_23S_NT</td>
<td>BG11739 &amp; BG11740</td>
<td>Plasmid with the ThermodCas9 and the non-targeting sgRNA, which is under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313.</td>
</tr>
<tr>
<td></td>
<td>BG11747 &amp; BG10829</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG10830 &amp; BG11746</td>
<td></td>
</tr>
<tr>
<td>p_1313_ldh_23S_P</td>
<td>BG11739 &amp; BG11741</td>
<td>Plasmid with the ThermodCas9 and the targeting sgRNA under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313. SgRNA with spacer 1 targeting promoter of lactate dehydrogenase gene.</td>
</tr>
<tr>
<td></td>
<td>BG10814 &amp; GB10829</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG10830 &amp; BG11746</td>
<td></td>
</tr>
<tr>
<td>p_1313_ldh_23S_S</td>
<td>BG11739 &amp; BG11742</td>
<td>Plasmid with the ThermodCas9 and the targeting sgRNA under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313. SgRNA with spacer 2 targeting start of the lactate dehydrogenase gene.</td>
</tr>
<tr>
<td></td>
<td>BG11419 &amp; GB10829</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG10830 &amp; BG11746</td>
<td></td>
</tr>
<tr>
<td>p_1313_pta_23S_P</td>
<td>BG11423 &amp; BG10829</td>
<td>Plasmid with the ThermodCas9 and the targeting sgRNA under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313. SgRNA with spacer 1 targeting promoter of phosphotransacetylase gene.</td>
</tr>
<tr>
<td></td>
<td>BG13665 &amp; BG10830</td>
<td></td>
</tr>
<tr>
<td>p_1313_pta_23S_S</td>
<td>BG11425 &amp; BG10829</td>
<td>Plasmid with the ThermodCas9 and the targeting sgRNA under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313. SgRNA with spacer 2 targeting start of phosphotransacetylase gene.</td>
</tr>
<tr>
<td></td>
<td>BG13666 &amp; BG10830</td>
<td></td>
</tr>
</tbody>
</table>

### Table S3 Spacers used in the present study

<table>
<thead>
<tr>
<th>Genes targeted</th>
<th>Spacer sequences (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-targeting</td>
<td>GCTAGTCTCAAGGTCATCATGACT</td>
</tr>
<tr>
<td>Lactate dehydrogenase (Clo1313_1160)</td>
<td></td>
</tr>
<tr>
<td>Spacer targeting the promoter (Spacer 1)</td>
<td>TTATCAAGTAGGATATAATATTACT</td>
</tr>
<tr>
<td>Spacer targeting the beginning of the gene (Spacer 2)</td>
<td>AAGCTGTGGTGGAAACCTACAAAGCC</td>
</tr>
<tr>
<td>Phosphate acetyltransferase (Clo1313_1185)</td>
<td></td>
</tr>
<tr>
<td>Spacer targeting the promoter (Spacer 1)</td>
<td>CTCACTATAAACAGCTATTTTACTG</td>
</tr>
<tr>
<td>Spacer targeting the beginning of the gene (Spacer 2)</td>
<td>AACTCATAAAATTCTTTTCTTTTCAT</td>
</tr>
</tbody>
</table>
Table S4 Primers used in the present study

The primers used to amplify the *recA* gene were BG10431 and BG10432
The primers used to amplify the *gyrA* gene were BG10435 and BG10436
The primers used to amplify the *pta* gene were BG14575 and BG14580
The primers used to amplify the *ldh* gene were BG15853 and BG15854

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG10431</td>
<td>GAGAAGCAGTGGCAAAAGG</td>
</tr>
<tr>
<td>BG10432</td>
<td>CGTCGCTTTACCGGATGATT</td>
</tr>
<tr>
<td>BG10435</td>
<td>TGCGAATGCAATGTGGTTT</td>
</tr>
<tr>
<td>BG10436</td>
<td>TGGCAATTCTAATCTTCC</td>
</tr>
<tr>
<td>BG10814</td>
<td>TTATCAAGTAGGATATATAATTACTGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG10829</td>
<td>TCAATGACCAAAATCCCTAAGG</td>
</tr>
<tr>
<td>BG10830</td>
<td>CACGTTAAGGGATTTCGTCATGA</td>
</tr>
<tr>
<td>BG11419</td>
<td>AAGCCTGTTGTTGGAACCTCACAAGGCGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11423</td>
<td>CTCACTATAAACGCTATTTTACTGGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11425</td>
<td>AACTCATAAAATCCCCTTCTCCTTTCATGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11586</td>
<td>CGGCACTACGTCTATCGG</td>
</tr>
<tr>
<td>BG11587</td>
<td>ACGTGGAAAGAGCTAGTG</td>
</tr>
<tr>
<td>BG11636</td>
<td>ATCGGCATTACGTCTATCGG</td>
</tr>
<tr>
<td>BG11637</td>
<td>AACACGAAAGCCAGCAC</td>
</tr>
<tr>
<td>BG11642</td>
<td>CAGCTGGGTGTCATAGTTC</td>
</tr>
<tr>
<td>BG11643</td>
<td>CCTAAGAGTGGGGAATGC</td>
</tr>
<tr>
<td>BG11734</td>
<td>CCATTCTTTGTTTTAATCCCTGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11735</td>
<td>CCAGCTGGATACCTAGCGTCATGGGATCCCTTCTAAAGGATATAATGACT</td>
</tr>
<tr>
<td>BG11739</td>
<td>CCATTCTTTGTTTTAATCCCTGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11740</td>
<td>CCAGCTGGATACCTAGCGTCATGGGATCCCTTCTAGGTGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG11741</td>
<td>AGTAATATATATATATCTACTGGTAAACCTACAGGTCTCTTCCCTTTG</td>
</tr>
<tr>
<td>BG11742</td>
<td>GGCTTTGTAGGGTTCACGCTTCTCAGGTCTCTTTCCCTTTG</td>
</tr>
<tr>
<td>BG11745</td>
<td>AAAAGTCGACAGGATTAACAAACAAATGG</td>
</tr>
<tr>
<td>BG11746</td>
<td>AGGCCTGCGACAGGATTAACAAACAAATGG</td>
</tr>
<tr>
<td>BG11747</td>
<td>GGATCCCATGAGCGTACGTATCCAGCGTGGGTCATGTCATGTCATGCCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG12577</td>
<td>GTTCAGGGCTTTTGTAGGTTC</td>
</tr>
<tr>
<td>BG12578</td>
<td>TTAGCCCTTTTTTGTTGG</td>
</tr>
<tr>
<td>BG13665</td>
<td>ATGAAAGGGAAGAAGAATTATGAGTTCTCAGGTCTCTTTCCCTTTG</td>
</tr>
<tr>
<td>BG13666</td>
<td>ATTAATGGAAGGGCGAAAATCGACCCCTACAGGTCTCTTTCCCTTTG</td>
</tr>
<tr>
<td>BG14575</td>
<td>TGAAAGGGCGGAATACAGACG</td>
</tr>
<tr>
<td>BG14580</td>
<td>CACTCCTTCAGATCAATATCC</td>
</tr>
<tr>
<td>BG15853</td>
<td>GGTGCAGGGCTTTGTAGGTTC</td>
</tr>
<tr>
<td>BG15854</td>
<td>TTAGCCCTTTTTTGTTGG</td>
</tr>
</tbody>
</table>
Chapter 4

Breaking the restriction barriers and applying a gene silencing tool in *Pseudoclostridium thermosuccinogenes*

Joyshree Ganguly\(^1\), Maria Martin-Pascual\(^2\)*, Alkan Bulut\(^3\), Bram Vermeulen\(^2\), Ivo Tjalsa\(^2\), Richard van Kranenburg\(^1,2\)

\(^1\)Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands

\(^2\)Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

\(^3\)Fontys University of Applied Sciences, 5612 AR Eindhoven, The Netherlands

*contributed equally

Manuscript in preparation:

Ganguly, J., Martin-Pascual, M., Bulut, A., Vermeulen, B., Tjalsa, I., and van Kranenburg, R.


ABSTRACT

*Pseudoclostridium thermosuccinogenes* is a thermophilic bacterium capable of producing succinate from lignocellulosic-derived sugars and has the potential to be exploited as a platform organism. However, exploitation of *P. thermosuccinogenes* has been limited partly due to the genetic inaccessibility and lack of genome engineering tools. In this study, we established the genetic accessibility for *P. thermosuccinogenes* DSM5809. By overcoming restriction barriers, transformation efficiencies of $10^2$ CFU/µg plasmid DNA were achieved. To this end, the plasmid DNA was methylated *in vivo* when transformed into an engineered *E. coli* HST04 strain expressing four native methylation systems of the thermophile. The protocol was used to introduce a ThermodCas9-based CRISPRi tool targeting the gene encoding malic enzyme in *P. thermosuccinogenes*, which resulted in a 75% downregulation of its expression. Moreover, the silencing of malic enzyme had an impact on the strain’s fermentation profile. This is the first example of genetic engineering in *P. thermosuccinogenes*, opening new possibilities for metabolic engineering of this bacterium.
INTRODUCTION

Metabolic engineering and synthetic biology are currently driving the development of genetically improved microorganisms to produce a compound of interest (Vickers, et al., 2012). To this end, two main different approaches have been applied. On one hand, engineering of mesophilic model organisms, such as *Escherichia coli* or *Saccharomyces cerevisiae*, to produce non-native compounds by introducing heterologous pathways in an efficient, rapid and high-throughput manner. However, challenges such as cofactor imbalance, genetic instability or toxicity due to heterologous gene expression need to be tackled for optimal performance (Wu, et al., 2016, Czajka, et al., 2017). Alternatively, non-model production strains with their native ability to produce a target compound can be exploited. In this case, the genetic accessibility and the availability of an extensive genetic toolbox are crucial factors when genetic engineering is needed to further improve production titers and yields.

Over the last decades, non-model thermophilic microorganisms such as *Clostridium* species, have been exploited as whole-cell biocatalysts (Joseph, et al., 2018). Thermophiles are beneficial over mesophiles with better substrate and product solubility, increased reaction rates as well as reduced downstream-processing and cooling costs (Jiang, et al., 2017). In addition, thermophilic anaerobes like clostridia are preferred over aerobes that suffer the decreasing productivities at low oxygen transfer rates and the costly aeration systems in industrial fermentations (Weusthuis, et al., 2011). Despite the above-mentioned advantages of using thermophilic clostridia as microbial cell factories, limitations like genetic inaccessibility and limited genome editing tools hinder research on clostridial physiology and metabolism and its metabolic engineering (Pyne, et al., 2014, Joseph, et al., 2018). The reasons for genetic inaccessibility are the inability of the plasmid DNA to penetrate the bacterial cell wall - notably the thick layer of peptidoglycan is characteristic to Gram-positive bacteria - or its intracellular degradation due to the presence of anti-phage defense systems, such as restriction-modification (RM) systems (Tock and Dryden, 2005, Vasu and Nagaraja, 2013), CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats loci, coupled to CRISPR-associated genes) (Marraffini, 2015, Mohanraju, et al., 2016, Koonin and Makarova, 2019) and the defense island system associated with restriction-modification (DISARM) system (Ofir, et al., 2018).

The anti-phage mechanisms avoid not only phage invasion or gene transfer among bacteria, but also genetic manipulation as the active RM systems cleave foreign DNA that is not correctly
methylation. The RM systems comprise two enzymes, a DNA methyltransferase (MTase) and a restriction endonuclease (REase). The MTase methylate specific nucleotides in a DNA sequence within the host’s genome, discriminating non-self and self-DNA and protecting the latter from digestion. The REase recognizes the same sequence as the MTase and cleaves the foreign DNA that lack proper methylation. Four different RM systems, type I, II, III and IV exists that differ in enzyme composition and mode of action. Type I is a multi-subunit complex consisting of two restriction (R), two modification (M) and a specificity (S) subunit (R$_2$M$_2$S), respectively. It requires all subunits (R$_2$M$_2$S) for restriction activity. In contrast, Type II RM system involve only two enzymes; restriction endonuclease (RE) and the methyltransferase (MT). Type III system comprise two R and M subunits. Contrarily to all other RM systems, type IV lacks the MT enzyme. The RE recognises non-specific and variable target sites, which are methylated, hydroxy-methylated or glucosyl-hydroxy methylated patterns (López-Larrea, 2011, Suzuki, 2012, Vasu and Nagaraja, 2013). As a result, RM systems are responsible for an all-or-nothing effect in the transformation efficiency of some difficult to transform bacteria such as *Clostridium* or *Bacillus*. Subsequently, they have greatly hindered the study of the clostridial physiology as well as its exploitation for industrial applications (Dong, et al., 2010).

Despite the difficulty to genetically transform clostridia, one successful strategy would be mimicking host methylation patterns to overcome the restriction barrier in the bacterium that harbours multiple RM systems (Zhang, et al., 2012, Riley, et al., 2019). Such strategy requires an *E. coli* strain deficient in characterized RM systems and orphan MTases (*dam*’, *dcm*’, *hsdRMS*), such as *E. coli* HST04; because the target bacteria might express not only type I, II and III RM systems but also type IV systems (Suzuki, 2012, Zhang, et al., 2012). Thus, the plasmid of interest can be propagated without being appropriately methylated. Otherwise, it presents a methylation pattern different to the native DNA, being consequently recognized as foreign DNA and cleaved. To produce host-mimicking DNA, prior identification of native RM systems as well as their recognition sequences was required. Once the recognition sequences on transforming plasmids have been identified, they can be protected via methylation. This methodology to overcome the restriction barrier for genetic accessibility of non-model organisms is described in detail in Figure 4.1.

There are two main approaches to methylate DNA, both *in vitro* and *in vivo*, depending on the type of RM systems (Suzuki, 2012). *In vitro* methylation uses commercially available methyl transferases prevalently used for type II RM system. Although it is a simple and highly efficient
strategy, the main drawback is the limited number of methyl transferases. In contrast, in vivo methylation as shown in Figure 4.1 consists of expressing the genes encoding MTases in the genome of an E. coli lacking all known RM systems and orphan MTases. While type I methylation can be reproduced by simultaneous expression of M and S subunit genes, type II and III methylation can be mimicked by expressing M subunit genes (Purdy, et al., 2002, Suzuki, et al., 2011, Suzuki, 2012). These approaches in combination with bioinformatics analysis using single-molecule real-time (SMRT) sequencing from Pacbio (Roberts, et al., 2013, Jensen, et al., 2019) or Nanopore Minion sequencing (Beaulaurier, et al., 2019) which are advantageous to get an idea on the methylome of the strains. Moreover, this method assists in overcoming restriction modification systems to enhance genetic accessibility in various bacteria (Sandoval, et al., 2015, McIntyre, et al., 2019, Riley, et al., 2019).

*Pseudoclostridium thermosuccinogenes* is a thermophilic anaerobe that produces succinate and acetate, as major products, and formate, lactate and ethanol as minor products, from inulin, glucose or xylose (Sridhar and Eiteman, 1999, Sridhar, et al., 2000, Koendjbiharie, et al., 2018). Accordingly, to become attractive as a succinate producer, *P. thermosuccinogenes* needs to be optimized by genome engineering and genetic accessibility, which are crucial aspects. Genetic engineering of *P. thermosuccinogenes* strains has not been described to date. *Hungateiclostridium thermocellum* is a close relative for which genetic tools have been developed. However, these are still laborious and inefficient; whilst we desire to develop simple, precise and high-throughput tools. Repurposing the CRISPR-Cas9 technology has paved the way for establishment of efficient genome editing tools in various organisms (Gasiunas, et al., 2012, Jinek, et al., 2012, Cong, et al., 2013, Jiang, et al., 2013, Qi, et al., 2013), but has not been applied for *H. thermocellum* so far. Two thermophilic Cas9 variants GeoCas9 (Harrington, et al., 2017) and ThermoCas9 (Mougiakos, et al., 2017) were applied in *H. thermocellum*, recently. A CRISPR interference (CRISPRi) system using the ThermoCas9 was adjusted for *H. thermocellum* to downregulate phosphotransacetylase and lactate dehydrogenase genes (Ganguly, et al., 2019). Another study developed the Type II (GeoCas9) CRISPR/Cas system together with recombineering machinery for efficient genome editing purposes in the same bacteria (Walker, et al., 2020). Nevertheless, CRISPRi is particularly suitable to target metabolic genes or pathways which require a basal expression. Besides, it serves as a powerful method for fast assessment of the possible impact of the intended genetic modifications in the microbial metabolism (Mougiakos, et al., 2017, Mougiakos, et al., 2018, Walker, et al., 2020).
In this study, we developed a protocol for transformation of *P. thermosuccinogenes* DSM5809 by electroporation. We applied an *in vivo* methylation pipeline mimicking DNA methylation patterns of *P. thermosuccinogenes* DSM5809 in an engineered *E. coli* HST04 strain expressing the four-native type I RM systems and lacking any other known RM systems to avoid unintended modifications. Plasmids prepared from this host showed increased transformation efficiency, apparently, escaping the clostridia’s restriction systems. After successfully overcoming the restriction barriers of this strain, as a proof-of-concept, we developed CRISPRi silencing tool for *P. thermosuccinogenes* DSM5809 (Figure 4.1). To conclude, this research demonstrates for the first-time the genetic accessibility of *P. thermosuccinogenes* followed with the development of a CRISPRi tool for downregulating the malic enzyme, unravelling a way for prevalent metabolic engineering of this thermophile.

**Figure 4.1.** Schematic overview for *in vivo* methylation and CRISPRi application for *P. thermosuccinogenes* DSM5809. a) The genome sequence and methylome of *P. thermosuccinogenes* was used to identify the native restriction modification (RM) systems. b) A set of *E. coli* HST04 derivative strains was constructed by cloning of the methyltransferase genes (orange and green) of *P. thermosuccinogenes*. c) The ThermodCas9 plasmid of interest (pCRISPRi) was introduced in these *E. coli* strains by electroporation. The methylated pCRISPRi plasmid DNA was isolated from this strain and d) electroporated into *P. thermosuccinogenes*. In this way, the pCRISPRi presents a methylation pattern like that of the genome of *P. thermosuccinogenes*, overcoming the restriction barrier. e) pCRISPRi expresses the ThermodCas9 and the sgRNA, downregulating the gene of interest upon recognizing and binding the target DNA (promoter region). Transcription is blocked since the Cas9 protein physically interferes with the RNA polymerase. f) Gene repression of desired gene was evaluated by qRT-PCR.
RESULTS

Initial transformation steps for *P. thermosuccinogenes* DSM5809

We started development of our transformation protocol using plasmid pNW33n, which was known to replicate in *H. thermocellum* DSM1313 (Olson and Lynd, 2012). First, we examined the effect of *dam* and *dcm* methylation on electroporation.

Plasmids were extracted from *E. coli* strains with different methylomes DH5α (*dam*+/*dcm*−), Acella (*dam*+/*dcm*−), JM110 (*dam*−/*dcm*−) and HST04 (*dam*−/*dcm*−/Δ(mrr-hsdRMS-mcrBC)) to prepare methylated/unmethylated plasmid DNA (Table S1, Supplementary data). Electroporation was performed using plasmid pNW33n prepared from *E. coli* DH5α, Acella, JM110 and HST04 strains with conditions adapted from Olson *et. al.* for *H. thermocellum* (Olson and Lynd, 2012), with modifications as mentioned in detail in the Experimental procedures section. For our transformation protocol, the major modifications were the preparation of electrocompetent cells at room temperature, using suitable electroporation settings (1.8 KV, 200 Ω and 25 μF, exponential decay pulse), recovery time of cells in CTFUD-Tm liquid medium after electroporation was restricted between 10-12 hours to obtain transformants. Afterwards the cells were plated on CTFUD-Tm agar medium, a few thiamphenicol (TmR) resistant colonies appeared only for plasmids prepared from *E. coli* JM110 and HST04 strains. This shows the importance of unmethylated DNA for successful transformation in *P. thermosuccinogenes*. The TmR colonies were verified by colony PCR and plasmid DNA was isolated from *P. thermosuccinogenes* and re-transformed in *E. coli* DH5α. After confirmation that the *P. thermosuccinogenes* transformants contained pNW33n, we aimed to improve transformation efficiency and reproducibility by various strategies aiming at overcoming restriction barriers.

Bioinformatics analysis of restriction modification systems in *P. thermosuccinogenes*

*P. thermosuccinogenes* DSM5809 genome sequence was assembled using data from Illumina HiSeq and Pacbio sequencing (Koendjibiharie, *et al.*, 2018). The annotation included genes annotated as putative restriction systems and DNA methyltransferases. Four type I restriction systems and one type III system were identified that could negatively impact the transformation efficiency. SMRT sequencing data were further analysed to identify all base modifications and to find modified sequence motifs (Table 4.1). The genome sequence contains three gene clusters of the predicted type I RM systems in different scaffolds. Scaffold 1 comprises of overlapping
gene clusters of two type I RM systems, Scaffold 2 and Scaffold 4 each has the other gene clusters, respectively, as illustrated in detail in Figure 4.2.

Table 4.1. Methylome data of *P. thermosuccinogenes* DSM5809

<table>
<thead>
<tr>
<th>RM system</th>
<th>RM type</th>
<th>Motifs</th>
<th>Modification Type</th>
<th>% Modified</th>
<th>Motifs in genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>CACNNNNNNNTNGC/ GCANNNNNNNGTG</td>
<td>m6A</td>
<td>95.09/88.47</td>
<td>876</td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td>GATNNNCTC</td>
<td>m6A</td>
<td>94.29</td>
<td>1630</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>DGAGNNNNATC</td>
<td>m6A</td>
<td>92.6</td>
<td>1461</td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>TCABNNNNNTARG/ CYTNNNNNTGA</td>
<td>m6A</td>
<td>90.24/85.79</td>
<td>697</td>
</tr>
<tr>
<td>5</td>
<td>III</td>
<td>CGAG</td>
<td>m6A</td>
<td>66.04</td>
<td>10352</td>
</tr>
</tbody>
</table>

Strategies to overcome restriction barrier to improve transformation efficiency

The pNW33n plasmid was examined for the presence of putative RM recognition sites retrieved from the methylome analysis of *P. thermosuccinogenes* DSM5809 (Table S5, Supplementary data). We identified three recognition sites of motifs 2 and 3 (Table 4.1) on the plasmid that might potentially hinder the transformation of the aforesaid plasmid. Since few transformants were obtained previously, the hypothesis was that some transforming plasmids can escape restriction and become methylated by the native methyltransferases, enabling propagation. This might also explain the low reproducibility and efficiency of the transformation experiments. To test said hypothesis two strategies were followed. Firstly, the detected recognition sites were mutated, so they cannot be recognized anymore by the corresponding RM systems. Secondly, the recognition sequences on pNW33n were protected by *in vivo* methylation. Thus, *E. coli* strains were engineered by introducing native genes encoding type I RM systems from *P. thermosuccinogenes* DSM5809. For the first approach, three different recognition sites in the backbone vector including the repB gene and the promoter region of the chloramphenicol acetyltransferase (*cat*) gene were detected. The *P. thermosuccinogenes* codon bias was established using the Codon Usage program of the Sequence Manipulation Suite (SMS) and considered to apply synonymous mutations with the codon usage closest to the original codon usage. The site that was present in the promoter region of the *cat* gene was not modified to avoid any side effect to the proper transcription of the chloramphenicol resistance gene. Thus,
the plasmid pNW33n_RM5809 (pNW33n lacking two RM recognition sites, except the one present in the promoter of the cat gene, Figure 4.3b) was electroporated into P. thermosuccinogenes to evaluate the transformation efficiency. The plasmids pNW33n_RM5809 and pNW33n (control) isolated from E. coli JM110 or HST04, did not show any significant difference in transformation efficiency, but the reproducibility was much better in comparison to control.

![Diagram of gene clusters](image)

**Figure 4.2.** Type I RM systems in the genome of P. thermosuccinogenes DSM5809. Genetic maps with gene functions of the different type I RM systems in P. thermosuccinogenes. Scaffold 1 harbours the gene clusters CDQ83_09145, CDQ83_09150, CDQ83_09155, CDQ83_09160 and CDQ83_09165. Scaffold 2 harbours the gene clusters CDQ83_16045 and CDQ83_16050. Scaffold 3 harbours the gene clusters CDQ83_18870 and CDQ83_18875.

As a second strategy, the host's methylation patterns were mimicked in E. coli and in vivo methylation of the RM recognition sequences on plasmid DNA was carried out. This strategy has been shown to significantly increase the transformation efficiency in various bacteria (Kwak, et al., 2002, Suzuki, et al., 2011, Suzuki and Yoshida, 2012, Zhang, et al., 2012). Linear cassettes, carrying different predicted type I RM systems identified in the genome of P. thermosuccinogenes DSM5809 were designed by making use of three different inducible promoters that would control the expression of the methylation genes. To know which methylation system had the strongest effect in the transformation efficiency, we engineered...
three different strains with individual RM systems or a single strain with the four RM systems together. Three cassettes (referred as S1, S2 and S4, Figure S2 and Table S7) comprised of the different type I RM systems from the three scaffolds of *P. thermosuccinogenes* were successfully generated via overlapping PCR and integrated in the genome of *E. coli* HST04 by Lambda/Red recombineering. The generated *E. coli* HST04 strains with the integrated cassettes were used next as intermediate strains to methylate the pH3n and pH3n_RM5809 (Figure 4.3b) plasmids with the native methylation patterns of *P. thermosuccinogenes* DSM5809 prior its electroporation (Figure 4.3a). The combination of the three methylation systems, S1, S2 and S4 on pH3n and pH3n_RM5809, showed the highest effect on the transformation efficiency with increase in two orders of magnitude, corresponding to 6.6*10^2 CFU/µg DNA and 3.2*10^2 CFU/µg DNA, respectively as shown in Figure 4.3c. Hence, the *in vivo* methylation has a positive influence on the increase of transformation efficiency.

**CRISPRi as a silencing tool for *P. thermosuccinogenes***

CRISPRi is a genetic perturbation tool that represses gene expression in bacteria. As a proof of concept, to study the efficacy of CRISPRi for gene suppression in *P. thermosuccinogenes*, we targeted the non-template strand of malic enzyme gene. We created pThermoCas9i vectors with ThermodCas9 under *xylL* promoter from *B. smithii* and sgRNA with the intergenic 16S/23S rRNA promoter from *H. thermocellum* (Figure 4.4a), respectively, which was previously demonstrated to be functional for gene silencing in *H. thermocellum* (Ganguly, et al., 2019). The sgRNA was with a non-targeting spacer or a spacer targeting the promoter region of malic enzyme gene. Targeting and non-targeting plasmids were electroporated to *P. thermosuccinogenes* DSM5809. The colonies of *P. thermosuccinogenes* harboring non-targeting and targeting plasmid showed expression of the ThermodCas9 and sgRNA genes using RT-PCR (Figure 4.4b and 4.4c). Consequently, qRT-PCR was performed to analyze the silencing efficacy of the transformants using 16S rRNA as house-keeping gene. The qRT-PCR analysis (Figure 4.4d) showed 75 % reduction of the targeting gene expression, in comparison to the non-targeting transformants. This indicates the suppression effectiveness of the targeting gene using CRISPRi.
Breaking the restriction barriers in *P. thermosuccinogenes*

**Figure 4.3.** Strategies to enhance transformation efficiency by breaking the restriction barriers. **a)** Host mimicking strategy to overcome restriction barriers by *in vivo* methylation of plasmid DNA. **b)** Schematic illustration of plasmid pNW33n_RM5809 with silent mutations marked (*); ori - pUC19 origin of replication for plasmid propagation in *E. coli*; oriB - pNW33n origin of replication for plasmid propagation in *P. thermosuccinogenes*; repB - replication protein for pNW33n origin of replication; cat - chloramphenicol acetyl-transferase, provides resistance to chloramphenicol and thiamphenicol. **c)** Transformation efficiency of *P. thermosuccinogenes* using the plasmids pNW33n and pNW33n_RM5809 isolated from modified *E. coli* HST04 strains consisting of all methylation patterns. Electroporation were performed in triplicates and the standard deviation.

Furthermore, we tested both the non-targeting and targeting transformants for HPLC analysis to get an insight into fermentation profile. The major products of *P. thermosuccinogenes* are succinate, acetate and formate, lactate and ethanol as minor products. It was evidently noted that repression of the gene encoding the malic enzyme led to an effect only on lactate and acetate production in comparison to non-targeting cells, respectively (Figure 4.5b). At OD$_{600}$ ~ 1.0 the acetate production decreased to 28% in contrast to lactate production, which increased by 33% in *P. thermosuccinogenes* with the malic enzyme gene repressed. The repression has affected the pyruvate node directly, which might have driven the flux towards the closest lactate branch. Other products such as succinate, formate and ethanol did not show any significant differences.
between the targeting and the non-targeting transformants (Supplementary Figure S3). In addition, no substantial variances were observed for malate and pyruvate production. In summary, qRT-PCR and HPLC analysis showed decrease in acetate and enhanced lactate production at both exponential and stationary growth phase of the fermentation. This shows that ThermodCas9, as a proof of concept, can silence metabolic genes in *P. thermosuccinogenes* with an impact on product formation.

Figure 4.4. Transcriptional repression of malic enzyme gene by CRISPRi of *P. thermosuccinogenes* transformants. a) Schematic representation of non-targeting plasmid pThermoCas9i (Mougiakos, et al., 2017). The *thermodcas9* gene under control of the *B. smithii* *xylL* promoter; sgRNA-expressing module under control of the native 16S/23S intergenic rRNA promoter from *H. thermocellum*; pNW33n backbone; asterisks (*) represent the regions with silent mutations. b) Non-targeting 1 and 2 clones, RT-PCR of thermodcas9 cDNA (lanes 2 and 3) with product size 124 bp; WT, RT-PCR with *P. thermosuccinogenes* DSM5809 wild-type cDNA (lane 4); *B. smithii*, RT-PCR with *B. smithii* harbouring pThermoCas9i_NT (Mougiakos, et al., 2017) cDNA (lane 1). c) Non-targeting 1 and 2 clones, RT-PCR of sgRNA with product size 169 bp (lanes 2 and 3); *B. smithii*, RT-PCR with *B. smithii* harbouring pThermoCas9i_NT (Mougiakos, et al., 2017) cDNA (lane 1). d) Effective downregulation of malic enzyme gene was evaluated using qRT-PCR in comparison to control non-targeting at an OD<sub>600</sub> ~ 1.0; silencing transformants for malic enzyme gene targeting the promoter. *P. thermosuccinogenes* DSM5809 cells were transformed with both non-targeting and targeting plasmids. The Tn<sup>8</sup> colonies of both the plasmids were transferred to 50 mL CP (5 g/L yeast extract) medium containing 6 µg/mL Tm and cultured to OD<sub>600</sub> ~ 1.0. 5 mL cells were sampled for qRT-PCR analysis with technical triplicates. The expression levels were normalized to those in the non-targeting control. Data represent the mean values of three biological replicates and the standard deviation. The level of significance of the differences when results were compared was estimated by means of analysis of variance (ANOVA), with α=0.05.
**DISCUSSION**

*P. thermosuccinogenes* DSM5809 is an anaerobic thermophile with potential for succinate production (Sridhar, et al., 2000, Koendjbiharie, et al., 2018). Development of a thermophilic production process for succinic acid using renewable resources could be achieved via an engineered *P. thermosuccinogenes*. Therefore, we developed a reproducible electroporation protocol to introduce plasmid DNA in *P. thermosuccinogenes*. A series of pilot experiments were conducted with *P. thermosuccinogenes* DSM5809, based on literature protocols for other clostridial species (Purdy, et al., 2002, Tyurin, et al., 2004, Song, et al., 2007, Guss, et al., 2012, Pyne, et al., 2013, Pyne, et al., 2014, Kolek, et al., 2016, Herman, et al., 2017). To study the genetic accessibility of *P. thermosuccinogenes*, *E. coli* DH5α-Bacillus shuttle vector pNW33n encoding chloramphenicol resistance marker (Olson and Lynd, 2012) was chosen based on two vital features: i) the origin of replication to propagate in both *E. coli* (cloning host) and *Clostridium* (host of interest), and ii) limited availability of antibiotic markers (being a thermophile). Another *E. coli*-Geobacillus shuttle vector pUCG3.8 encoding the thermostable
kanamycin nucleotidyl transferase gene (Bartosiak-Jentys, et al., 2013), was also tested but failed to give transformants. The reasons might be that the plasmid with *Geobacillus* origin of replication was unable to propagate in clostridia or the presence of recognition sites 1, 2 and 3 (Table 4.1) in the plasmid, hampered DNA uptake. Furthermore, electroporation was done using various growth states of cells (OD$_{600}$ 0.4-1.2), electroporation buffers (SMP, glycerol, ddH$_2$O), electroporation cuvettes (0.1 and 0.2 cm gap), and electrical parameters (field strength 2.5-15 kV/cm, time constant 5-20 ms). Unfortunately, none of the settings yielded transformants, suggesting the presence of RM systems correlated with improper DNA methylation or other critical factors like thick cell wall of Gram-positive bacteria, electroporation settings and physiological state of the cells that hinder DNA uptake.

The plasmid pNW33n, previously introduced and isolated from *E. coli* DH5α (*dam*+/*dcm*+), was used to transform different intermediate *E. coli* strains (Acella, JM110 and HST04) prior their transformation in *P. thermosuccinogenes*. As a result, only pNW33n plasmid extracted from *E. coli* JM110 and HST04 yielded colonies with the electroporation protocol. Both strains have the *dam* and *dcm* genes deleted. The difference between the two strains is the additional deletion of genes encoding type I and IV RM systems in *E. coli* HST04 (*Δmrr-hsdRMS-mcrBC*). The absence of type I RM genes is crucial to avoid different methylation pattern of plasmids to the clostridia methylation. Moreover, the presence of plasmids with varied methylation pattern from *P. thermosuccinogenes*, could lead to the activation of native type IV RM systems in clostridia, which cleaves methylated DNA (Suzuki and Yoshida, 2012, Zhang, et al., 2012). Therefore, the fact that only plasmids isolated from *E. coli* HST04 could be transformable in this thermophile, in contrast to other *E. coli* strains, hint at the presence of unidentified type IV RM systems in *P. thermosuccinogenes*.

To gain insight into the RM systems of *P. thermosuccinogenes*, Pacbio methylation analysis was used to identify the presence of RM motifs in the genome. Four type I and one type III RM systems were annotated in the genome sequence, and five methylated restriction sites were identified by the methylation analysis. This allowed us to search for the presence of putative RM recognition sites of *P. thermosuccinogenes* DSM5809 in pNW33n plasmids. We identified three recognition sites that might potentially abolish the acquisition of pNW33n by *P. thermosuccinogenes*. Since few transformants were obtained when *E. coli* JM110 and HST04 were used as intermediate hosts, we hypothesized an escape strategy of the plasmids from the native RM systems enabling propagation into the host. This might also explain the low reproducibility and efficiency of the transformation protocol. To test said hypothesis two
strategies were followed. On one hand, the detected recognition sites were mutated, so they cannot further be recognized by the identified RM systems. On the other hand, the transforming plasmids were protected by in vivo methylation. In this approach, the native genes encoding type I RM systems from *P. thermosuccinogenes* DSM5809 were mimicked in *E. coli* strains and plasmid DNA was isolated from these specific strains to overcome restriction barrier. In the first strategy, the plasmid pNW33n_RM5809 (pNW33n lacking the RM recognition sites, Figure 4.3b) was electroporated into *P. thermosuccinogenes* with improved reproducibility compared with the pNW33n. Regarding the second strategy, we observed improvements on reproducibility and on transformation efficiency. The transformation efficiency of $10^2$ CFU/μg DNA was achieved when plasmid was in vivo methylated in an engineered *E. coli* HST04 strain, expressing the four native methylation systems of *P. thermosuccinogenes*. This strategy has also significantly increased the transformation efficiency in other bacteria (Kwak, et al., 2002, Suzuki, et al., 2011, Suzuki and Yoshida, 2012, Riley, et al., 2019).

Once, the transformation protocol was improved on transformation efficiency and reproducibility, we aimed at developing CRISPRi tool as a proof of concept. Our recent publication on the adaptation of CRISPRi on *H. thermocellum* using a thermostable Cas9 (Ganguly, et al., 2019), opened new possibilities for transcriptional regulation of the thermophile *P. thermosuccinogenes*. Due to the close relatedness of the two clostridial species, we hypothesized that the CRISPRi plasmid used to downregulate metabolic genes from *H. thermocellum* might be functional in *P. thermosuccinogenes*. Dead Cas9 based CRISPRi allows the transcriptional regulation of the gene of interest without completely disrupting its function, resulting in less pleiotropic effects than gene knockouts (Dominguez, et al., 2016, Peters, et al., 2016). We targeted the central metabolic gene encoding NADP$^+$-dependent malic enzyme. With this, we successfully proved the functionality of the ThermoCas9i plasmid in *P. thermosuccinogenes*, achieving 75% reduction of the malic enzyme gene expression. Moreover, the silencing has an impact on the fermentation profile. We observed that the downregulation of the malic enzyme gene, results in increased lactate production and declined acetate production. The malic enzyme is part of the malate shunt pathway together with malate dehydrogenase and phosphoenolpyruvate carboxykinase for the conversion of pyruvate from phosphoenolpyruvate in *P. thermosuccinogenes* as depicted in Figure 4.5a (Koendjibiharie, et al., 2019). Due to the silencing the pyruvate accumulation was less, which seems to direct the flux towards lactate production at the cost of acetate. We expected effects on malate or pyruvate concentrations, but no significant variations were observed. In addition, prominent fluctuations
in other organic acids or ethanol were also not noticed. To improve the yields of desired products such as succinate it can be achieved with the knockout or knockdown of genes involved in the malate shunt pathway or lactate dehydrogenase gene, as indicated by previous studies. For instance, _H. thermocellum_ YD02 strain was created with heterologous expression of pyruvate kinase and deletion of malic enzyme plus lactate dehydrogenase genes. This strain showed increase in ethanol and formate production (Deng, et al., 2013). Hence, we can apply both active and dCas9 CRISPR system for _P. thermosuccinogenes_ to manipulate other candidate genes and study its impact on succinate production with elimination of by-products.

In summary, the work demonstrates the successful development of a genetic transformation procedure for _P. thermosuccinogenes_. We have established a straightforward workflow to handle non-model organisms from genetic accessibility to developing silencing tool for transcriptional suppression of metabolic genes. The strategies made to overcome restriction barriers were the introduction of silent mutations and _in vivo_ methylation of plasmid DNA. Moreover, we used CRISPRi to efficiently down-regulate the expression of the metabolic gene that encodes the malic enzyme affecting the metabolism of the organism. As a result, this is the first study of an effective genetic tool development for _P. thermosuccinogenes_, signifying an advancement for this industrially-relevant bacterium for green chemicals production.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains and growth conditions**

All the strains with their respective characteristics used in the present study are described in Table S2, Supplementary data. Chemically competent _E. coli_ DH5α was used for cloning purposes. _E. coli_ JM110 (Yanisch-Perron, et al., 1985), Acella (Chemically Competent cells, Edge Bio) and HST04 (Stellar™ Competent Cells, _dam⁻/dcm⁻_, catalogue number: 636764) were used for plasmid propagation and isolation before _P. thermosuccinogenes_ DSM5809 transformation. _E. coli_ strains were cultured in Lysogeny-Broth (LB) medium at 37 °C and 200 rpm, unless otherwise specified. Antibiotics were added when required, at the following concentrations: chloramphenicol, 20 μg/mL; ampicillin, 150 μg/mL; kanamycin, 20-50 μg/mL. _P. thermosuccinogenes_ DSM5809 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). _P. thermosuccinogenes_ DSM5809 wild type and transformants were grown anaerobically in CP medium adapted from (Plugge, 2005). The final volume of 50 mL medium was dispensed in serum bottles under 80:20 N₂/CO₂ atmosphere with
~70 kPa overpressure and then autoclaved. The wild type and transformants were obtained in CTFUD agar plates adapted from Olson et al., and grown at 50-60 °C range (Olson and Lynd, 2012). Glucose as a carbon source was also autoclaved separately and added later to a final concentration of 5.0 g/L. If needed thiamphenicol, 6 µg/mL was added.

For *P. thermosuccinogenes* spore purification, it was grown anaerobically at 60 °C overnight in CP medium. In order to trigger sporulation, 1 mL of pre-culture was added to sporulation medium adapted from Yang et al. and B Mearls et al., (Yang, et al., 2009, B Mearls, et al., 2012), cultured at 55 °C for 2 days.

**Preparation and purification of *P. thermosuccinogenes* DSM5809 spores**

*P. thermosuccinogenes* DSM5809 was grown anaerobically at 60 °C overnight in CP medium. For sporulation, the Modified AEA Sporulation Medium Base (catalogue number 17170, Sigma Aldrich) was prepared and 30 mL dispensed into anaerobic bottles under 80:20 N₂/CO₂ with ~70 kPa overpressure and then autoclaved. After autoclavage the medium was supplemented with 4 % (v/v) of glucose solution, freshly prepared and filter sterilized 1.33 % (v/v) of sodium carbonate (Na₂CO₃) and 1.33 % (v/v) cobalt chloride hexahydrate (CoCl₂·6H₂O), respectively. Then, 1.33 % v/v of freshly prepared sodium ascorbate solution (C₆H₃O₆Na) was added. 1 mL of the pre-culture was added to the sporulation medium and incubated at 55 °C. Once spores were formed 1 mL of pre-culture from sporulation medium was inoculated into 50 mL CP medium and grown for 2 days at 60 °C. When mature spores were highly predominant, spore suspension was harvested aerobically by centrifugation at 4800 g for 15 minutes. Cell pellets were resuspended, gently layered on 50% Histodenz solution and centrifuged at 3000 g, 20 °C for 60 minutes. Spore pellet was collected and resuspended in 15 mL Milli-Q water. Centrifuged at 1500 g, 20 °C for 30 minutes and supernatant was discarded. Mature spores pellets were resuspended in 10 mL Milli-Q water. Samples were examined under a phase contrast microscope (Figure S1, Supplementary data). 1 mL aliquots of mature spores were stored at -80 °C.

**Plasmid construction**

All PCR reactions for cloning were performed with the NEB Q5® High-Fidelity DNA polymerase, according to the manufacturer’s instructions (M0491). PCR fragments were subjected to 1% w/v agarose gel electrophoresis, and isolated using Zymoclean™ Gel DNA Recovery kit. Plasmids were built NEBuilder® HiFi DNA assembly cloning kit according to the manufacturer’s protocol. The plasmids developed in this study, as well as the primers used
to construct them, together with a detailed overview of the setup, are presented in Table S3, Supplementary data. For genetic accessibility experiments, *E. coli*-Bacillus shuttle vector pNW33n containing origin of replication that functions in *H. thermocellum* and chloramphenicol/thiamphenicol antibiotic resistance was used. For the CRISPRi experiments the pThermoCas9i plasmid with rRNA promoter for sgRNA expression, developed in a previous work (Ganguly, et al., 2019) was used as a template for the construction of a plasmid targeting the malic enzyme (CDQ83_17115_S2: 987621_988793) in *P. thermosuccinogenes* DSM5809. The sequence of the non-targeting and targeting spacers is presented in Table S4, Supplementary data.

*P. thermosuccinogenes* DSM5809 genome encodes multiple RM systems, specifically four type I and one type III RM systems have been identified. To increase the strain’s genetic transformability, pThermoCas9i plasmids were analysed to find possible recognition sites of such RM systems. The only found recognition sites by the type I RM systems of *P. thermosuccinogenes* DSM5809 were abolished by introducing point mutations. Five nucleotides were replaced by the most suitable synonymous substitutions (Table S5, Supplementary data), taking into consideration the codon usage bias of *P. thermosuccinogenes* DSM5809 (Table S6, Supplementary data).

All plasmids were introduced by heat-shock in chemically competent *E. coli* DH5α cells. Single colonies present in the selective medium were examined. They were grown in LB cultures with the appropriate antibiotic. Plasmids were isolated by using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) and confirmed by standard sequencing from Macrogen (MACROGEN Inc. DNA Sequencing Service; Amsterdam, The Netherlands) using the primers as shown in Table S7, Supplementary data. Plasmids were electroporated in *E. coli* HST04 and *E. coli* JM110 before being electroporated in *P. thermosuccinogenes*. Due to the exceptionally high concentrations of plasmid DNA required for the transformations in clostridial species, specifically in *P. thermosuccinogenes*, midiprep was performed to efficiently extract the plasmid from the intermediate *E. coli* strains, according the manufacturer’s instructions of ZymoPURE II™ plasmid isolation kit.

**Electroporation of *P. thermosuccinogenes* DSM5809**

Prior to genetic transformation by electroporation, 1 mL of dormant mature spores, stored at -80 ºC was heat activated at 80 ºC for 15 minutes. Optimal heat activation decreases the lag phase, allowing more spores to germinate (Wang, et al., 2011). The activated spores were in
turn inoculated into 50 mL CP medium. Spores were germinated overnight at 60 °C. Next day, 4-6 mL of the overnight culture was re-inoculated into fresh CP medium, with OD_{600} ~ 0.1. The bacterial culture was grown anaerobically at 60 °C to mid-log phase till OD_{600} ~ 0.6 and the culture was kept at room temperature for 20 minutes. To harvest the cells, the culture was brought into the anaerobic chamber and transferred to an anaerobic 50 mL Greiner tube. Cells were harvested aerobically by centrifugation at room temperature and 4800 g for 20 minutes. 20 mL of the 10% (v/v) glycerol wash buffer, was added without disturbing the pellet and centrifuged at room temperature, 4800 g for 15 minutes. The supernatant was removed, and the above washing steps were repeated. Consequently, cells were resuspended with 100-500 μL of 10% (v/v) glycerol. 80 μL aliquots of cells were made aerobically in 1.5 mL Eppendorf tubes. 1-4 μL, containing 4 μg of plasmid DNA, was added to each aliquot. The plasmids were previously introduced and isolated from *E. coli* HST04 (dam'/dcm') to ensure the transfer of unmethylated plasmids. The samples were taken into the anaerobic chamber, and the mixture was pipetted into 2-mm gap electroporation cuvette. A single exponential decay pulse was applied using a Gene Pulser X-Cell (Bio-Rad) set at 1.8 KV, 200 Ω and 25 μF. Immediately after the electrical pulse, cells were resuspended in 4 mL CP media, and incubated anaerobically at 50 °C inside a dry incubator for 10-12 hours. 100 μL and 2 mL of the transformed cells were plated in CTFUD media containing 6 μg/mL of thiamphenicol and 5 g/L of glucose. The agar plates were placed inside a 2.5 L anaerobic jar (ThermoFisher) with an anaerobic bag (ThermoFisher). The anaerobic jar, containing the transformed plates was placed inside a dry incubator at 50 °C. If the transformation succeeded, after 3-5 days colonies were visible. For the transformation of CRISPRi plasmids the protocol was modified with the electrical settings for electroporation. A single square wave pulse was applied at 1000 V and 5 milliseconds.

**Generating *E. coli* HST04 strains with methyltransferases genes of *P. thermosuccinogenes***

Genes encoding four different putative type I RM systems in *P. thermosuccinogenes* DSM 5809 were introduced in the genome of *E. coli* HST04 harbouring the Red/lambda system in the pKD46 plasmid. Three functional cassettes were constructed. The primers used to construct them, together with a detailed overview of the setup, are presented in Table S8, Supplementary data. Cassette S1 comprised the genes of an operon, encoding two type I RM systems, annotated in scaffold 1 (CDQ83_09145-CDQ83_09165). Cassette S2 contains the genes encoding the M and S subunits from a type I RM system, annotated in scaffold 2 (CDQ83_16045 and CDQ83_16050). Cassette S4 harbours the genes encoding the M and S subunits from the type
I RM system, annotated in scaffold 4 (CDQ83_18870-CDQ83_18875). The R subunits were not included in the cassettes. All the genes present in the cassettes are indicated in Figure 4.2. The functional cassettes were flanked by 100 bp long homology arms for the disruption of the targeted locus in the *E. coli* chromosome. The homology arms correspond to the upstream (5’ end) and downstream (3’ end) sequence, separated from each other ~ 1 kb in the insertion site. The synthetic linear cassettes are composed by two elements. The DNA methyltransferase expression module and the FRT-antibiotic resistance unit. The DNA methyltransferase expression module carries an inducible promoter, which controls the expression of the methylation genes and the T7 terminator, allowing efficient transcription termination. Aside from the methylation module, the FRT-antibiotic resistance unit is composed by an antibiotic resistance gene which is flanked by the FRT sites (Figure S2, Supplementary data). Therefore, the FRT-flanked antibiotic selection unit was excised by transformation of the flippase (FLP) expression vector, pCP20 into the *E. coli* HST04, harbouring the cassette. These engineered *E. coli* HST04 hosts were further used for transformations in *P. thermosuccinogenes*.

**RNA isolation and first strand cDNA synthesis**

RNA isolation of *P. thermosuccinogenes* transformants, harbouring the pThermoCas9i plasmids was performed using 5 mL of overnight cultures at an OD$_{600}$ ~ 1.0. The RNA isolation for qRT-PCR was adapted from Ganguly *et al.* (Ganguly, et al., 2019). Maxwell 16 LEV Total RNA Cells Kit was used to obtain RNA from the transformants. The purified RNA was measured in the NanoDrop spectrophotometer to determine the quality and the concentration of the elution and stored at -20 °C. The first strand cDNA synthesis was done following the manufacturer’s instruction of the SuperScript™ III Reverse Transcriptase protocol (Invitrogen). The primers BG11642 and BG11643 were used to amplify 169 bp of the sgRNA and the primers BG11636 and BG11637 were used to amplify 282 bp of the ThermodCas9 using the NEB Q5® High-Fidelity DNA polymerase.

**Quantitative Real Time PCR**

The qRT-PCR was performed by using the iQ™ SYBR® Green Supermix from Bio-Rad. The cDNA samples were diluted in sterile Milli-Q water. The amount of cDNA used as a template was equivalent to 50 ng of RNA. The house-keeping gene used to measure the relative expression was 16S rRNA of *P. thermosuccinogenes* DSM5809. The primers used to amplify 16S rRNA and malic enzyme gene of *P. thermosuccinogenes* were BG10427, BG10428 and...
BG18976, BG18977, respectively (Table S7, Supplementary data). The qRT-PCR was run in a Bio-Rad C1000 Thermal Cycler.

**High-pressure liquid chromatography**

A high-pressure liquid chromatography (HPLC) system ICS-5000 was used for the organic acids and ethanol quantification. The system has Aminex HPX 87H column from Bio-Rad Laboratories and is equipped with a UV1000 detector operating on 210 nm and a RI-150 40 °C refractive index detector. The mobile phase consisted of 0.16 N H₂SO₄ and the column was operated at 0.8 mL/minute. All samples were diluted 4:1 with 10 mM DMSO in 0.01 N H₂SO₄. *P. thermosuccinogenes* non-targeting control and malic enzyme gene silencing transformants were grown in CP medium (5 g/L yeast extract) for 2 days and samples were taken at different time points from OD₆₀₀ > 1.0 for analysis with HPLC. Sugars (cellobiose, glucose, ethanol and glycerol) and organic acids (malic acid, pyruvic acid, acetic acid, lactic acid, succinic acid and formic acid) were used as standards with a concentration range between 1.25-25 mM.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGMENTS**

Funding: This work was supported by Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068].
SUPPLEMENTARY DATA

S1 Promoter intergenic 16S/23S rRNA of *H. thermocellum* DSM1313

GCCTATGTCGGCTTTGAGAAGAATTGCAATATATAATTTGCACCTTGAAAAATATATAATGCTGTA
AAGCAAGATGAGAAAAGGGTAACAGGCATGACCGGAAGCTTAAAGAAGTGACCGGGATGCAAGTG
ACTCTTAGGATCTCATTAAGATGGTGAAAAGACATCACTTTAAGGAGGACCAAAGGAAAGGACC
TGAGG

Figure S1. Percent sporulation of *P. thermosuccinogenes* DSM5809 under phase contrast microscopy

a) Sporulation solution after 1 day in CP medium. b) Sporulation solution after 2 days in CP medium. c) Purified spores using 50% (w/v) solution of Histodenz, depicting > 95% purity. d) Purified spores by washing them 8-10 consecutive times with 20 mL of Milli-Q water, depicting < 90% purity. e) Spores appearance immediately before electroporation.

Figure S2. Schematic representation of the cassettes of RM systems from *P. thermosuccinogenes* to be integrated in *E. coli* HST04.

a) The cassette is composed by the T7 promoter, the S1 type I RM system, the T7 terminator and the kanamycin resistance gene flanked by the FRT sites. The 100 bp long homology arms directs the cassette to the bifunctional isocitrate dehydrogenase kinase/phosphatase gene (08395).

b) The cassette is composed by the XylS/Pm promoter, the S4 type I RM system, the T7 terminator and the kanamycin resistance gene flanked by the FRT sites. The upstream 100 bp region (5'US) targets the ArsR family transcriptional regulator (11130) and the downstream 100 bp homology region (3'DS) targets the arsenic transporter (11125).

c) The cassette is composed by the arabinose promoter, the S2 type I RM system, the T7 terminator and the chloramphenicol resistance gene flanked by the FRT sites. The 100 bp long homology arms directs the cassette to the lacZ gene. Cassettes a and b were equally built. Three individual PCR reactions (1a, 1b, 2b) were performed using primers with the appropriate overhangs. Followed by an overlap-extension PCR (2a) and an extra single reaction (3). The products yielded by the last two reactions were combined in the last overlap-extension PCR (4) to obtain the complete cassette. Cassette c followed the same strategy than cassettes a and b with some slight differences. The reaction 1c is added and the reaction 3 is not necessary to be performed.
Figure S2. Schematic representation of the cassettes of RM systems from *P. thermosuccinogenes* to be integrated in *E. coli* HST04. **a)** The cassette is composed by the T7 promoter, the S1 type I RM system, the T7 terminator and the kanamycin resistance gene flanked by the FRT sites. The 100 bp long homology arms directs the cassette to the bifunctional isocitrate dehydrogenase kinase/phosphatase gene (08395). **b)** The cassette is composed by the XylS/Pm promoter, the S4 type I RM system, the T7 terminator and the kanamycin resistance gene flanked by the FRT sites. The upstream 100 bp region (5’US) targets the ArsR family transcriptional regulator (11130) and the downstream 100 bp homology region (3’DS) targets the arsenic transporter (11125). **c)** The cassette is composed by the arabinose promoter, the S2 type I RM system, the T7 terminator and the chloramphenicol resistance gene flanked by the FRT sites. The 100 bp long homology arms directs the cassette to the *lacZ* gene. Cassettes a and b were equally built. Three individual PCR reactions (1a, 1b, 2b) were performed using primers with the appropriate overhangs. Followed by an overlap-extension PCR (2a) and an extra single reaction (3). The products yielded by the last two reactions were combined in the last overlap-extension PCR (4) to obtain the complete cassette. Cassette c followed the same strategy than cassettes a and b with some slight differences. The reaction 1c is added and the reaction 3 is not necessary to be performed.
Figure S3. Effects of CRISPRi-mediated repression on product formation of malic enzyme silencing transformants of both non-targeting and targeting plasmids. *P. thermosuccinogenes* DSM5809 cells were transformed with both non-targeting and targeting plasmids. The Tm<sup>R</sup> colonies of both the plasmids were transferred to 50 mL CP medium (5 g/L yeast extract) containing 6 µg/mL Tm and cultured to OD<sub>600</sub> > 1.0. 1 mL cells were sampled for 2 days for HPLC analysis. Data represent the mean values of three biological replicates and the standard deviation. The level of significance of the differences when results were compared was estimated by means of analysis of variance (ANOVA), with α=0.05.
Table S1. DNA methylation (dam/dcm) effects on electro- transformation efficiency when plasmids prepared from different E. coli hosts

<table>
<thead>
<tr>
<th>E. coli host</th>
<th>DNA methylation</th>
<th>CFU/ µg of plasmid pNW33n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>dcm&lt;sup&gt;-&lt;/sup&gt;/dam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>Acella</td>
<td>dcm&lt;sup&gt;-&lt;/sup&gt;/dam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>JM110</td>
<td>dcm&lt;sup&gt;-&lt;/sup&gt;/dam&lt;sup&gt;-&lt;/sup&gt;</td>
<td>3</td>
</tr>
<tr>
<td>HST04</td>
<td>dcm&lt;sup&gt;-&lt;/sup&gt;/dam&lt;sup&gt;-&lt;/sup&gt;/hsdRMS</td>
<td>5</td>
</tr>
</tbody>
</table>

Table S2. Bacterial strains used in the present study

<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> DH5α</td>
<td><em>fluA2</em> (argF-lacZ) U169 <em>phoA</em> glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi&lt;sup&gt;1&lt;/sup&gt; hsdR17</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td><em>Escherichia coli</em> Acella</td>
<td>F&lt;sup&gt;+&lt;/sup&gt;ompT hsdSB (rB&lt;sup&gt;+&lt;/sup&gt; mB&lt;sup&gt;-&lt;/sup&gt;) gal dcm (DE3) ΔendA ΔrecA</td>
<td>Edge Bio</td>
</tr>
<tr>
<td><em>Escherichia coli</em> JM110</td>
<td><em>psL</em> (Strr) thr leu thi&lt;sup&gt;1&lt;/sup&gt; lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) [F&lt;sup&gt;+&lt;/sup&gt; traD36 proAB lacIq ZAM15]</td>
<td>(Yanisch-Perron, et al., 1985)</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04</td>
<td>F&lt;sup&gt;+&lt;/sup&gt;, Δ(lac-proAB) [Φ80d lacZAM15]. rpsL(str), thi, Δ(mrr-hsdRM-S-mcrBC), ΔmcrA, dam, dcm</td>
<td>Clontech Laboratories (Takara)</td>
</tr>
<tr>
<td><em>Escherichia coli</em> CGSC 7632</td>
<td><em>E. coli</em> CGSC 7632 with the pKD4-kan plasmid (FRT-kanamycin-FRT cassette)</td>
<td>(Datsenko and Wanner, 2000)</td>
</tr>
<tr>
<td><em>Escherichia coli</em> CGSC 7632</td>
<td><em>E. coli</em> CGSC 7632 with the pKD3-Chl plasmid (FRT-chloramphenicol-FRT cassette)</td>
<td>(Datsenko and Wanner, 2000)</td>
</tr>
<tr>
<td><em>Pseudoclostridium thermosuccinogenes</em></td>
<td>DSM5809</td>
<td>DSMZ</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S1</td>
<td><em>E. coli</em> HST04 with the S1 type I RM system. The bifunctional isocitrate dehydrogenase kinase/phosphatase gene (08395) was disrupted</td>
<td>This study</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S2</td>
<td><em>E. coli</em> HST04 with the S2 type I RM system. The lacZ gene was disrupted</td>
<td>This study</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S4</td>
<td><em>E. coli</em> HST04 with the S4 type I RM system. Transcriptional regulator gene <em>arsR</em> (11130) and the arsenic transporter gene (11125) was disrupted</td>
<td>This study</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S1-S2-S4</td>
<td><em>E. coli</em> HST04 with the S1, S2 and S4 type I RM systems</td>
<td>This study</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S1-S2-S4</td>
<td><em>E. coli</em> HST04 with pThermoCas9i plasmid-16S/23S rRNA-non-targeting</td>
<td>This study</td>
</tr>
<tr>
<td><em>Escherichia coli</em> HST04-S1-S2-S4</td>
<td><em>E. coli</em> HST04 with pThermoCas9i plasmid-16S/23S rRNA-targeting</td>
<td>This study</td>
</tr>
</tbody>
</table>
Chapter 4

Table S3. Plasmids used in the present study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Primers</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNW3n</td>
<td>-</td>
<td><em>E. coli-Bacillus</em> shuttle vector pNW33n containing origin of replication functional in <em>H. thermocellum</em> and chloramphenicol/thiamphenicol antibiotic resistance (Olson and Lynd, 2012)</td>
</tr>
<tr>
<td>pNW3n_RM5809</td>
<td>BG12510 and BG12511 BG12512 and BG12513 BG12514 and BG12515 BG12516 and BG12517</td>
<td>Plasmid without the RM recognition sites of <em>P. thermosuccinogenes</em> DSM5809</td>
</tr>
<tr>
<td>pThermoCas9i-16S/23S rRNA-non-targeting</td>
<td>BG12508 and BG12518 BG12510 and BG12511 BG12512 and BG12513 BG12514 and BG12515 BG12516 and BG12517 BG12519 and BG12526 BG12527 and BG12509</td>
<td>Plasmid with the ThermodCas9 and the non-targeting sgRNA, which is under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313</td>
</tr>
<tr>
<td>pThermoCas9i-16S/23S rRNA-targeting</td>
<td>BG12508 and BG12518 BG12509 and BG12523 BG12510 and BG12511 BG12512 and BG12513 BG12514 and BG12515 BG12516 and BG12517 BG12519 and BG12520</td>
<td>Plasmid with the ThermodCas9 and the targeting sgRNA under the 16S/23S intergenic promoter of <em>H. thermocellum</em> DSM1313. The sgRNA with spacer targeting promoter of malic enzyme gene</td>
</tr>
</tbody>
</table>

Table S4. Spacers used in the present study

<table>
<thead>
<tr>
<th>Description</th>
<th>Spacer sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-targeting</td>
<td>5’GCTAGTCTCAAGGTCTCAGTAGCT3’</td>
</tr>
<tr>
<td>Spacer targeting the promoter of the NAD dependent malic enzyme</td>
<td>5’ATTAAGATTTTCCTATATAGATCTTA 3’</td>
</tr>
</tbody>
</table>
### Table S5. Point mutations introduced in ThermoCas9i / pNW33n plasmid

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Plasmid feature</th>
<th>Recognition site</th>
<th>Location inside ThermodCas9 plasmid (bp)</th>
<th>Mutation</th>
<th>Location of mutation inside the protein (bp)</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNW33N and pThermoCas9i</td>
<td>RepB protein</td>
<td>GATNNNNNCTC</td>
<td>2124-2133</td>
<td>GCT → GCC</td>
<td>568</td>
<td>Alanine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DGAGNNNNNATC</td>
<td></td>
</tr>
<tr>
<td>pNW33N and pThermoCas9i</td>
<td>pNW33N backbone</td>
<td>GATNNNNNCTC</td>
<td>3124-3133</td>
<td>A → C</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pThermoCas9i</td>
<td>ThermodCas9 protein</td>
<td>TCABNNNNNTARG</td>
<td>5079-5092</td>
<td>CCT → CCA</td>
<td>927</td>
<td>Proline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYTANNNNNVTGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GATNNNNNCTC</td>
<td>6016-6025</td>
<td>GGC → GGT</td>
<td>1860</td>
<td>Glycine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DGAGNNNNNATC</td>
<td>6016-6026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GATNNNNNCTC</td>
<td>6359-6368</td>
<td>CCG → CCA</td>
<td>2196</td>
<td>Proline</td>
</tr>
<tr>
<td>pNW33N and pThermoCas9i</td>
<td>Promoter of the cat</td>
<td>GATNNNNNCTC</td>
<td>1454-1563</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>protein</td>
<td>DGAGNNNNNATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S6. Codon usage of *P. thermosuccinogenes* DSM5809

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>Fraction</th>
<th>Amino acid</th>
<th>Codon</th>
<th>Fraction</th>
<th>Amino acid</th>
<th>Codon</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>0.12</td>
<td>Pro</td>
<td>CCG</td>
<td>0.22</td>
<td>Gly</td>
<td>GGG</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>GCA</td>
<td>0.33</td>
<td></td>
<td>CCA</td>
<td>0.29</td>
<td></td>
<td>GGA</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>GCT</td>
<td>0.32</td>
<td></td>
<td>CCT</td>
<td>0.30</td>
<td></td>
<td>GGT</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>0.24</td>
<td></td>
<td>CCC</td>
<td>0.19</td>
<td></td>
<td>GCC</td>
<td>0.25</td>
</tr>
<tr>
<td>Thr</td>
<td>ACG</td>
<td>0.13</td>
<td>Gln</td>
<td>CAG</td>
<td>0.46</td>
<td>His</td>
<td>CAT</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>0.33</td>
<td></td>
<td>CAA</td>
<td>0.54</td>
<td></td>
<td>CAC</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>ACT</td>
<td>0.25</td>
<td>Arg</td>
<td>AGG</td>
<td>0.23</td>
<td>Ile</td>
<td>ATA</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>ACC</td>
<td>0.29</td>
<td></td>
<td>AGA</td>
<td>0.23</td>
<td></td>
<td>ATT</td>
<td>0.35</td>
</tr>
<tr>
<td>Glu</td>
<td>GAG</td>
<td>0.34</td>
<td></td>
<td>CGG</td>
<td>0.19</td>
<td></td>
<td>ATC</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>GAA</td>
<td>0.66</td>
<td></td>
<td>CGA</td>
<td>0.12</td>
<td>Lys</td>
<td>AAG</td>
<td>0.37</td>
</tr>
<tr>
<td>Phe</td>
<td>TTT</td>
<td>0.62</td>
<td></td>
<td>CGT</td>
<td>0.11</td>
<td></td>
<td>AAA</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>TTC</td>
<td>0.38</td>
<td></td>
<td>CGC</td>
<td>0.12</td>
<td>Met</td>
<td>ATG</td>
<td>1.00</td>
</tr>
<tr>
<td>Tyr</td>
<td>TAT</td>
<td>0.70</td>
<td>Cys</td>
<td>TGT</td>
<td>0.46</td>
<td>Asn</td>
<td>AAT</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>TAC</td>
<td>0.30</td>
<td></td>
<td>TGC</td>
<td>0.54</td>
<td></td>
<td>AAC</td>
<td>0.37</td>
</tr>
<tr>
<td>Leu</td>
<td>TTG</td>
<td>0.19</td>
<td>Asp</td>
<td>GAT</td>
<td>0.67</td>
<td>Ser</td>
<td>AGT</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>TTA</td>
<td>0.20</td>
<td></td>
<td>GAC</td>
<td>0.33</td>
<td></td>
<td>AGC</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>CTG</td>
<td>0.17</td>
<td>Val</td>
<td>GTG</td>
<td>0.19</td>
<td></td>
<td>TCG</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>CTA</td>
<td>0.09</td>
<td></td>
<td>GTA</td>
<td>0.28</td>
<td></td>
<td>TCA</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>CTT</td>
<td>0.23</td>
<td></td>
<td>GTT</td>
<td>0.32</td>
<td></td>
<td>TCT</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>CTC</td>
<td>0.13</td>
<td></td>
<td>GTC</td>
<td>0.22</td>
<td></td>
<td>TCC</td>
<td>0.22</td>
</tr>
<tr>
<td>Trp</td>
<td>TGG</td>
<td>1.00</td>
<td>Stop</td>
<td>TGA</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TAG</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TAA</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S7. Primers used in the present study

Primers used for plasmid and cassette construction

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG10829</td>
<td>TCATGACCAAAAATCCCTTAACGTG</td>
</tr>
<tr>
<td>BG10830</td>
<td>CACGTTAAGGGATTGTTGTCATGA</td>
</tr>
<tr>
<td>BG11739</td>
<td>CCATTGTTGTTAATCCTGTGACGCTATGTCGCTGCTTGTGAG</td>
</tr>
<tr>
<td>BG11740</td>
<td>CCAGCTGAGTACTAGCGTACGGATCCCTACAGGTCCCTTTCTTTG</td>
</tr>
<tr>
<td>BG11746</td>
<td>AGGCCGTGACGGATTAACAAAAATGG</td>
</tr>
<tr>
<td>BG11747</td>
<td>GGAATCCATGACGCTATATCCAGCTGGGTCATATGCCCTCCTGAGATTATCG</td>
</tr>
<tr>
<td>BG12493</td>
<td>AACCTATAGTAGAAATCATTGATATCTCTCTCTAAAGGTAA</td>
</tr>
<tr>
<td>BG12494</td>
<td>TTAACCTTGAAGGAGATATACAAATGATTTTTCATCTATAGATTATTTCCT</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Primer code</th>
<th>Codon</th>
<th>Fraction</th>
<th>Amino acid</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG11747</td>
<td>GTG</td>
<td>0.22</td>
<td>Met</td>
<td>0.18</td>
</tr>
<tr>
<td>BG11746</td>
<td>CTA</td>
<td>0.32</td>
<td>Asn</td>
<td>0.18</td>
</tr>
<tr>
<td>BG10829</td>
<td>TAA</td>
<td>0.28</td>
<td>Val</td>
<td>0.63</td>
</tr>
<tr>
<td>BG11747</td>
<td>GTC</td>
<td>0.67</td>
<td>Arg</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>0.54</td>
<td>Cys</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG11746</td>
<td>CTA</td>
<td>0.54</td>
<td>Asn</td>
<td>0.32</td>
</tr>
<tr>
<td>BG10829</td>
<td>TAA</td>
<td>0.46</td>
<td>Val</td>
<td>0.46</td>
</tr>
<tr>
<td>BG11747</td>
<td>GTC</td>
<td>0.12</td>
<td>Arg</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>0.19</td>
<td>Cys</td>
<td>0.19</td>
</tr>
</tbody>
</table>

**Remark:**
- The table lists the codon usage of BG11747, BG11746, and BG10829, along with the amino acid sequences and their corresponding fractions.
- The codon usage is provided for each amino acid, showing the fraction of codons used for each amino acid.
Chapter 4

Primers used in the Quantitative Real Time PCR

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG12568</td>
<td>TACGCTGGCAGGGTCAGGAAATCAATTAACCTCATCGGAAAGTGGTGATCTGCA TATGAATATCCCTCCTTAGTTCTATTC</td>
</tr>
<tr>
<td>BG12572</td>
<td>TGTCATTCTTGATTACCACATCATTGTTTGTGATGAAACCCAG AAGGAAACCCTCATTCG</td>
</tr>
<tr>
<td>BG12573</td>
<td>CGCTTCGGGAGATTTCAGCCAGCATATCGTGTGTTCTCGGGAATATC TATGAATATCCTCCTTAGTTCTATTC</td>
</tr>
</tbody>
</table>

Primers used to sequence pThermodCas9i plasmids

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5’-3’)</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG11015</td>
<td>GATGAATTAGCCGCATCC</td>
<td></td>
</tr>
<tr>
<td>BG11016</td>
<td>ACCCGTCTTAAAGGTTTTAT</td>
<td></td>
</tr>
<tr>
<td>BG11017</td>
<td>ACCGAAAAGAAACGAAACTG</td>
<td></td>
</tr>
<tr>
<td>BG11018</td>
<td>CGAGTCACCGCCTTCTATC</td>
<td></td>
</tr>
<tr>
<td>BG11019</td>
<td>GATGAAAGGATCTTGCCAAAC</td>
<td>Particularly used to sequence the spacer</td>
</tr>
<tr>
<td>BG11021</td>
<td>GCTGTAATAATGGGTAGAAGGTA</td>
<td></td>
</tr>
<tr>
<td>BG11022</td>
<td>CCACCATAAGGAGATTAACCTTT</td>
<td></td>
</tr>
<tr>
<td>BG11023</td>
<td>CTTTTGTCTTGTCCATAAAC</td>
<td></td>
</tr>
<tr>
<td>BG11024</td>
<td>TGCCGTTTTTCCATAGGC</td>
<td></td>
</tr>
<tr>
<td>BG11025</td>
<td>CACGTTAAGGGATTTTGGTCATGA</td>
<td></td>
</tr>
</tbody>
</table>

Primers used in the RNA reverse transcriptase PCR

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG11636</td>
<td>ATCGGCTATTACGTCTATCGG</td>
</tr>
<tr>
<td>BG11637</td>
<td>AACACGAAGCTGCCAGAC</td>
</tr>
<tr>
<td>BG11642</td>
<td>CAGCTGGGTCCATAGTTCC</td>
</tr>
<tr>
<td>BG11643</td>
<td>CCTAAGAGTGGGGAAATGC</td>
</tr>
</tbody>
</table>

Primers used in the Quantitative Real Time PCR

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG10427</td>
<td>GCAATGGAGGAAACTCTGAC</td>
</tr>
<tr>
<td>BG10428</td>
<td>ACCCAGTAATTCCGGACAAC</td>
</tr>
<tr>
<td>BG18976</td>
<td>GGAAGGGCAAAATTGAAGTAT</td>
</tr>
<tr>
<td>BG18977</td>
<td>CGAGGTTCATCTTCTTGT</td>
</tr>
</tbody>
</table>
Table S8. Cassette construction

<table>
<thead>
<tr>
<th>Cassette</th>
<th>Primers</th>
<th>Template</th>
<th>Fragment</th>
<th>Overlap primers</th>
<th>Overlap PCR templates</th>
<th>Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>US0835 T7 S1 F</td>
<td>BG12567 and BG12493</td>
<td>BG11816 (GenArt)</td>
<td>T7 promoter</td>
<td>BG12569 and BG12499</td>
<td>T7 promoter and S1 gene</td>
<td>US0835 T7 S1</td>
</tr>
<tr>
<td>RT kan FRT DS 0835</td>
<td>BG12494 and BG12495</td>
<td>gDNA <em>P. thermosuccinogenes</em> DSM5809</td>
<td>S1 gene</td>
<td>BG12570 and BG12571</td>
<td>US0835 T7 S1 FRT kan FRT S1 FRT kan FR T_DS0835</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG12496 and BG12568</td>
<td>pKD4 plasmid</td>
<td>FRT-kan-FRT</td>
<td>BG12570 and BG12571</td>
<td>US0835 T7 S1 FRT kan FRT S1_FRT_kan_FR</td>
<td></td>
</tr>
<tr>
<td>USlacZ araB _S2 FRT Chl FR T_DSlacZ</td>
<td>BG12556 and BG12557</td>
<td>BG11818 (GenArt)</td>
<td>araB promoter</td>
<td>BG12563 and BG12499</td>
<td>AraC promoter and S2 gene</td>
<td>USlacZ araB _S2</td>
</tr>
<tr>
<td></td>
<td>BG12558 and BG12559</td>
<td>gDNA <em>P. thermosuccinogenes</em> DSM5809</td>
<td>S2 gene</td>
<td>BG12565 and BG12566</td>
<td>USlacZ AraC pm _S2 and FRT-chl-FRT S2 and FRT-chl-FRT DSlacZ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG12496 and BG12562</td>
<td>pKD3 plasmid</td>
<td>FRT-chl-FRT</td>
<td>BG12565 and BG12566</td>
<td>USlacZ araB _S2 FRT_chl_FRT_DSlacZ</td>
<td></td>
</tr>
<tr>
<td>US11130 Pm S4 _FRT kan FRT D S11125</td>
<td>BG12572 and BG12502</td>
<td>BG11820 (GenArt)</td>
<td>Pm promoter</td>
<td>BG12574 and BG12499</td>
<td>Pm promoter and S4 gene</td>
<td>US11130 Pm S4</td>
</tr>
<tr>
<td></td>
<td>BG12503 and BG12504</td>
<td>gDNA <em>P. thermosuccinogenes</em> DSM5809</td>
<td>S4 gene</td>
<td>BG12575 and BG12576</td>
<td>US11130 Pm S4 FRT kan FRT S11125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BG12496 and BG12573</td>
<td>pKD4 plasmid</td>
<td>FRT-kan-FRT</td>
<td>BG12575 and BG12576</td>
<td>US11130 Pm S4 FRT kan FRT S11125</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Heterologous expression of the *Pseudoclostridium thermosuccinogenes* xylose gene cluster in *Hungateiclostridium thermocellum*

Joyshree Ganguly1, Bram Vermeulen2, Richard van Kranenburg1, 2

1Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
2Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

Manuscript in preparation: Ganguly, J., Vermeulen, B., and van Kranenburg, R.
Chapter 5

Heterologous expression of the *Pseudoclostridium thermosuccinogenes* xylose gene cluster in *Hungateiclostridium thermocellum*

Joyshree Ganguly¹, Bram Vermeulen², Richard van Kranenburg¹, ²

¹Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
²Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

Manuscript in preparation:

Ganguly, J., Vermeulen, B., and van Kranenburg, R.
Chapter 5

ABSTRACT

Consolidated Bioprocessing (CBP) is the direct bioconversion of lignocellulosic feedstocks into desired products with a single microorganism or a microbial consortium. In this work, we focus on *Hungateiclostridium thermocellum* DSM1313 that has excellent cellulolytic activity, but is unable to grow on xylose. A GTP-dependent plasmid-based xylose gene cluster from *Pseudoclostridium thermosuccinogenes* was introduced into *H. thermocellum*. The expression of the xylose utilization genes allowed for growth on xylose. Transcriptional analysis showed upregulation expression of the xylulokinase (*xylB*) and transcriptional regulator (*xylR*) genes when grown on xylose compared to cellobiose. Unexpectedly, introduction of the plasmid-based xylose gene cluster into *H. thermocellum* also impacted the cellobiose fermentation profile. Ethanol and acetate production was improved by 24% and 19%, respectively, for the xylose-plasmid bearing cells compared to the empty vector. In conclusion, introducing the *P. thermosuccinogenes* xylose gene cluster is a useful step towards CBP with *H. thermocellum* to produce fuels and high value chemicals.
INTRODUCTION

A cost-effective strategy for biofuel production by a cellulolytic micro-organism combines cellulolytic enzymes production, biomass hydrolysis and fermentation of six-carbon (C6) substrates to preferred products in a single step process. This process is referred to as consolidated bioprocessing (CBP) (Lynd, 2016, Lynd, et al., 2017). The thermophile Hungateiclostridium thermocellum has been studied extensively for CBP, because of its native capability to ferment cellulose (Lynd, et al., 2005, Olson, et al., 2012). So far, H. thermocellum has been engineered mainly for ethanol and iso-butanol production via CBP (Lin, et al., 2015, Tian, et al., 2016, Hon, et al., 2017, Hon, et al., 2018). Although H. thermocellum has the advantage of its cellulolytic capacity and thermophilic nature, optimal use of biomass feedstock through CBP also requires consumption of the hemicellulose fraction, and H. thermocellum DSM1313 cannot grow on hemicellulose or its derived C5 sugars (Demain, et al., 2005). In few studies, co-culture of the hemicellulolytic Thermoanaerobacter and Ruminiclostridium stercorarium in combination with the cellulolytic H. thermocellum was used to ferment lignocellulose into ethanol in an CBP approach (Lynd, et al., 2002, Argyros, et al., 2011, Olson, et al., 2015, Froese, et al., 2019). Yet, a single micro-organism CBP arrangement has benefits over a microbial consortium with different optima for growth, pH, substrate availability, redox potential and product inhibition. Recently, the attempt of single organism CBP production was achieved with H. thermocellum DSM1313 by targeted integration of xylAB genes from Thermoanaerobacter ethanolicus into the H. thermocellum genome. The resulting strain showed increased H2 and ethanol production for co-consumption of xylose and cellulose in comparison to cellulose only (Argyros, et al., 2014, Xiong, et al., 2018).

In this study, we introduced a plasmid into H. thermocellum containing a GTP-dependent xylose gene cluster derived from the close relative Pseudoclostridium thermosuccinogenes. The latter is a thermophilic spore-forming anaerobe that converts sugars to succinate. The enzymes involved in xylose metabolism in P. thermosuccinogenes are xylose isomerase and GTP-dependent xylulokinase which are in a gene cluster together with a transcriptional regulator (Koendjbiharie, et al., 2018). Expression of the xylose metabolizing genes in H. thermocellum was quantified using qRT-PCR. In addition, we characterized the product profiles of the modified strain on C5 and C6 sugars in anaerobic bottles as well as fermenters. These results will demonstrate the usefulness of introduction of the P. thermosuccinogenes xylose gene
cluster into \textit{H. thermocellum} in the development towards CBP for production of biofuels and biochemicals.

\section*{RESULTS}

\subsection*{Creation of a plasmid-based xylose genes system and expression in \textit{H. thermocellum}}

\textit{H. thermocellum} can utilize cellobiose and cellulose, but is unable to metabolize xylose, as it lacks xylose isomerase and xylulokinase genes. Plasmid pNW33n\_Xyl was constructed with the GTP-dependent xylose gene cluster consisting of \textit{xylA} (CDQ83\_14305), \textit{xylR} (CDQ83\_14310) and \textit{xylB} (CDQ83\_14315) genes with two predicted promoters and a terminator from \textit{P. thermosuccinogenes} (Figure 5.1a). The plasmids pNW33n\_Xyl and pNW33n (control) were transformed to \textit{H. thermocellum}. Transformants were selected on CTFUD plates containing 3.5 g/L cellobiose and selecting for thiamphenicol resistance (Tm$^R$). These transformants were re-streaked to CTFUD-Tm plates with either cellobiose or xylose (3.5 g/L). The clones that grew on both plates were verified using colony PCR and further grown in CTFUD-Tm liquid cultures containing cellobiose or xylose (Figure S1, Supplementary data). The \textit{H. thermocellum} (pNW33n\_Xyl) and \textit{H. thermocellum} (pNW33n) transformants grew on cellobiose agar plates and in liquid cultures. The control \textit{H. thermocellum} (pNW33n) did not grow on xylose CTFUD plates and in liquid cultures. A PCR-amplified fragment containing the gene cluster was sent for standard sequencing from Macrogen (MACROGEN Inc. DNA Sequencing Service; Amsterdam, The Netherlands) to confirm the integrity of the plasmid in several colonies. In addition, we tested the expression of xylose utilization genes in \textit{H. thermocellum} (pNW33n\_Xyl) transformants when grown on xylose or cellobiose. The qRT-PCR analysis showed upregulation of \textit{xylR} and \textit{xylB} genes between 5-15\% in three independent transformants in xylose cultures compared to cellobiose cultures (Figure 5.1b).
The GTP-dependent xylose gene cluster consisting of

**Chapter 5**

**RESULTS**

Creation of a plasmid

**Figure 5.1.** a) The xylose gene cluster from *P. thermosuccinogenes* cloned in a pNW33n backbone; *xylA* (CDQ83_14305) - xylose isomerase, *xylR* (CDQ83_14310) - xylose regulatory protein and *xylB* (CDQ83_14315) - xylulokinase. Predicted promoters and terminators indicated as arrows and hairpin loop, respectively. b) Relative *xylA*, *xylR* and *xylB* gene expression of *H. thermocellum* (pNW33n_Xyl) evaluated by qRT-PCR in xylose compared to cellobiose grown conditions at an OD_{600} ~ 0.6, respectively. The strains were transferred to 50 mL CP (5 g/L yeast extract) medium containing 6 µg/mL Tm and 3.5 g/L cellobiose or xylose. 5 mL cells were sampled for qRT-PCR analysis with biological and technical triplicates. The expression levels were normalized to the control. Data represent biological triplicates and the standard deviation.

**Product profile for xylose transformants in anaerobic bottles**

*H. thermocellum* transformants harbouring plasmid pNW33n_Xyl or pNW33n (control) were grown on CTFUD medium in anaerobic bottles with xylose or cellobiose as carbon sources. *H. thermocellum* (pNW33n_Xyl), displayed an approximately 24 hr lag period for growth on xylose compared to growth on cellobiose. Both strains were similar in growth and cellobiose consumption, signifying that *H. thermocellum* (pNW33n_Xyl) did not compromise its growth capacity on cellobiose compared to the control *H. thermocellum* (pNW33n). However, the results indicated differences in product formation between the cellobiose-grown *H. thermocellum* (pNW33n_Xyl) and the control strain for lactate and formate yields (Figure 5.2). Furthermore, the major products ethanol and acetate were improved by 24% and 19%,

---

**Figure 5.2**
respectively, in xylose-grown *H. thermocellum* (pNW33n_Xyl) transformants compared to empty vector. This was again at the cost of lactate and formate production.

![Fermentation products from anaerobic bottle incubations of *H. thermocellum* (pNW33n_Xyl) grown on cellobiose (blue), xylose (orange) and *H. thermocellum* (pNW33n) grown on cellobiose (grey). Organic acids yields are expressed as mol of product produced per mol of substrate consumed. Colonies of both strains were transferred to 50 mL CP medium (1 g/L yeast extract) containing 6 µg/mL Tm and cultured to late stationary phase for 2 days.1 mL of broth was sampled for HPLC analysis. Data represent the mean values of biological triplicates and the standard deviation are indicated.](image)

**Figure 5.2.** Fermentation products from anaerobic bottle incubations of *H. thermocellum* (pNW33n_Xyl) grown on cellobiose (blue), xylose (orange) and *H. thermocellum* (pNW33n) grown on cellobiose (grey). Organic acids yields are expressed as mol of product produced per mol of substrate consumed. Colonies of both strains were transferred to 50 mL CP medium (1 g/L yeast extract) containing 6 µg/mL Tm and cultured to late stationary phase for 2 days.1 mL of broth was sampled for HPLC analysis. Data represent the mean values of biological triplicates and the standard deviation are indicated.

### The effect of xylose genes on product formation in fermenter studies

Anaerobic bottle experiments indicated an increase in ethanol and acetate yields from xylose for *H. thermocellum* (pNW33n_Xyl). This was further investigated in fermenter runs that are more representative for industrial conditions. The fermentation profiles (Figure 5.3) for *H. thermocellum* (pNW33n_Xyl) are in line with those observed in the anaerobic bottle cultures with acetate and ethanol as major products and formate and lactate as minor ones. Acetate was 30% higher for the xylose-grown strain compared to cellobiose-grown cultures. For ethanol the values probably underestimate the true values due to potential evaporation issues at elevated temperature, but still, compared to cellobiose cultures an approximate 8.5% increased production was observed. The minor fermentation products lactate and formate were decreased in xylose fermentations, which was also observed in the bottle cultures.
Chapter 5

Heterologous expression of the xylose gene cluster in *H. thermocellum*

**DISCUSSION**

Biofuels and green chemicals production in *H. thermocellum* require modulation of metabolic pathways, primarily based on genetic engineering or transcriptional regulation (Mohr, et al., 2013, Biswas, et al., 2014, Papanek, et al., 2015, Ganguly, et al., 2019). Effective bioprocesses may be obtained by using microbial consortia. Defined cellulosytic co-cultures of *H. thermocellum* for the breakdown of lignocellulosic biomass have been reported (Ng, et al., 1981, Argyros, et al., 2011). *H. thermocellum*'s unique capacity to hydrolyze cellulose without the utilization of C5 sugars (Zhang and Lynd, 2005, Lu, et al., 2006, Blumer-Schluette, et al., 2008), opens ways to establish co-cultures with pentose-consuming organisms for effective CBP. This has been demonstrated by co-culturing of a metabolically engineered *H. thermocellum* and *T. saccharolyticum* providing high ethanol titers of 38 g/L (Argyros, et al., 2011).

Figure 5.3. Effect on fermentation product yields [mol of product produced/mol of substrate consumed] of *H. thermocellum* (pNW33n_Xyl) grown on cellobiose (blue) or xylose (red), respectively. A single clone was further tested in fermenter in 500 mL CP medium (5 g/L yeast extract) containing 6 µg/mL Tm. Xylose or cellobiose (5 g/L) as carbon sources and cultured to late stationary phase for 2 days, and 1 mL cells were sampled for HPLC analysis. Data represent technical triplicates and the standard deviation.
Chapter 5

Another strategy is to introduce the capacity to co-ferment both cellulose and hemicellulose from plant biomass in *H. thermocellum* for an enhanced CBP strain (Xiong, et al., 2018). To this end, we developed a *H. thermocellum* strain that can consume xylose. After introduction of the xylose genes from *P. thermosuccinogenes*, we observed more than 20-hour lag phase in growth compared to cellobiose grown strains, after a similar growth rate as observed in previous studies (Argyros, et al., 2014, Xiong, et al., 2018). Although we have no clear explanation for this phenomenon, xylose transport and expression of the heterologous xylose genes in *H. thermocellum* could be contributing to the delay. The xylose transport is likely to be facilitated by ATP-binding cassette (ABC) transporters in *H. thermocellum* (Nataf, et al., 2009, Verbeke, et al., 2017). Preculture on xylose plates does not eliminate the lag phase of xylose fermentations, eliminating the xylose induced expression of the *xylB* gene as potential cause. Further studies are required to understand, and potentially prevent, the lag-phase in xylose fermentation. These could include metabolomics studies linking the C5 and C6 sugars pathways.

The introduction of xylose genes impacted the product profile when grown on cellobiose. This raises the question on how xylose pathway may impact the C6 sugars metabolism and consequently, affect ethanol and acetate production. Recently, a study revealed a sedoheptulose 1,7-bisphosphate pathway that links pentose metabolism to the rest of glycolysis for both *P. thermosuccinogenes* and *H. thermocellum*. In this reversible pathway, pyrophosphate-dependent phosphofructokinase (PPI-PFK) converts sedoheptulose-7-phosphate to sedoheptulose-1,7-bisphosphate, then fructose bisphosphate aldolase cleaves the latter into dihydroxyacetone phosphate and erythrose 4-phosphate. Lastly, sedoheptulose-7-phosphate can be converted to ribose-5-phosphate and xylulose-5-phosphate (Koendjibiharie, et al., 2019). However, further research involving silencing of genes in this pathway and stable-isotope labeling plus kinetic flux analysis is required to measure intermediates and predict the effect of xylose pathway on C6 sugars metabolism.

Studies on xylose metabolism in *P. thermosuccinogenes* showed the presence of a GTP-dependent xylulokinase in xylose conversion pathway. Enzyme assays demonstrated that xylulokinase has preference for GTP over ATP (Koendjibiharie, et al., 2018). The *T. saccharolyticum* xylulokinase that Xiong *et al.* (Xiong, et al., 2018) introduced has not been characterized for its ATP or GTP co-factor usage. Earlier studies showed preference of GTP rather than ATP during glycolysis in *H. thermocellum* (Zhou, et al., 2013). Though no distinct explanation has been presented for the dependence of the central metabolism on GTP.
Conceivably, GTP permits the presence of an extra energy charge other than ATP, to accomplish different activities in the metabolism (Bhuyan and Simpson, 1962, Martín del Campo, et al., 2013, Koendjibiarie, et al., 2018). This would favor the choice for a GTP-dependent xylulokinase, like the one we used, for *H. thermocellum*.

The introduction of the *P. thermosuccinogenes* genes into *H. thermocellum* has impacted the product profiles of this strain as studied both in anaerobic bottles and controlled batch fermentations. For anaerobic bottles, unpredictably, the xylose gene cluster has impacted the cellobiose fermentation profile with reduced lactate and formate and slight increase in ethanol production. Furthermore, when the cellobiose cultures were compared to the xylose cultures we observed enhanced ethanol and acetate production. The anaerobic bottle product formation correlated with the fermentation data for the xylose cultures. In comparison with the *H. thermocellum* KJC335 strain with *xylAB* from the thermophile *T. ethanolicus* integrated into its genome, our strain exhibited equivalent xylose consumption profiles, acetate and ethanol as major products with reduced lactate and formate concentrations. Xiong *et al.*, tested co-consumption of xylose with avicel and concluded that the engineered strain has doubled H2 and ethanol production with co-consumption rather than only avicel consumption (Xiong, *et al.*, 2018). Hence, co-consumption of xylose with C6 sugars would be a relevant test for our strain to confirm if this also results in an increased flux towards ethanol.

In summary, we created a *H. thermocellum* strain that can utilize xylose which allows for utilization of both C5 and C6 sugars derived from plant biomass feedstocks. To achieve this, we have introduced a plasmid with the xylose gene cluster from the close relative *P. thermosuccinogenes*. The resulting strain can grow both in cellobiose or xylose. Introduction of the xylose genes improved *H. thermocellum* ethanol yield, compared to the empty-plasmid control strain and that the effect was more pronounced for the xylose-grown cells, but also present for cellobiose-grown cells. Therefore, this is the starting point to further develop this strain for co-fermentation of five-carbon and six-carbon substrates derived from lignocellulosic biomass to improve ethanol production.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains and growth conditions**

All bacterial strains used in this study are listed in supplementary Table S1. High efficiency NEB (New England Biolabs, CFU ~10^9/µg pUC19) chemically competent *E. coli* DH5α was
used for cloning purposes. *E. coli* Acella (Acella™ Chemically Competent cells, Edge Bio) was used for plasmid propagation and isolation for *H. thermocellum* DSM1313 transformation purposes to ensure the generation of Dam-methylated plasmid DNA. *E. coli* strains were cultured in Lysogeny-Broth (LB) medium (1% tryptone, 1% NaCl, 0.5% yeast extract) at 37 ºC and 200 rpm. When appropriate the medium was supplemented with 20 μg/mL chloramphenicol.

*H. thermocellum* DSM1313 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). *H. thermocellum* wild type and transformants were grown anaerobically in CTFUD broth and agar media at 50-60 ºC range adapted from Olson and Lynd (Olson and Lynd, 2012). *P. thermosuccinogenes* DSM5809 and the *H. thermocellum* transformants were grown in CP medium adapted from (Plugge, 2005). The final volume of 50 mL medium was dispensed in serum bottles under 80:20 N2/CO2 atmosphere with ~70 kPa overpressure and then autoclaved. A stock solution comprising sodium bicarbonate and l-cysteine was autoclaved separately and added later. Likewise, vitamin stock solution was added to calcium chloride dihydrate solution after it was autoclaved. The transformants were grown with 5.0 g/L (RNA isolation) or 1.0 g/L (HPLC) yeast extract (BD Bacto), respectively. Cellobiose or xylose carbon sources were also autoclaved separately and added later to a final concentration of 3.5 g/L. When appropriate, 6 μg/mL thiamphenicol was added.

**Xylose-utilization plasmid construction**

The plasmids and primers used in this study are listed in Table S1 and S2, Supplementary data, respectively. Plasmids were built using fragments with the appropriate overhangs via NEBuilder™ HiFi DNA assembly cloning kit according to the manufacturer’s protocol. All PCR reactions for amplification of fragments were performed with the NEB Q5® High-Fidelity DNA polymerase (M0491). PCR fragments were subjected to 1% w/v agarose gel electrophoresis, and isolated using Zymoclean™ Gel DNA recovery kit. Plasmid pNW33n was the empty vector used as a control. The genes of interest involved in xylose conversion are *xylA*, *xylR* and *xylB*. Plasmid pNW33n_Xyl was designed by cloning genes of native xylose gene cluster from *P. thermosuccinogenes* DSM5809 to aid xylose consumption for *H. thermocellum* DSM1313. Genomic DNA (gDNA) of *P. thermosuccinogenes* was used as a template for PCR to amplify the xylose gene cluster. The gDNA was isolated using the MasterPure™ Gram positive DNA Purification Kit (Lucigen), according to the manufacturer’s instructions. All plasmids were introduced by heat-shock into chemically competent *E. coli* DH5α cells (Sambrook, et al., 1989). Plasmids were isolated from selected single transformants by using the GeneJET
plasmid miniprep kit® (Thermo Scientific). Plasmid sequences were confirmed by standard sequencing from Macrogen (MACROGEN Inc. DNA Sequencing Service; Amsterdam, The Netherlands). After sequence confirmation, plasmids were electroporated into *E. coli* Acella (Sambrook, et al., 1989) before being introduced in *H. thermocellum* by electroporation as described by Olson and Lynd (Olson and Lynd, 2012).

**Gene expression of xylose metabolism genes using qRT-PCR**

RNA isolation of *H. thermocellum* transformants, harbouring the pNW33n or pNW33n_Xyl plasmids, was performed using 5 mL of overnight cultures in the mid-log phase (OD$_{600}$ nm reached 0.6). The procedure for RNA extraction with cDNA synthesis was followed as described in Ganguly et al. (Ganguly, et al., 2019). The cDNA was used for qRT-PCR experiments. To assess the functionality of the three xylose metabolizing genes grown in cellobiose and xylose, a quantitative real time PCR (qRT-PCR) was performed using the cDNA synthesized from the RNA of *H. thermocellum* DSM1313 transformants harbouring plasmids pNW33n (control) and pNW33n_Xyl. As an additional control, the cDNA was also synthesized from the RNA of the wild type *H. thermocellum* DSM1313 strain. It was used to measure the relative expression level of the *xylA*, *xylB* and *xylR* genes of the transformant *H. thermocellum* grown in xylose, compared to grown in cellobiose with the xylose gene cluster and pNW33n (control). The qRT-PCR was performed by using the iQ™ SYBR® Green Supermix (Bio-Rad) in a Bio-Rad C1000 Thermal Cycler. The final volume of the reaction was set to 20 µL with the amount of cDNA relative to 50 ng of RNA as a template. The house-keeping gene used to measure the relative expression was the recombinase A (*recA*) of *H. thermocellum* DSM1313. The primers used to amplify the *xylA*, *xylR* and *xylB* genes of *H. thermocellum* DSM1313 were BG18803, BG18804, BG18805, BG18806, BG18807 and BG18808 (Table S3, Supplementary data).

**Fermentation and product profile analysis using HPLC**

*H. thermocellum* strains were cultivated in CP medium adapted from (Plugge, 2005). The strains harbouring pNW33n_Xyl were grown in DASGIP® BioBlock reactors (Eppendorf). Culture conditions were kept anaerobic with 500 mL medium, that contained cellobiose or xylose as carbon sources to a final concentration of 5 g/L with 6 µg/mL thiamphenicol for plasmid selection. Bioreactors with CP medium was autoclaved, xylose or cellobiose (5 g/L), sodium bicarbonate plus L-cysteine (4 g/L) was autoclaved separately and added later. Similarly, vitamin stock solution was added to calcium chloride dihydrate solution (0.11 g/L)
after it was autoclaved. After that inoculation with 1 mL overnight culture from an anaerobic bottle was done. The reactors were sparged with 80:20 N2/CO2 at a rate of 1 L/h. The cultures were grown at 50 °C, stirred at 200 rpm and pH-7.0 by titration with 3M KOH. Samples were taken at different time points for 3 days for analysis with high-pressure liquid chromatography (HPLC) using the ICS-5000 HPLC system (Thermo Scientific). To this end, the system was equipped with an Aminex HPX 87H column (Bio-Rad Laboratories) and peak detection was via UV1000 detector operating at 210 nm and a RI-150 40 °C refractive index detector. The mobile phase consisted of 0.16 N H2SO4 and the column was operated at 0.8 mL/minute. All samples were diluted 4:1 with 10 mM DMSO in 0.01 N H2SO4. Sugars (cellobiose, xylose glucose, ethanol and glycerol) and organic acids (acetic acid, lactic acid, succinic acid and formic acid) were used as standards with a concentration range between 1.25-25 mM.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Funding: This work was supported by Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068].
SUPPLEMENTARY DATA

Figure S1. a and b) *H. thermocellum* transformants (green) and pNW33N_WT control (magenta) on CTFUD agar plates and liquid cultures with 3.5 g/L and 6 μg/mL thiamphenicol grown on cellobiose (a) and xylose (b), respectively.
Chapter 5

Table S1 Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strains or Plasmid</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Escherichia coli DH5α</strong></td>
<td><em>fluA2</em> (argF-lacZ) U169 phoA glnV44 80 (lacZ)M15</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td></td>
<td><em>gyrA96 recA1 relA1 endA1 thi-1 hsdR17</em></td>
<td></td>
</tr>
<tr>
<td><strong>Escherichia coli Acella</strong></td>
<td><em>F comPThsdSB (r6 mB) gal dcm (DE3) ΔendA ΔrecA</em></td>
<td>Edge Bio</td>
</tr>
<tr>
<td><strong>Hungateiclostridium thermocellum</strong></td>
<td></td>
<td>DSM1313</td>
</tr>
<tr>
<td><strong>Pseudoclostridium thermosuccinogenes</strong></td>
<td></td>
<td>DSM5809</td>
</tr>
<tr>
<td><strong>pNW 3n</strong></td>
<td><em>E. coli</em>-Hungateiclostridium shuttle vector, 4.2 kb, Tm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Bacillus Genetic Stock Center</td>
</tr>
<tr>
<td><strong>pNW 3n_Xyl</strong></td>
<td>8.1 kb, Tm&lt;sup&gt;R&lt;/sup&gt;, pNW33n containing the <em>xylA</em>, <em>xylR</em> and <em>xylB</em> gene clusters of <em>P. thermosuccinogenes</em></td>
<td>This study</td>
</tr>
</tbody>
</table>

Table S2 Primers used in the present study

The primers used to amplify the *recA* gene were BG10431 and BG10432.
The primers used to amplify the *xylA*-xylR-*xylB* gene cluster for cloning into pNW33n were BG17402 and BG17403.
The primers used to amplify the pNW33n backbone for cloning of the *xylA*-xylR-*xylB* gene cluster were BG17400 and BG17401.
The primers used to amplify the *xylA* gene were BG18803 and BG18804.
The primers used to amplify the *xylR* gene were BG18805 and BG18806.
The primers used to amplify the *xylB* gene were BG18807 and BG18808.

<table>
<thead>
<tr>
<th>Primer code</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG10431</td>
<td>GAGAAGCAGTTTGCGGAAGG</td>
</tr>
<tr>
<td>BG10432</td>
<td>CGTCGTCTTACCCGATGATT</td>
</tr>
<tr>
<td>BG17400</td>
<td>ATATCGGGATCTTTTACTTTCCGTTCTTATACATCAAAAG</td>
</tr>
<tr>
<td></td>
<td>GTTATTCATTCT</td>
</tr>
<tr>
<td>BG17401</td>
<td>GGCACAGCAAATCCGATTC</td>
</tr>
<tr>
<td>BG17402</td>
<td>TGTCAAAAGAGTGAATCGGATGCTGCTGCCTCTGTGTGAAAT</td>
</tr>
<tr>
<td></td>
<td>TGTTATCC</td>
</tr>
<tr>
<td>BG17403</td>
<td>AGAACGGGAAAAGTAAAACG</td>
</tr>
<tr>
<td>BG18803</td>
<td>GCAGCAGTTGCGGTATTG</td>
</tr>
<tr>
<td>BG18804</td>
<td>TCTACGACATAGGAGTGGAAT</td>
</tr>
<tr>
<td>BG18805</td>
<td>GGACTACAAGCCCAATATTCG</td>
</tr>
<tr>
<td>BG18806</td>
<td>CCCTATAATCATGAGCTCATTTTC</td>
</tr>
<tr>
<td>BG18807</td>
<td>TGCGGAATCTGTTCGAT</td>
</tr>
<tr>
<td>BG18808</td>
<td>ATCACAGCGCGGATCATA</td>
</tr>
</tbody>
</table>
Chapter 6

Characterization of sporulation dynamics of *Pseudoclostridium thermosuccinogenes* using flow cytometry

Joyshree Ganguly¹, Marcel Tempelaars², Tjakko Abee², Richard van Kranenburg¹,³

¹Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
²Laboratory of Food Microbiology, Wageningen University and Research, 6708 WG Wageningen, The Netherlands
³Laboratory of Microbiology, Wageningen University and Research, 6708 WE Wageningen, The Netherlands

Manuscript under revision:

Ganguly, J., Tempelaars, M., Abee, T., and van Kranenburg, R.
Chapter 6

ABSTRACT

Single-cell analysis of microbial population heterogeneity is a fast growing research area in microbiology due to its potential to identify and quantify the impact of subpopulations on microbial performance in, for example, industrial biotechnology, environmental biology, and pathogenesis. Although several tools have been developed, determination of population heterogeneity in anaerobic bacteria, especially spore-forming clostridia species has been amply studied. In this study, we applied single cell analysis techniques such as flow cytometry (FCM) and fluorescence-assisted cell sorting (FACS) on the spore-forming succinate producer *Pseudoclostridium thermosuccinogenes*. By combining FCM and FACS with fluorescent staining, we differentiated and enriched all sporulation-related morphologies of *P. thermosuccinogenes*. To evaluate the presence of metabolically active vegetative cells, a blend of the dyes propidium iodide (PI) and carboxy fluorescein diacetate (cFDA) tested best. Side scatter (SSC-H) in combination with metabolic indicator cFDA dye provided the best separation of sporulation populations. Based on this protocol, we successfully determined culture heterogeneity of *P. thermosuccinogenes* by discriminating between mature spores, forespores, dark and bright phase endospores, and vegetative cells populations. Henceforth, this methodology can be applied to further study sporulation dynamics and its impact on fermentation performance and product formation by *P. thermosuccinogenes*. 
INTRODUCTION

Microbial population heterogeneity is an outcome from genetic diversity of the population and/or phenotypic variance among isogenic cells. Population diversity results from the response of cells to their environment, microbial interactions, or processes such as cell division or gene regulation. Recent developments showed the existence and relevance of cellular heterogeneity leading to research on single microbial cell analysis (Davey and Kell, 1996, Veal, et al., 2000, Davey and Winson, 2003, Müller and Nebe-von-Caron, 2010, Ambriz-Avina, et al., 2014). Therefore, single-cell analysis techniques focused on subpopulations will advance the understanding of the bacterial population dynamics.

The class clostridia are obligate anaerobe, gram-positive, sporulating Firmicutes that are phylogenetically diverse. These species are important to human or animal health as well as related to the industrial production of bio-based products (Paredes, et al., 2005). Pathogenic species such as Clostridium difficile, Clostridium tetani, Clostridium perfringens, and Clostridium botulinum produces bacterial toxins (Raffestin, et al., 2005). The cellulolytic and solventogenic species such as Clostridium saccharobutylicum, Clostridium cellulolyticum, Hungateiclostridium thermocellum, Clostridium acetobutylicum, Clostridium beijerinckii and Clostridium pasteurianum have the potential for sustainable biofuel or biochemical production (Raffestin, et al., 2005, Ni, et al., 2013, Gaida, et al., 2016, Jiang, et al., 2016, Cheng, et al., 2019). Pseudoclostridium thermosuccinogenes is the only known thermophilic spore-forming bacteria that ferments sugar to succinate. Anaerobic thermophilic production of succinic acid may become an attractive green alternative to the existing chemical processes (Sridhar, et al., 2000, Koendjibiharie, et al., 2018). Furthermore, other nontoxigenic, proteolytic species such as Clostridium sporogenes and Clostridium novyi, serve as chemotherapeutic vehicles for intravenous delivery of spores, to treat solid tumors (Minton, 2003).

Sporulation is common to all clostridia, but knowledge on its mode of action and regulation at molecular and cellular level is still limited. The sporulation process of clostridia is studied extensively for several industrial applications using their enzymes and metabolites impacting product formation and their use in the medical field for drug delivery (M. Num and Useh, 2014, Green, et al., 2017). Sporulation can be linked to relevant characteristics such as enterotoxin production in C. perfringens and C. botulinum (Nakamura, et al., 1978, Freedman, et al., 2016) or solvent production in C. acetobutylicum (Jang, et al., 2013) and C. beijerinckii (Zhang, et al.,...
2018), respectively. *C. sporogenes* and *C. novyi* spores are administered as therapeutics (Minton, 2003).

Sporulation in clostridial species can be studied with single-cell analysis techniques including Fourier transform infrared spectroscopy, Raman spectroscopy coupled with microscopy, and electro-optical measurement techniques (Schuster, et al., 2000, Schuster, et al., 2001, Junne, et al., 2008). However, these techniques need custom-built equipment, lack the option for high-throughput analysis and are unable to quantify bulk population heterogeneity with different phenotypes. Flow cytometry (FCM), potentially combined with fluorescence activated cell sorting (FACS), is ideal to characterize microbes on single cell level. It is a quick method to obtain multi-parametric data of individual cells from thousands of them within a sample (Cronin and Wilkinson, 2010). Light scattering (LS) measurements such as Forward-scatter (FSC) offer data on the cell size, whereas side-scatter (SSC) is linked to diffraction of light based on cellular content and granularity (Ambriz-Avina, et al., 2014, Buysschaert, et al., 2016). Moreover, fluorescent dyes can provide insights into other cell properties such as viability or physiological state of cells (Hoefel, et al., 2003, Stocks, 2004, Berney, et al., 2007). For example, *Bacillus licheniformis* spores were differentiated from viable cells based on size and fluorescence using the SYBR green I dye (Zheng, et al., 2017). Several other applications of FCM include discrimination of microbial species in co-cultures, viability assays of bacteria and yeasts, studies on sporulation cycles, metabolic activity and membrane integrity (Davey and Kell, 1996, Veal, et al., 2000, Hornbæk, et al., 2002, Davey and Winson, 2003, Laflamme, et al., 2005, Cronin and Wilkinson, 2010).

Here we report the development of FCM methods for investigation of sporulation aspects of anaerobic thermophilic clostridia exemplified by *P. thermosuccinogenes*. Knowledge on the initiation and dynamics of the sporulation process may provide clues to further optimize succinate production. LS parameter (SSC-H) and cFDA/PI dyes were evaluated to discriminate cellular morphologies and viability. Microscopy and heat inactivation verified the identities of the sorted populations of mature spores, dark and bright phase endospores, forespores and vegetative cells of *P. thermosuccinogenes*. This paves the way for possible use of the described approach in monitoring industrial fermentations and to link sporulation dynamics to metabolite production.
RESULTS AND DISCUSSION

Microscopic characterization of *P. thermosuccinogenes*

*P. thermosuccinogenes* is a thermophilic, anaerobic and spore-forming bacterium. The genome of *P. thermosuccinogenes* DSM5809 was sequenced by Illumina shotgun sequencing (Koendjibiharie, et al., 2018). Bioinformatic analysis revealed the major sporulation and germination genes to show homology to that of other *Clostridium* and *Bacillus* species. To further characterize *P. thermosuccinogenes*, we performed phase contrast microscopy and Transmission Electron Microscopy (TEM) of cultures in the late exponential phase (20h).

Microscopic analysis showed that *P. thermosuccinogenes* DSM5809 exhibited diverse populations as shown in Figure 6.1a, including 1) vegetative cells, 2) bright phase endospores, 3) dark phase endospores and 4) mature free spores. The morphology of the vegetative cells and mature spores were evaluated with TEM. Thin sections of fixed spores showed a highly structured, multi-layered surface as also indicated by Drent *et al.* (J. Drent, et al., 1991). The rigid unique structural layers of the spores provide resistance to these organisms against adverse conditions (Setlow, 2007). The innermost spore part, called the spore core (Figure 6.1b) comprises the spore DNA, RNA and other enzymes (Setlow, 2007). The core is enclosed by an inner membrane protein with similar phospholipid composition like vegetative cells (Cowan, et al., 2004). This membrane safeguards the core DNA from damaging chemicals (Setlow, 2006). Adjacent to the spore’s inner membrane in *P. thermosuccinogenes* is the germ cell wall, composed of a peptidoglycan layer and at a later stage the cell wall of the bacterium. The peptidoglycan layer is surrounded by a spore cortex derived from the mother cell. This protects the spore against cellular cortex hydrolases but confers no resistance properties. The surrounding spore cortex has a proteinaceous coat (Figure 6.1c) that prevent spore lysis by protozoal enzymes (Riesenman and Nicholson, 2000, Klobutcher, et al., 2006, Setlow, 2006, Laaberki and Dworkin, 2008). The spore coat is the outer layer in spore-formers such as *Bacillus subtilis*. In contrast, some pathogenic and thermophilic species like *Bacillus anthracis*, *Bacillus cereus*, *C. difficile*, *H. thermocellum* and *P. thermosuccinogenes*, the exosporium (Figure 6.1d) is the outermost layer (Giorno, et al., 2007, Oliva, et al., 2008, Xue, et al., 2010, B Mearls, et al., 2012, Paredes-Sabja, et al., 2014, Stewart, 2015).
Different fluorescence dyes for detecting the culture heterogenity

In earlier studies the discrimination between free spores, endospores, viable and damaged/dead cells within a sample has been done using fluorochromes, based on respiratory activity, membrane integrity, enzymatic action and DNA/RNA content. These assays have been used extensively as a viability marker for several bacterial species, which differentiates live and damaged/dead cells or spores under limiting conditions (Fiksdal and Tryland, 1999, Bunthof and Abe, 2002, Hoefel, et al., 2003, Tracy, et al., 2008). For detecting cell viability and sporulation heterogeneity, *P. thermosuccinogenes* DSM5809 was grown in CP medium starting from spore suspension (t0). Samples were taken at 5 and 20 hours referred as t5 and t20, respectively. The population heterogeneity of the strain was analysed using combinations of two dyes either carboxyfluorescein diacetate (cFDA)/propidium iodide (PI) or Syto9/PI dyes. Application of the cFDA/PI dyes enabled the division of *P. thermosuccinogenes* DSM5809 into different populations with overlapping areas. The predicted populations based on microscopy...
and uptake of cFDA/PI dyes were grouped as mature spores (H3), endospores (H2) and vegetative cells (H1) as differentiated by BD Accuri C6 flow cytometry (Supplementary data, Figure S1). At stationary phase, a clear extra population appeared suggesting dead/damaged cells. Patakova et al. discussed the double staining (PI and cFDA) technique, suitable for assessing the number of non-sporulating vegetative cells during exponential phase of \textit{C. beijerinckii} and \textit{C. tetanomorphum} and concluded no distinct relation between cFDA/PI staining and morphologically diverse cell forms could be obtained (Patakova, et al., 2013). Another study displayed different morphological states of \textit{C. pasteurianum} and \textit{C. beijerinckii} inclusive of active cells, doubly stained cells, damaged cells and spores were discriminated based on combination of light scattering and fluorescent staining (PI/cFDA) (Kolek, et al., 2016). Likewise, for \textit{P. thermosuccinogenes} our initial results of flow cytometry hinted at better separation of the sporulating populations by cFDA/PI double staining. Conversely, when a combination of Syto9 and PI dyes was used no clear separation of populations was observed, conceivably due to overlapping excitation/emission spectra of Syto9 and PI. Therefore, we selected the combination of cFDA/PI dyes as an activity/viability marker to distinguish different populations including metabolically active vegetative cells and mature dormant spores.

**cFDA/PI double staining allows identification of \textit{P. thermosuccinogenes} phenotypes at different stages of sporulation**

\textit{P. thermosuccinogenes} DSM5809 was cultured from a spore suspension to synchronize growth. Three different time points of cultivation were analysed from pure spores at the start (t0), early growth stages (t5) and late exponential phase (t20). These samples were then examined by flow cytometry analysis (FACSAria III, BD Biosciences, USA) for their LS characteristics including FSC-H and SSC-H in combination with cFDA and PI double staining. FSC is the scattering of light axial to the laser and is mostly due to light diffraction around the cell. The intensity of the signal obtained via photodiode can often be used as an indicator of particle size. SSC is perpendicular light scattering measured in an orthogonal direction to the laser showing particle properties related to cell morphology, cellular content and granularity (Vembadi, et al., 2019). The bacterial vegetative cell and spore populations were gated to make a clear separation from cell debris. Notably, the spores could be separated from cell debris, based on their increased intensity in the FSC-H plot as shown in Figure 6.2a (IV). Inside the gated region for FSC-H/SSC-H plot (I) and cFDA/FSC-H (IV), the first time point t0 containing a pure spore suspension showed a high FSC-H, minor side scattering and no cFDA fluorescence. Notably, cFDA/PI dyes stained vegetative cells but not spores as shown in SSC-H/cFDA (II) and SSC-
H/PI (III) plots (Figure 6.2a), in line with previous results from Kolek et. al. (Kolek, et al., 2016). Therefore this methodology offering options for discrimination of sporulation heterogeneity of P. thermosuccinogenes.

![Figure 6.2. LS investigation with cFDA/PI dyes at start t0 (0 hour), early t5 (5 hours) and late exponential phase t20 (20 hours) cultures of P. thermosuccinogenes DSM5809. Contour plots FSC/SSC (I), SSC/cFDA (II), SSC/PI (III) and FSC/cFDA (IV) of a) pure spore suspension (t0), b) at t5 and c) at t20 of cultivation in CP medium at 60 °C using LS (FSC-H/SSC-H) analysis in comparison to cFDA/PI.](image)

At t5 a shift along the SSC-H axis for FSC-H/SSC-H (I) was observed in Figure 6.2b. The SSC-H higher intensity illustrates complexity or altered diffraction of light of the cell in comparison to spores. Figure 6.2b displays the SSC-H/cFDA plot (II) with three divergent populations and two subpopulations. This includes: 1) lower region - cFDA low/SSC-H low with another extended subpopulation (high SSC-H), 2) intermediate region - SSC-H middle/cFDA middle and 3) upper region - SSC-H high/cFDA high involves overlapping metabolically active subpopulations, respectively. The SSC-H/PI (III) plot presents an extra population containing dead/damaged cells or spores. FSC-H/cFDA (IV) showed three cFDA based populations without any FSC subpopulations. The mature free spores were separated based on their increased autofluorescence in the FSC-H channel. These spores had lower signal intensities than stained cells but greater than debris or non-stained vegetative cells. The exact mechanism for spore autofluorescence is unknown, but it could be due to the composition of spore coat or
its oxidase or peroxidase activity (Mohapatra and La Duc, 2013). Spore autofluorescence as a sorting condition was used for Bacillus spores (Laflamme, et al., 2006) or via a combined strategy of LS and fluorescence staining discriminated Paenibacillus spores from vegetative cells (Comas and Vives-Regó, 2002).

Furthermore, after cultivating *P. thermosuccinogenes* DSM5809 for 20 hours (t20, Figure 6.2c) the various populations were better defined compared to the pattern seen at t5 (Figure 6.2b). An extra population appeared with high FSC-H/SSC-H (Figure 6.2c, I), revealing longer cells and significantly different diffraction of light showing changed cell complexity. FSC-H/cFDA (IV) stain showed the same five major populations composed of metabolically active/inactive cells of different sizes or intensities in their signals. This will be further discussed in details in the next section. Therefore, all populations reflecting the different cell morphologies and stages during sporulation of *P. thermosuccinogenes* DSM5809 were identified.

**FACS analysis for detailed insights on various populations of *P. thermosuccinogenes***

Analysis of the spore populations during late exponential phase provided insights into population diversity with the best separation using SSC-H/cFDA contour plots (Figure 6.3, main plot). The FSC-H/SSC-H plot also showed different populations based on both size and LS effects. In this case, LS only allows to differentiate spores from vegetative cells but not other subpopulations. To enable this, it was necessary to use the metabolic dye cFDA as a mean of adding a new dimension to the analysis. SSC-H/cFDA main plot revealed five populations with overlapping regions. To characterize these groupings we decided to sort them using FACS Aria III. We also compared this main plot with other sub-plots FSC-H/cFDA (I), FSC-H/PI (II) and SSC-H/FSC-H (III), respectively, to get a correlation between them. The various distinguished populations were as follow (Figure 6.3, main plot):

**A1 - SSC-H low/cFDA high (vegetative cells):** The uppermost left corner was shown to represent vegetative cells with high metabolic activity as shown in Figure 6.3 (main plot A1). We sorted one million cells, observed them using phase contrast microscopy and heat-treated them to select for the presence of spores. This resulted in killing of all cells confirming it to be a vegetative cell population absent of spores. The subplots FSC-H/cFDA highlighted the A1 (I) region with high cFDA uptake. The absence of PI staining (subplot, A1 (II)) for this population indicated active cells with intact membranes.
A2 - SSC-H high/CFDA high (overlap area, endospores and forespores): Two overlapping populations based on the refractive index of light due to similar SSC properties were identified as shown in Figure 6.3 (main plot A2). Interestingly, the sorted populations depict dark/bright phase endospores and forespores, as confirmed by microscopy. The main plot SSC-H/cFDA gave a better separation of overlaying populations between forespores and vegetative cells when compared to FSC-H/cFDA (Figure 6.3, subplot A1 (I)). This shows no significant size differences, but alterations in light diffraction due to cellular complexity. Moreover, this specific population apart from the dormant and germinating spores had (partial) outgrowth after heat inactivation. Partial, due to the existence of spores that are still in the germination process and have not yet acquired (complete) heat resistance.

A3 - SSC-H middle/CFDA middle (germinating spores): The middle area when sorted displayed populations consisting of germinating spores (Figure 6.3, main plot A3). We observed under the phase contrast microscope germinating dark swollen spores slightly stained with cFDA dye. They are similar to spores based on their size with no shift in FSC-H, but move in SSC-H axis to the mid region (Supplementary data, Figure S2, subplot A3 (I)).

A4 - SSC-H low/CFDA low (free spores): This single population resembles the starting spore suspension with mature free spores which are dormant (Figure 6.3, main plot A4). Particles do not stain with cFDA or PI and are smaller in size than the endospores. Previous studies showed similar outcomes with mature free spores stained with cFDA/PI to understand culture heterogeneity in C. beijerinckii NCIMB 8052 and C. pasteurianum NRRL B-598 during solvent production (Kolek, et al., 2016).

A5 - SSC-H high/CFDA low (dead or damaged spores): The last population in the rightmost lower corner which are extending from the mature spores has two overlying areas as depicted in Figure 6.3 (main plot, A5). It has no cFDA staining, indicating the presence of inactive cells (Supplementary data, Figure S2, subplot A5 (I)). One of the populations is PI stained showing dead or damaged cells with high side and forward scatter (Supplementary data, Figure S2, subplot A5 (II)). We could not obtain clear microscopical conformation of this population due to its low population size resulting in low sort counts. Based on the parameters, we assume it is a mixed population of dead/damaged cells and spores.

Finally, to give a supporting overview, FSC-H/cFDA and SSC-H/FSC-H contour plots were obtained by overlaying the five individual populations (Supplementary data, Figure S3). The FSC-H/cFDA plot shows large overlap between vegetative cells (A1, red) and forespores/endospores (A2, light blue). Furthermore, A4 (light green) partially overlaps with
A5 (dark green) subpopulations based on FSC-H. For the SSC-H/FSC-H plot, it does not provide a better representation as the A1 population is partially hidden due to an overlap with the A4 population. Therefore, when these two FSC based plots were compared to SSC-H/cFDA, we concluded that the highest resolving power for all the diverse populations was attributed to the latter one. Hence, these results showcase the use of flow cytometry in combination with SSC-H/cFDA parameter, acts as a potential tool to effectively assess the complex culture heterogeneity of thermophilic clostridia.

CONCLUSIONS

Our study using succinate producer *P. thermosuccinogenes* as a model, shows that FCM can be an efficient and quick method for assessment of population heterogeneity and sporulation dynamics of anaerobic spore-forming bacteria. The developed flow cytometry-based method combines SSC-H/cFDA parameter analysis to quantify culture heterogeneity of *P. thermosuccinogenes*, discriminating mature spores, dark and bright phase endospores, forespores and vegetative cells. It can be applied to study sporulation dynamics during the fermentation, optionally after sorting, to determine physiological parameters that may provide leads for further optimization of the process.
Figure 6.3. Detailed insights of LS investigation with cFDA/PI staining for late exponential phase cultures of *P. thermosuccinogenes* DSM5809 at t20 (20 hours). Populations were sorted based on the combination of SSC-H and cFDA staining characteristics. Five distinct populations were sorted by FACS in the main plot: A1 - SSC-H low/cFDA high, vegetative cells; A2 - SSC-H high/cFDA high, endospores and forespores; A3 - SSC-H middle/cFDA middle, germinating spores; A4 - SSC-H low/cFDA low, mature free spores and A5 - SSC-H high/cFDA low, dead or damaged spores. SSC-H and cFDA plot were correlated with cFDA/FSC-H (I), PI/FSC-H (II) and SSC-H/FSC-H (III) contour subplots to get a clear picture of major morphologies of each sorted population. *P. thermosuccinogenes* DSM5809 was grown in CP medium at 60 °C for 20 hours. At t20 samples were sorted and analysed using a FACS Aria III, BD Biosciences, USA.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains and growth conditions**

*P. thermosuccinogenes* DSM5809 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). *P. thermosuccinogenes* was grown anaerobically in CP medium adapted from (Plugge, 2005). The final volume of 50 mL medium was dispensed...
in serum bottles under 80:20 N₂/CO₂ atmosphere with ~70 kPa overpressure and then autoclaved. Glucose as a carbon source was autoclaved separately and added later to a final concentration of 5.0 g/L.

**Phase contrast and Fluorescence Microscopy**

Phase contrast and fluorescence microscopy (Olympus BX51) were used to determine the morphological status of cells and their staining patterns. Bacterial cells were observed by phase contrast microscopy at 400x and 1000x magnifications. Counting of spores was performed with 400x magnification with an oil immersion lens.

**Transmission Electron Microscopy (TEM)**

A suspension of spores of each strain was pelleted for 2 minutes at 10,000 rpm and the spores were suspended and fixed in 4% (v/v) gluteraldehyde (EMS, Washington D.C., USA), buffered at pH 7.2 with 0.1M phosphate buffer for 18 hours at 4 °C. After three times washing with buffer (20 minutes), the samples were post-fixed for 64 hours at 4 °C with 1% (w/v) OsO₄ (EMS, Washington D.C., USA) in 0.1M phosphate buffer pH 7.2. Next, the samples were washed 5 times (10 minutes) with dH₂O and dehydrated in a graded ethanol series (10, 30, 50, 70, 90 and 100% ethanol) at 10 minutes per step. Between each step the spore samples were pelleted for 2 minutes at 10,000 rpm and vortexed again. Subsequently, the samples were infiltrated with quetol resin (EMS, Washington D.C., USA) by adding small amounts of resin during 6 hours after which the quetol concentration was raised till 65%. Finally, for three days and by three changes of 100% quetol the samples were fully infiltrated with quetol resin. Each day the samples were centrifuged for 3 minutes at 14,000 rpm and the pelleted spores were resuspended in fresh quetol mixture. Hereafter a few droplets of quetol containing spores were put in conical tip BEEMTM capsules (EMS, Washington D.C., USA), covered with fresh quetol resin, centrifuged for 3 minutes at 14,000 rpm and polymerized in an embedding oven (Agar scientific, B702, England) for 48 hours at 60 °C. Ultra-thin sections (80 nm) were made with an ultra-microtome (Reichert Ultra Cut S) using a diamond knife. During sectioning, one micrometer distance was left between the sectioned areas. The sections were examined with a TEM (JEM 1011 Jeol, Tokyo, Japan).

**Spore suspension preparation**

*P. thermosuccinogenes* DSM5809 was grown anaerobically at 60 °C overnight in CP medium. In order to trigger sporulation, 1 mL of pre-culture was added to sporulation medium adapted
from Yang et al. and B Mearls et al. (Yang, et al., 2009, B Mearls, et al., 2012), cultured at 55 ºC for 2 days. The Modified AEA Sporulation Medium Base (catalogue number 17170, Sigma Aldrich) was prepared and 30 mL dispensed into anaerobic bottles under 80:20 N₂/CO₂ with ~70 kPa overpressure and then autoclaved. After autoclavation the medium was supplemented with 4 % (v/v) of glucose solution, freshly prepared and filter sterilized 1.33 % (v/v) of sodium carbonate (Na₂CO₃) and 1.33 % (v/v) cobalt chloride hexahydrate (CoCl₂·6H₂O), respectively. Then, 1.33 % v/v of freshly prepared sodium ascorbate solution (C₆H₇O₆Na) was added. 1 mL of the pre-culture was added to the sporulation medium and incubated at 55 ºC. Once spores were formed 1 mL of pre-culture from sporulation medium was inoculated into 50 mL CP medium and grown for 2 days at 60 ºC. When mature spores were highly predominant, spore suspension was harvested aerobically by centrifugation at 4700 rpm for 15 minutes. Cell pellets were resuspended, gently layered on 50 % Histodenz solution and centrifuged at 3000 g, 20 ºC for 60 minutes. Spore pellet was collected and resuspended in 15 mL Milli-Q water. Centrifuged at 1500 rpm, 20 ºC for 30 minutes and supernatant was discarded. Mature spores pellets were resuspended in 10 mL Milli-Q water. 1 mL aliquots of mature spores were stored at -80 ºC.

**Flow cytometry and FACS analysis**

*P. thermosuccinogenes* was grown anaerobically in CP medium and 1 mL culture were collected at different time points (0, 5, 20 hours). Cells were collected and immediately washed twice for 1 minute, at 1000 rpm) with sterile, microfiltered (0.22 µm) 0.9% NaCl solution. Aliquots of 200 µL cells were stained with either cFDA (22 µM)/PI (15 µM) or Syto-9/PI (100 µM) combination of dyes. Samples were kept at room temperature in the dark and incubated for 10 minutes. 20 µL of sample were diluted in 2 mL of ultrapure H₂O and immediately analyzed in FCM. This protocol was adapted from Kolek et al. (Kolek, et al., 2016). For the FCM, 10,000 events were analyzed. FSC, SSC, green (FL1; 515–565 nm) and red (FL3; > 605 nm) fluorescence were measured using an Accuri C6 cytometer (BD Accuri Cytometer Inc., USA). It is equipped with a 20 mW, 488 nm, Solid State Blue Laser. FSC and SSC characteristics were used as trigger signals. Samples were stained at close range to the flow cytometer and analysed immediately following incubation. For the FACS analysis the same sample preparation was followed. We have used a FACSARia III (BD Biosciences, USA) with lasers 488 nm for FSC/SSC and cFDA/PI double staining (502 LP with 530/30 filter; 488 nm and 600 LP with 610/20 filter; 561 nm). Pre-selection of data occurred using FSC and SSC signals and the background noise removed by gating of bacterial populations. Finally, 50,000
events were collected, and analysis was carried out using software FlowJO 10 (FLOWJO, LLC and BD Biosciences, USA; https://www.flowjo.com/solutions/flowjo).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank Tiny Franssen-Verheijen for the electron microscopy work which was done at the Wageningen Electron Microscopy Centre, Wageningen, The Netherlands. We are grateful to Remco Hermans and Victoria Lisenko for their suggestions on the initial flow cytometry experiments.

Funding: This work was supported by Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068].
SUPPLEMENTARY DATA

**Figure S1** Histogram plots of cFDA staining of early exponential t5 (5 hours) and late exponential phase t20 (20 hours) cultures of *P. thermosuccinogenes* DSM 5809. At t5 the plot has two overlapping peaks predicted as vegetative cells and endospores referred as H1 and H2, respectively. At t20 peak H3 can be potentially mature free spores. The strain was grown in CP medium at 60 °C for 24 hours. At t5 and t20 samples were analysed using the Accuri C6 cytometer (BD Accuri Cytometer Inc., USA).

**Figure S2** Detailed insights of LS investigation with cFDA/PI staining for late exponential phase cultures of *P. thermosuccinogenes* DSM5809 at t20 (20 hours). Populations were sorted based on the combination of SSC-H and cFDA staining characteristics. Five distinct populations were sorted by FACS: A1 - SSC-H low/cFDA high, vegetative cells; A2 - SSC-H high/cFDA high, endospores and forespores; A3 - SSC-H middle/cFDA middle, germinating spores; A4 - SSC-H low/cFDA low, mature free spores, and A5 - SSC-H high/cFDA low, dead or damaged spores. SSC-H and cFDA plot were correlated with cFDA/FSC-H (I), PI/FSC-H (II) and SSC-H/FSC-H (III) contour plots to get a clear picture of major morphologies of each sorted population. *P. thermosuccinogenes* DSM5809 was grown in CP medium at 60 °C for 20 hours. At t20 samples were sorted and analysed using the FACSAria III, BD Biosciences, USA.
**Figure S2** Detailed insights of LS investigation with cFDA/PI staining for late exponential phase cultures of *P. thermosuccinogenes* DSM5809 at t20 (20 hours). Populations were sorted based on the combination of SSC-H and cFDA staining characteristics. Five distinct populations were sorted by FACS: A1 - SSC-H low/cFDA high, vegetative cells; A2 - SSC-H high/cFDA high, endospores and forespores; A3 - SSC-H middle/cFDA middle, germinating spores; A4 - SSC-H low/cFDA low, mature free spores and A5 - SSC-H high/cFDA low, dead or damaged spores. SSC-H and cFDA plot were correlated with cFDA/FSC-H (I), PI/FSC-H (II) and SSC-H/FSC-H (III) contour plots to get a clear picture of major morphologies of each sorted population. *P. thermosuccinogenes* DSM5809 was grown in CP medium at 60 °C for 20 hours. At t20 samples were sorted and analysed using the FACS Aria III, BD Biosciences, USA.
Figure S3 Contour plots of FSC-H/cFDA and SSC-H/FSC-H of all the five populations depicted by A1 (red), A2 (blue), A3 (orange), A4 (light green) and A5 (dark green). For FSC-H/cFDA plot A1 (red), A2 (light blue) and A4 (light green), A5 (dark green) are overlapping, respectively. SSC-H/FSC-H plot shows A4 (light green), A1 (red) and A2 (light blue), A5 (dark green) overlaying subpopulations. *P. thermosuccinogenes* DSM5809 was grown in CP medium at 60 °C for 20 hours.
Chapter 7

Summary and general discussion
Chapter 7

SUMMARY

Sustainable production of bio-based products can significantly reduce our dependency on fossil resources. Thus, economic production of green chemicals as building blocks for bio-based products has become a key research target. The ability of thermophilic clostridia to ferment lignocellulose biomass, make them attractive as potential production organisms. However, the limitation of genetic accessibility and genetic tools for clostridial manipulation pose a major barrier for exploiting them as industrial platform organisms (chapter 1). This thesis explores the genetic accessibility and tool development for two thermophilic clostridia. Moreover, we studied aspects such as xylose consumption, sporulation and CRISPRi-mediated silencing of central metabolism to improve strain engineering.

For development of improved microbial cell factories, high-throughput genetic tools are desired. To enhance strain engineering efforts, CRISPR-Cas-based technologies have been used in bacteria. In chapter 2 we review bacterial metabolic engineering for production of bio-based products, related to the exploitation of new CRISPR-Cas variants. This technology expedited metabolic engineering of non-model organisms like bacilli, clostridia and cyanobacteria to exploit them as platform organisms. In addition, we discuss about the importance of screening novel Cas9-like systems from diverse microbial environments to extend the currently available technologies.

In chapter 3 we evaluate the use of ThermodCas9 in the thermophile Hungateiclostridium thermocellum DSM1313. The ThermoCas9-based CRISPRi was utilized to silence central metabolic genes targeting lactate dehydrogenase (ldh) and phosphotransacetylase (pta) gene expression. The guide RNA was under the control of native intergenic 16S/23S rRNA promoter from H. thermocellum. Both the ldh and pta genes targeted at the promoter were repressed up to 67% and 62%, respectively. This resulted in 24% and 17% decrease in lactate and acetate production, respectively. Hence, these data established the possibility of employing CRISPRi-mediated silencing of metabolic genes to modulate pathways without genome engineering.

Genetic manipulation of the thermophilic anaerobe Pseudoclostridium thermosuccinogenes is desired to improve its succinate production. In chapter 4 we established the genetic system by electroporation-mediated transformation with efficiency up to $10^2$ CFU/μg plasmid DNA. To this end, we overcame the restriction barriers by in vivo methylation of plasmid DNA when transformed into an engineered E. coli HST04 strain expressing four native methylation systems of the thermophile. Furthermore, the same approach was used to apply the CRISPRi silencing
tool targeting the malic enzyme gene in *P. thermosuccinogenes*. This led to 75% downregulation of the gene expression with an impact on lactate and acetate production. This genetic tool development for *P. thermosuccinogenes*, now paved the way for metabolic engineering of this bacterium.

Apart from development of genetic tools, this thesis also deals with heterologous expression of genes from *P. thermosuccinogenes* into *H. thermocellum*. This thermophile has been of interest for consolidated bioprocessing (CBP), because of its capability to ferment cellulose without any pre-treatment or additional enzymes. *H. thermocellum* has excellent cellulolytic activity for six-carbon substrates, but has the inability to grow on five-carbon substrates such as xylose. A plasmid-based GTP-dependent xylose gene cluster from *P. thermosuccinogenes* was introduced into *H. thermocellum*. This facilitated the growth of the latter on xylose. In chapter 5, we have shown that this resulted in increased ethanol and acetate production for the xylose grown strains.

In chapter 6, population heterogeneity of *P. thermosuccinogenes* was studied with single-cell analysis techniques such as flow cytometry (FCM) and fluorescence-assisted cell sorting (FACS). FCM with FACS plus fluorescent staining by propidium iodide and carboxy fluorescein diacetate (cFDA) enabled the discrimination and enrichment of all sporulation populations. Side scatter (SSC-H) with metabolic indicator cFDA dye provided the best separation of varied populations. Based on this method, we successfully differentiated between mature spores, forespores, dark and bright phase endospores, and vegetative cells of *P. thermosuccinogenes*. Henceforth, this protocol can be applied to study further the sporulation dynamics and its impact on fermentation performance and product formation of the organism.

In summary, this PhD thesis focuses on the non-model thermophilic organisms such as *H. thermocellum* and *P. thermosuccinogenes* as bacterial cell factories. Taking the succinate producer *P. thermosuccinogenes* as a model, we described strategies to overcome restriction barriers by *in vivo* methylation of plasmid DNA to have a genetically accessible strain. As a proof-of-concept, CRISPRi silencing tool was developed for both the thermophilic clostridia and applied to impact their transcriptional expression and fermentation profiles. Lignocellulolytic capacities of *H. thermocellum* were improved by the introduction of xylose genes from *P. thermosuccinogenes*. Additionally, tools to study population heterogeneity and sporulation dynamics of *P. thermosuccinogenes* were developed for future optimization of industrial fermentations.
GENERAL DISCUSSION

Non-model microbes as platform organisms have been exploited for the conversion of biomass to fuels and high value chemicals to reduce the dependency on unsustainable processes. To develop a platform organism, the crucial aspects considered are genetic accessibility plus genome editing tools, to modulate genes for better understanding of their metabolism and fermentation efficiency to enhance green chemicals production. In this thesis, we described the first steps to develop non-model thermophile *P. thermosuccinogenes* as platform organism by addressing genetic accessibility. Then we applied CRISPRi as a transcriptional modulator. This CRISPRi system was also applied to another thermophile, *H. thermocellum*, to modulate metabolic gene expression. For both the organisms silencing has an impact on fermentation profile as discussed in the respective chapters 3 and 4 in detail. Still, there remain many open questions about their metabolism that need to be dealt with. The introduction of xylose genes from *P. thermosuccinogenes* into *H. thermocellum* as the latter is unable to grow on xylose has affected product formation (chapter 5). This last chapter will discuss the strategies to develop a better genetic system and the effects on metabolism, due to metabolic or physiological conditions in non-model thermophiles.

GENETIC ACCESSIBILITY FACTORS

DNA transfer mechanisms

Genetic accessibility is facilitated by the transfer of exogeneous DNA into bacteria, which is quite challenging and requires developing strain specific and, preferably, easy and high-throughput procedures for non-model organisms. The major approaches for DNA transfer have been conjugation, electroporation or induced natural competence. Electroporation is mostly employed in clostridia due to simple technicality, more reliability, no dependence on donor, and high transformation efficiency (Pyne, et al., 2014). Transformation by electroporation involves creation of pores using electrical pulses on the cell envelope becoming permeable for DNA uptake (Chassy, et al., 1988, Teissie and Rols, 1988). Electroporation was optimized for the non-model thermophile *P. thermosuccinogenes* using plasmid pNW33n and yielded transformants as discussed in detail in chapter 4. As an alternative to electroporation, conjugation was also tested, which includes direct cell-to-cell transfer of plasmid DNA through the pilus from donor to recipient species (Willettts, 1993). Conjugation was done with *E. coli* HB101 containing transferrable plasmid pJS28 (a pNW33n-derivative with oriT) plus pRK
helper plasmid (incompatibility (Inc) group called Inc-P, extrachromosomal genetic elements that aids in transfer and replication) and \textit{P. thermosuccinogenes} DSM5809 as acceptor. Conjugation was performed at 37 ºC, 42 ºC and 45 ºC under anaerobic conditions using different ratios of donor to recipient (1:1, 5:1, 10:1), after which the mixtures were plated to select for transconjugants. The experiments did not result in colonies. The probable reason could be the need to have optimal temperature for both the donor and recipient strains as \textit{E. coli} grows at 37 ºC and \textit{P. thermosuccinogenes} grows best at 55 ºC. Despite the fact that conjugation theoretically circumvents the restriction barrier (Purdy, et al., 2002), numerous reports displayed a dramatical drop in the conjugal transfer efficiency due to the presence of active restriction-modification (RM) systems in the recipient host (Elhai, et al., 1997, Purdy, et al., 2002). Thus, we used \textit{E. coli} JM110 and HST04 as donor strains to bypass the restriction activities but no conjugants were obtained. Furthermore, the precise mechanisms of DNA transport combined with the exact factors to trigger these systems remains mostly elusive (Llosa, et al., 2002, Mell and Redfield, 2014, Muschiol, et al., 2015). Hence, electroporation was preferred for effective DNA transfer.

\textbf{Genome sequencing for methylome analysis}

For genetic accessibility, a critical step entails the genome analysis of the target organism to indicate the presence of native RM systems. Single-molecule real-time (SMRT) sequencing aided the identification of bacterial methylome with the finding of methylated nucleotides, identification of recognition sites (motifs) and methyltransferases present, (Eid, et al., 2009, Flusberg, et al., 2010, Fang, et al., 2012, Krebes, et al., 2014, Blow, et al., 2016, Beaulaurier, et al., 2019) available in the centralized REBASE database (Roberts, et al., 2010). In addition, SMRT sequencing provided major types of DNA methylation in bacteria such as N6-methyladenine (\textit{\textsuperscript{6m}A}), 4-methylcytosine (\textit{\textsuperscript{4m}C}) and 5-methylcytosine (\textit{\textsuperscript{5m}C}) (Flusberg, et al., 2010, Blow, et al., 2016, Beaulaurier, et al., 2019). The \textit{\textsuperscript{6m}A} modifications are abundant in bacteria and SMRT sequencing unveiled \textit{\textsuperscript{6m}A} MTase activity required for methylation-based host defense mechanisms. For \textit{P. thermosuccinogenes} DSM5809, SMRT Pacbio sequencing revealed five motifs with \textit{\textsuperscript{6m}A} methylation. The motifs confirm the presence of type I and type III restriction-modification systems in \textit{P. thermosuccinogenes} DSM5809 (\textbf{chapter 4}). This permitted to search for these putative motifs in plasmids such as pNW33n used for transformation of this strain. We identified three motifs that might potentially limit the transformation of pNW33n plasmid. This explains the low reproducibility and transformation efficiency of this strain. To test this hypothesis, plasmid pNW33n\textunderscore RM5809 (pNW33n lacking
all but one of the detected motifs) was constructed with mutated recognition sites, so they cannot be recognized by the identified native restriction systems of *P. thermosuccinogenes* DSM5809. This approach improved reproducibility among several electroporation attempts compared to pNW33n plasmid. Though the transformation efficiency was not enhanced, maybe as one site is still not mutated. This shows that for a quick check, using genome and methylome analysis can overcome RM systems, but *in vivo* methylation of plasmid DNA is necessary to improve transformation efficiency as discussed in detail later.

Other types of methylation are $5^m$C modification with cytosine methylase (Dcm) in *E. coli*, with the target sequence 5’-CCWGG-3’. Dcm methylation has been associated with Tn3 transposition (Yang, et al., 1989) and lambda phage recombination (Korba and Hays, 1982). However, the biological role of $5^m$C in bacteria remains unclear and recent studies revealed this modification facilitates overcoming restriction barriers in few strains (O’Connell Motherway, et al., 2014). Also, $4^m$C modification in bacteria have unknown functionality (Beaulaurier, et al., 2019). Detailed studies are required to understand the role of these modifications. Recently, new sequencing platforms like MinION, nanopore sequencing by Oxford Nanopore Technologies (Clarke, et al., 2009, Laszlo, et al., 2014, Jain, et al., 2016, de Lanno, et al., 2017) have provided insights in detection of DNA methylation patterns. The detection method depends on base calling the canonical bases and when signals differ between canonical bases and modified nucleotides, predicts the specific DNA modifications in both prokaryotic and eukaryotic methylomes (Rand, et al., 2017, Stoiber, et al., 2017, McIntyre, et al., 2019). The DNA modifications with trained hidden Markov model (HMM) and hierarchical Dirichlet process (HDP) can identify methylases of known specificity with expected $4^m$C, $5^m$C and $6^m$A motifs in bacteria (Rand, et al., 2017). However, the detection accuracy remains a challenge with various types of methylation and motif specificities (Stoiber, et al., 2017). To conclude, both SMRT and Minion sequencing can probe an organism’s methylation signature, that can allow identification of putative RM systems with motifs, which has valuable implications for strain manipulation. For instance, we performed *in vivo* methylation of the plasmid DNA by mimicking the *P. thermosuccinogenes* strain’s methylation pattern in *E. coli* HST04, that increased transformation efficiency (chapter 4). Moreover, we predict the motif ‘GATNNNNNCTC’ of *P. thermosuccinogenes* found in plasmid pNW33n is responsible for hampering genetic accessibility by being recognized by one of the RM systems that will cleave the plasmid DNA. To validate this, we performed Minion sequencing on the engineered *E. coli* HST04 strains and with preliminary methylome analysis we found *E. coli* HST04_S1 strain
with two type I system has $\text{6m}A$ methylated significantly in the sequence ‘GATNNNNNCTC’ compared to the wild type \textit{E. coli} HST04 strain (data not shown). Hence, the fast pace advances in the sequencing technology make both SMRT and nanopore sequencing a promising approach to identify different methylation patterns to overcome genetic inaccessibility.

### Role of native bacterial defense systems

In general, bacteria have native defense systems to prevent exogeneous DNA uptake. These defense systems support the organism to discriminate foreign DNA from self-DNA and may permit them to adapt to specific environments and control evolution (Jeltsch, 2003, Vasu and Nagaraja, 2013). However, these roles are important for the bacteria but can affect genetic accessibility by restricting the foreign DNA transfer. Therefore, to make non-model organisms genetically accessible and engineer them further, it is vital to know the native defense systems and ways to overcome them.

#### Restriction modification systems

The defense mechanism of microbes against foreign DNA has been RM systems which have been described in \textbf{chapter 1}. These systems are composed of methyltransferases (MTases) and restriction endonucleases (REases). Both these enzymes recognize a specific target DNA sequence. The MTase methylates or hemimethylates an adenine or a cytosine in the target sequence whereas the REase will cleave unmethylated DNA in the target sequence. In contrast, type IV REases cleave methylated DNA. Thus, it is necessary to circumvent these RM systems for successful transformation. As mentioned in \textbf{chapter 4}, Pacbio sequencing recognized the RM systems in \textit{P. thermosuccinogenes} DSM5809. Previous studies as described in Table 7.1, showed \textit{in vivo} methylation of the plasmid DNA by mimicking the host’s methylation pattern increases transformation efficiency (Kwak, et al., 2002, Suzuki, 2012, Zhang, et al., 2012, Riley, et al., 2019).

For mimicking the methylation system, MTases from the target organism must be expressed in \textit{E. coli}. Consequently, from the same \textit{E. coli}, plasmids must be isolated and transformed to the target organism (Mermelstein and Papoutsakis, 1993, Yasui, et al., 2009). Expressing multiple MTase in \textit{E. coli} increases transformation efficiency compared to expressing only a single MTase (Zhang, et al., 2012). Considering this approach, native genes encoding type I RM systems from \textit{P. thermosuccinogenes} DSM5809 were expressed in \textit{E. coli} HST04 (Figure 7.1, Genetic accessibility). The plasmid DNA was isolated from engineered \textit{E. coli} HST04 and transformed to the target strain. The transformation efficiency of $10^2$ CFU/\textmu g plasmid DNA
was achieved compared to without \textit{in vivo} methylated plasmids. Though the transformation efficiency was improved but the unknown effect of the temperature difference between the optima for the two species can still affect the activity of the methyltransferases \textit{in vivo}. Therefore, \textit{in vivo} methylation is effective compared to silent mutations as in our case the site close to promoter of chloramphenicol resistance marker in the plasmid is not mutated, which is crucial for increasing transformation efficiency.

Another possibility would be \textit{in vitro} methylation of plasmid DNA. This method involves incubation of plasmid DNA with cell-free extract from the target organism or commercial MTase or CpG MTase before transformation to enhance the efficiency (Chen, et al., 2008, Kim, et al., 2010, Fu, et al., 2017). We incubated the plasmid pNW33N with cell-free extract of \textit{P. thermosuccinogenes} and did not get any transformants. This might be due to the procedure being suitable only for type II RM systems but \textit{P. thermosuccinogenes} lacks those systems, thus this approach was not applicable.

The discrimination from self to foreign DNA can also be facilitated in bacteria by stand-alone MTases such as Dam and Dcm. The Dam has the recognition sequence ‘GATC’ with a modified adenine whereas Dcm has CC(A/T)GG with internal cytosine methylation. Matching the Dam/Dcm methylation of target organism could enhance transformation efficiency by circumventing type IV RM systems, that cleaves methylated DNA (Guss, et al., 2012, Kolek, et al., 2016). With genome analysis we have confirmed that \textit{P. thermosuccinogenes} was lacking \textit{dam} and \textit{dcm} homologues. After the Dam/Dcm methylation pattern of target organism was known, the plasmid DNA to be transformed could be isolated from a selected \textit{E. coli} strain to match the correct methylation: \textit{E. coli} HST04 (\textit{dam''/dcm''}). In contrast, DNA isolated from \textit{E. coli} DH5a (\textit{dam''/dcm''}) or Acella (\textit{dam''/dcm''}) did not give any transformants as explained in chapter 4. The successful transformation with plasmid \textit{E. coli} HST04 indicates the presence of an unknown type IV RM system in \textit{P. thermosuccinogenes}, as it cleaves methylated DNA. Similarly, for \textit{H. thermocellum} (\textit{dam''/dcm''}) has GATC methylation, probably due to a type II RM system and that Dam mimics this methylation pattern. So transformation plasmids prepared from \textit{E. coli} BL21 or Acella (\textit{dam''/dcm''}) improved transformation efficiencies drastically (Guss, et al., 2012). Hence, these above-mentioned RM systems should be studied in detail to overcome genetic inaccessibility for non-model organisms.

<table>
<thead>
<tr>
<th>Host Organism</th>
<th>Method</th>
<th>Transfer Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Clostridium difficile}</td>
<td>\textit{in vivo}</td>
<td>(HG/MR)</td>
</tr>
<tr>
<td>\textit{Clostridium thermocellum}</td>
<td>\textit{in vivo}</td>
<td>(IG)</td>
</tr>
<tr>
<td>\textit{Clostridium cellulovorans}</td>
<td>\textit{in vivo}</td>
<td>(MG)</td>
</tr>
<tr>
<td>\textit{Clostridium acetobutylicum}</td>
<td>\textit{in vivo}</td>
<td>(HG)</td>
</tr>
<tr>
<td>\textit{Geobacillus kaustophilus}</td>
<td>\textit{in vivo}</td>
<td>(MG)</td>
</tr>
<tr>
<td>\textit{Bacillus anthracis}</td>
<td>\textit{in vitro}</td>
<td>(MF)</td>
</tr>
<tr>
<td>\textit{Bacillus cereus}</td>
<td>\textit{in vitro}</td>
<td>(EX)</td>
</tr>
<tr>
<td>\textit{Bacillus weihenstephanensis}</td>
<td>\textit{in vitro}</td>
<td>(EX)</td>
</tr>
<tr>
<td>\textit{Borrelia burgdorferi}</td>
<td>\textit{in vitro}</td>
<td>(MT)</td>
</tr>
</tbody>
</table>

General discussion
Chapter 7

E. coli was achieved compared to without (Guss, et al., 2012) RM system and that Dam mimics this methylation pattern. So transformation plasmids prepared

Similarly, for in

E. coli known, the plasmid DNA to be transformed could be isolated from a selected

MTases th

Another possibility would be close to promoter of chloramphenicol resistance marker in the plasmid is not mutated, which is therefore, crucial for increasing transformation efficiency.

in vivo methylation of plasmid DNA. This method involves incubation of plasmid DNA with cell-free extract from the target organism or commercial

free extract from the target organism or commercial

P. thermosuccinogenes, P. thermosuccinogenes, HST04 (Guss, et al., 2008, Kim, et al., 2008, Kolek, et al., 2016). With genome analysis we have confirmed that dam homologues. After the

methylation of target organism could enhance transformation efficiency by matching the

Dcm methylation is effective compared to silent mutations as in our case the site 

has CC(A/T)GG with internal cytosine methylation. Matching the

Dcm +Į

HST04 (+Į

+/Į

-Į

-Į

tradicates the presence

in vivo

methylation using a crude extract of the bacterium to be transformed. MR: transforming plasmid with mutated recognition sites. Adapted from (Suzuki, 2012).

Table 7.1. Transformation of different bacterial strains mimicking host methylation patterns

<table>
<thead>
<tr>
<th>Host Organism</th>
<th>Host-mimicking Strategy</th>
<th>DNA-transfer Method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clostridium difficile</td>
<td>In vivo (HG/MR)</td>
<td>Conjugation</td>
<td>(Purdy, et al., 2002)</td>
</tr>
<tr>
<td>Clostridium thermocellum</td>
<td>In vivo (IG)</td>
<td>Electroporation</td>
<td>(Tyurin, et al., 2004)</td>
</tr>
<tr>
<td>Clostridium cellulovorans</td>
<td>In vivo (MG)</td>
<td>Electroporation</td>
<td>(Yang, et al., 2015)</td>
</tr>
<tr>
<td>Clostridium acetobutylicum</td>
<td>In vivo (HG)</td>
<td>Electroporation</td>
<td>(Croux, et al., 2016)</td>
</tr>
<tr>
<td>Geobacillus kaustophilus</td>
<td>In vivo (MG/IG)</td>
<td>Conjugation</td>
<td>(Suzuki and Yoshida, 2012)</td>
</tr>
<tr>
<td>Bacillus anthracis</td>
<td>In vivo (MF)</td>
<td>Electroporation</td>
<td>(Sitaraman and Leplla, 2012)</td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>In vitro (EX)</td>
<td>Electroporation</td>
<td>(Groot, et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>In vivo (MG)</td>
<td>Electroporation</td>
<td>(Zhang, et al., 2012)</td>
</tr>
<tr>
<td>Bacillus weihenstephanensis</td>
<td>In vitro (EX)</td>
<td>Electroporation</td>
<td>(Groot, et al., 2008)</td>
</tr>
<tr>
<td>Bacillus amyloliquefaciens</td>
<td>In vivo (MG)</td>
<td>Electroporation</td>
<td>(Zhang, et al., 2012)</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td>In vitro (MT)</td>
<td>Electroporation</td>
<td>(Chen, et al., 2008)</td>
</tr>
</tbody>
</table>
Chapter 7

Transcription is blocked since the Cas9 protein physically interferes with the RNA polymerase. The ThermodCas9 plasmid of interest (pCRISPRi) was introduced in these H. thermocellum strains by electroporation. The methylated pCRISPRi plasmid DNA was isolated from this strain and electroporated into P. thermosuccinogenes. In this way, the pCRISPRi presents a methylation pattern like that of the genome of P. thermosuccinogenes, overcoming the restriction barrier.

P. thermosuccinogenes was analyzed to identify the native restriction modification (RM) systems. The methylome and genome sequence of P. thermosuccinogenes was analyzed to identify the native restriction modification (RM) systems.

Type I-B system. Previously, subtype I-B Cas protein was characterized for P. thermosuccinogenes. CRISPR-Cas acts as an adaptive immune system that defends microbes from specific Defense Island System Associated with Restriction-Modification systems are both present in bacteria. CRISPR-Cas acts as an adaptive immune system that defends microbes from specific DNA uptake. The CRISPR regions consist of spacers in between the repeats, and the spacer sequences are identical to foreign DNA that the bacteria have been exposed to earlier. When the bacteria are re-exposed to DNA containing a spacer sequence together with proto-spacer adjacent motifs (PAMs), that DNA is cleaved by CRISPR-Cas systems (Horvath and Barrangou, 2010). Both H. thermocellum and P. thermosuccinogenes have the native CRISPR Type I-B system. Previously, subtype I-B Cas protein was characterized for H. thermocellum.

**Figure 7.1.** Schematic overview for in vivo methylation and CRISPRi application for P. thermosuccinogenes DSM5809 (same as Figure 4.1, chapter 4). a) The methylome and genome sequence of P. thermosuccinogenes was analyzed to identify the native restriction modification (RM) systems. b) A set of E. coli HST04 derivative strains was constructed by cloning of the methyltransferase genes (orange and green) of P. thermosuccinogenes. c) The ThermodCas9 plasmid of interest (pCRISPRi) was introduced in these E. coli strains by electroporation. The methylated pCRISPRi plasmid DNA was isolated from this strain and d) electroporated into P. thermosuccinogenes. In this way, the pCRISPRi presents a methylation pattern like that of the genome of P. thermosuccinogenes, overcoming the restriction barrier. e) pCRISPRi expresses the ThermodCas9 and the sgRNA, downregulating the gene of interest upon recognizing and binding the target DNA (promoter region). Transcription is blocked since the Cas9 protein physically interferes with the RNA polymerase. f) Gene repression of desired gene was evaluated by qRT-PCR.

**CRISPR-Cas and DISARM systems**

The CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats) and DISARM (Defense Island System Associated with Restriction-Modification) systems are both present in bacteria. CRISPR-Cas acts as an adaptive immune system that defends microbes from specific DNA uptake. The CRISPR regions consist of spacers in between the repeats, and the spacer sequences are identical to foreign DNA that the bacteria have been exposed to earlier. When the bacteria are re-exposed to DNA containing a spacer sequence together with proto-spacer adjacent motifs (PAMs), that DNA is cleaved by CRISPR-Cas systems (Horvath and Barrangou, 2010). Both H. thermocellum and P. thermosuccinogenes have the native CRISPR Type I-B system. Previously, subtype I-B Cas protein was characterized for H. thermocellum.
that consists of the universal Cas1, 2 and 4 proteins that aid in spacer integration, Cas3, 5, 7 and 8b form the cascade complex and Cas6 protein required for crRNA processing (Richter, et al., 2012, Pyne, et al., 2016). CRISPR-Cas systems do not hinder bacterial transformation, but if a plasmid has a similar DNA sequence to an existing CRISPR spacer, then the plasmid will be cleaved, and transformation will fail (Marraffini and Sontheimer, 2008, Freed, et al., 2018). To overcome this hindrance, single point mutations in the recognition sequence of the plasmid could be introduced, thus avoiding plasmid degradation. In addition, rare events such as plasmid loss after transformation can occur if the microbe acquires spacers from the plasmid due to the presence of CRISPR-Cas systems (Garneau, et al., 2010). Since CRISPR system only attacks specific DNA sequences with PAM recognition, the probability to degrade plasmids are rare.

DISARM is a novel type of multi-gene RM system that defends prokaryotes against their viruses. It is divided into, Class 1 and Class 2 types, consisting of a DNA methylase, helicase domain, phospholipase D (PLD) domain, DUF1998 domain and a hypothetical protein (Ofir, et al., 2018). We examined these domains in \textit{P. thermosuccinogenes} DSM5809; the presence of SNF2 helicase and methyltransferase completely based on annotations hinted the presence of Class 1 DISARM systems. Hence, the \textit{P. thermosuccinogenes} potential DISARM system could still be a limitation for transformation of the \textit{in vivo} methylated DNA. Though the exact reason of this limitation is unknown, the DISARM resembling the RM system, allows that the \textit{in vivo} or \textit{in vitro} strategies can be applied to improve transformation.

\textbf{Transformation parameters}

Apart from native defense systems which pose a challenge for transformation, the other hurdle is the physical barrier imposed by the Gram-positive bacteria for plasmid DNA transfer (Aune and Aachmann, 2010). Gram-positive bacteria have a thick cell wall, so the cells were treated with weakening and membrane solubilizing agents. Studies have shown use of cell-wall-weakening compounds such as glycine, penicillin G, ampicillin, isoniacin and DL-threonine, have an impact on clostridial transformation (Pyne, et al., 2014). Similarly, for \textit{P. thermosuccinogenes}, glycine with sucrose (osmoprotectant) was tested and improved transformation efficiency by 16-fold in comparison to untreated cells. These additives affect cell viability and requires osmotic stabilization, as clostridial cells with damaged cell walls show inhibited growth (Aune and Aachmann, 2010). Similarly, the bacteria’s membrane can be damaged by partial solubilization with ethanol in some organisms (Pyne, et al., 2014), though this did not show any effect on \textit{P. thermosuccinogenes} transformation efficacy. Other
optimization factors, such as electrical parameters, plasmid, amount of DNA, and physiological state of the cells, can vary widely and result in different transformation efficiencies (Nakotte, et al., 1998, Tyurin, et al., 2004, Aune and Aachmann, 2010, Pyne, et al., 2014). As a result, protocol development for optimal transformation efficiency may not be trivial. Taking all this information into consideration, we developed a protocol best for *P. thermosuccinogenes* transformation as described in detail in chapter 4. Therefore, all these parameters entail the development of robust method for improving transformation efficiency in clostridia and other Gram-positive organisms.

**DEVELOPMENT OF GENETIC TECHNIQUES AND TOOLS**

**Host-vector system and selection of promoters**

The development of a DNA transfer method requires construction of a host-vector system, with selection of suitable antibiotic markers and functional origin of replication in clostridia, intended for genetic engineering. Previous studies have presented series of plasmids with four functional origins of replication and clostridial selectable markers such as erythromycin/clarithromycin or thiamphenicol/chloramphenicol, spectinomycin/streptomycin and tetracycline (Heap, et al., 2009). For thermophilic clostridia, pNW33n plasmid that contains an origin of replication from *Geobacillus stearothermophilus* plasmid pTHT15 was suitable (chapter 3, 4 and 5). First, we tested two different plasmids, the *E. coli*-Bacillus shuttle vector pNW33n encoding a chloramphenicol acetyltransferase (Olson and Lynd, 2012) and the *E. coli-*Geobacillus shuttle vector pUCG3.8 encoding a thermostable kanamycin nucleotidyl transferase gene with origin of replication from *G. stearothermophilus* plasmid pBST1 (Bartosiak-Jentys, et al., 2013). These specific plasmids were selected based on two vital features: i) the origin of replication to propagate both in *E. coli* (cloning host) and *Clostridium* (host of interest), and ii) limited availability of antibiotic markers at elevated temperatures. The plasmid pUCG3.8 was unable to propagate due to the presence of several recognition sites or origin of replication not compatible for clostridia (chapter 4, Table 4.1), so the plasmid pNW33n was best suited for transformation.

In addition, access to both constitutive and inducible promoter systems were desired for (plasmid-based) gene (over)expression. A limited number of promoters and gene reporter systems have been tested in clostridia before. For example, promoters like 2638, 2926, 966 and 815 showed reliable expression when driving the expression of *adhB* and *lacZ* genes in *H. thermocellum* (Olson, et al., 2015). In this thesis we have managed to show activity of native
intergenic 16S/23S rRNA promoter from *H. thermocellum* for the expression of the chimeric sgRNA (*chapter 3*). The 16S/23S intergenic promoter sequence in *E. coli* has a role in the transcription of ribosomal RNAs (Mankin, et al., 1987, Zacharias and Wagner, 1989). Similarly, the characterized *xylL* promoter from *Bacillus smithii* allows the expression of ThermodCas9 protein but the *Bacillus coagulans pta* promoter does not work for unknown reasons in both *H. thermocellum* and *P. thermosuccinogenes* (Mougiakos, et al., 2017, Ganguly, et al., 2019). We also showed activity of native xylose promoters from *P. thermosuccinogenes* functional in *H. thermocellum* (*chapter 5*). For inducible promoters, only *celC* promoter has been characterized in *H. thermocellum* which is induced by laminaribiose. The induction levels are best between 1 and 2 hours and expression drops drastically after 3 hours (Mearls, et al., 2015). The above-mentioned xylose promoters can be also tested for inducible expression systems. Hence, these constitutive promoters are best available now for gene expression in thermophiles.

**Plasmid-based gene overexpression**

Plasmid-based gene overexpression involves both homologous and heterologous expression approaches. Homologous expression deals with the native chromosomal gene activity in a host from where it originates, whereas heterologous expression impart a foreign or non-native activity to the host strain. Plasmid-based gene overexpression strategies have been reported in several clostridial species. The cellulolytic *C. cellulolyticum* have been modified by heterologous overexpression of operons or genes from *E. coli, B. subtilis, Lactococcus lactis,* and *Zymomonas mobilis* to improve ethanol and isobutanol production (Guedon, et al., 2002, Higashide, et al., 2012). Similarly, we have expressed the xylose metabolism genes from *P. thermosuccinogenes* into *H. thermocellum* to enable its growth on xylose. This approach has increased ethanol and acetate production in *H. thermocellum* as described in the next section and in *chapter 5*.

**ClosTron or ThermoTargetron**

In contrast to plasmid-based overexpression strategies, gene knockouts can be performed in clostridia to abolish gene functionality. Gene knockouts can be achieved either by random mutagenesis or site-specific gene deletion. Random mutagenesis uses transposons generated *C. difficile* and *C. perfringens* mutants for toxin production (Awad and Rood, 1997), quorum sensing (Vidal, et al., 2009), and sporulation/germination studies (Cartman and Minton, 2010). Alternatively, site-specific gene knockouts were created in bacteria to study metabolic and
physiological pathways for industrial and medical applications. The established method for site-directed gene disruption in mesophilic clostridia was ClosTron or Targetron, a group II intron-based technology (Heap, et al., 2007, Heap, et al., 2010). This technology is based on gene targeting vectors derived from mobile Lactococcus lactis L1.ltrB group II intron. Their utility for gene targeting deals with ribozyme-based DNA integration mechanism, named as retro-homing, aided by a ribonucleoprotein (RNP) complex that has excised intron lariat RNA and an intron-encoded protein (IEP) with reverse transcriptase (RT) activity (Lambowitz and Zimmerly, 2011) as described in Figure 7.2a. Gene targeting using mesophilic and thermophilic group II introns (ThermoTargetron), have been established in several bacteria (Karberg, et al., 2001, Heap, et al., 2007, Heap, et al., 2010, Lambowitz and Zimmerly, 2011, Mohr, et al., 2013, Pyne, et al., 2014, Little, et al., 2018). Recently, the thermophilic cyanobacterium Thermosynechococcus elongatus group II introns, was characterized and applied for efficient chromosomal gene targeting in H. thermocellum at elevated temperatures (Mohr, et al., 2013). However, even with its proven utility in clostridia, group II intron mediated gene disruptions have certain limitations with the inability to have clean knockouts, target specificity is reduced with lack of IEP recognition sites and random intron insertion events (Table 7.2) (Mohr, et al., 2013, Song, et al., 2015, Bruder, et al., 2016, Joseph, et al., 2018). Due to these limitations it was not preferred for our organisms of interest, as we desired to develop a simple, precise, efficient and preferably high-throughput tool.

CRISPR-Cas system

The previously discussed techniques for genome engineering and transcriptional regulation purposes in chapter 2 are laborious, often inefficient and technically complex in the anaerobic bacteria, which are generally known to be challenging for their genetic manipulation. The advent of CRISPR-Cas system as a gene editing tool to manipulate clostridia is gaining attention. The type II CRISPR-Cas system spCas9 was applied in various clostridial species as a counter-selection tool for genome engineering (Bruder, et al., 2016, Joseph, et al., 2018). There is also an option to explore endogenous type I-B CRISPR-Cas systems. For example, successful genome editing was established in mesophilic and thermophilic clostridia using endogenous type I CRISPR-Cas systems (Pyne, et al., 2016). Yet, making the native CRISPR-Cas machinery functional is strain-specific and requires identifying the correct PAM sequences. Taking this into account we tried rather to apply the Type II spCas9 for H. thermocellum which has the specific ‘NGG’ as PAM. In previous studies, a temperature-controlled recombination/counterselection spCas9 tool was established in thermophile B. smithii for
genome editing purposes. The spCas9 activity was controlled by transformation at 45 °C, and growth at 37 °C so the non-edited cells are eliminated by counterselection (Mougiakos, et al., 2017). When applying the same approach for \textit{H. thermocellum}, we could not generate mutants probably due to the inactivity of mesophilic Cas9 at > 42 °C, and it could not be used for genetic engineering. Recently, three thermophilic Cas9 variants were characterized and referred to as GeoCas9, AceCas9, and ThermoCas9 (Harrington, et al., 2017, Mougiakos, et al., 2017, Tsui, et al., 2017), that could potentially be utilized for \textit{H. thermocellum}. We applied the active ThermoCas9 for the thermophiles, but it was impossible to clone the homology arms in \textit{E. coli}. The cloning issues is related to toxicity of the genes in \textit{E. coli} which can be tackled by using yeast as a cloning host. Some \textit{Clostridium} species have toxicity due to expression of activeCas9 with functional sgRNA, so a Cas9 nickase (Cas9n) variant was preferred (Li, et al., 2016, Xu, et al., 2017, Li, et al., 2019). Cas9n systems have either the HNH or RuvC nuclease domains inactivated with ssDNA breaks which are less lethal and increases the rate of homology directed repair (HDR) compared to dsDNA. Yet, the cell needs sufficient time to integrate the provided template or repair the break before cell death. The difficulty to get mutants could be due to low HDR efficiency in clostridia (Li, et al., 2016, Wang, et al., 2016, Walker, et al., 2020). Recently, Walker and colleagues showed for the first time the use of a Type II GeoCas9 system together with recombineering machinery improved genome editing efficiency to 94%, compared to without recombination system in \textit{H. thermocellum} (Walker, et al., 2020). However, low plasmid transformation efficiencies, inefficient HDR events, lack of recombineering and NHEJ tools posing hurdles for tool development in thermophilic clostridia, led us to apply the catalytically dead Cas9 (ThermodCas9) variant (Mougiakos, et al., 2017) as alternative. The ThermodCas9 targets a 20 nt spacer sequence in genome by a guide RNA and blocks transcription as depicted in Figure 7.2b. CRISPRi has enabled simple, reversible gene knockdown at a transcriptional level with no modifications in the genome (Qi, et al., 2013). CRISPRi has been used previously for gene knockdowns for both native and heterologous genes in mesophilic clostridia (Bruder, et al., 2016, Wen, et al., 2017). As a first attempt we applied CRISPRi as a silencing tool in \textit{H. thermocellum} DSM1313 (Ganguly, et al., 2019) and \textit{P. thermosuccinogenes} DSM5809 (Figure 7.1, dCas9 based CRISPR technology (\textbf{chapter 3} and \textbf{4})). This resulted in redistribution of the organic acids and ethanol production in the strains which will be discussed in the next section.

CRISPRi-based gene regulation can be used for multiplex silencing. This can be achieved by using a dual RNA (crRNA/tracrRNA) approach for a rapid CRISPR-array assembly method
represented as CRISPathBrick (Cress, et al., 2015) or applying a DNase-dead Cpf1 (ddCpf1) variant to regulate gene expression as exhibited in other studies (Zhang, et al., 2017). These tools could enable multiplex silencing in non-model organisms with limited genetic toolboxes. Although, CRISPRi works effectively in some clostridial species, it would be interesting to apply a transcriptional activator coupled to dCas9 to improve the transcription of specific genes. Gene knockdowns previously have been established on the translational level using antisense RNA (asRNA) in Clostridium. The asRNA knockdown includes targeting mRNA transcript using its asRNA to characterize various genes including essential ones where gene knockout is not feasible (Tummala, et al., 2003, Sillers, et al., 2009). However, asRNA needs large constructs for gene silencing whereas CRISPRi has repression specificity with a 20 nt sgRNA (Table 7.2). CRISPRi may have disadvantages such as reduced expression of downstream or sometimes upstream genes in operons, that may be transcript or organism-specific (Peters, et al., 2015). Another limitation for selective pressure of silenced transformants necessary for growth, results in suppressor mutations that inactivate the CRISPR system. We observed also a temporary silencing effect on ldh and pta genes at stationary phase and this CRISPRi loss-of-function might be due to the presence of suppressor mutations in the dead Cas9 coding sequence (Zhao, et al., 2016, Liu, et al., 2017). It has been previously shown that even a single mutation in the seed region of the sequence decreased repression intensely for both type I and II CRISPR systems (Jinek, et al., 2012, Qi, et al., 2013). Further studies in our case is required to demonstrate that this is the case. During the silencing experiments we observed insertion elements (IS) in CRISPRi plasmids in E. coli DH10β. Sequencing results showed that the insertion (fragment size 777bp) was in the middle of the thermodcas9 gene. A 9 bp sequence motif (5’-AGGAAACCG-3’) found on either side of the transposable element seems to be the target sequence by the transposase. How this transposable element was constantly inserted in all plasmids including non-targeting at the same position is still unclear. The template used to make the ThermodCas9 fragment was sequenced and did not have the insert. Moreover, this insertion was only observed when plasmids were cloned in E. coli DH10β. The plausible explanations would be systematic recognition of the transposase to a specific motif (5’-AGGAAACCG-3’) is occurring. This motif was also found in the plasmid sequence at another site in ThermodCas9, but that was not targeted. We can thus rule out the systematic recognition possibility as the cause of the transposition. If there is indeed recognition of a specific motif, it is also possible to mutate those specific sequences, so no future transposition can occur. The other explanation is selective pressure that may induce the transposable element to insert at that locus. The insertion in the middle of the dcas9 gene would act as a knock-out. This in turn
would negate the potential silencing effect on the desired genes. However, if selective pressure is the cause for the transposable element targeting the same locus, then why was the second motif not targeted, since it is also present within the ThermodCas9 sequence. Though further studies are desired to find the cause of such an event using CRISPRi systems, we have at least solved the IS element problem by using different *E. coli* DH5α and TG90 strains to clone the plasmids. Hence, the CRISPRi genetic system developed here demonstrates a step forward towards manipulation of thermophilic strains.

![Figure 7.2](image)

**Figure 7.2.** a) ThermoTargetron, Targetron plasmid with a group II intron RNA and reverse transcriptase (RT). The intron RNA is expressed with exon sequences (E1 and E2) and is spliced by intron-encoded RT to form a ribonucleoprotein (RNP) complex. The RNPs recognize DNA target sites primarily by base pairing of the intron RNA to the DNA. Subsequently, the intron RNA is inserted by reverse splicing into one strand of the DNA. The reverse transcriptase protein cleaves the opposite strand and synthesizes the second strand. The cDNA is integrated into the genomic DNA by host DNA repair mechanisms, disrupting the target gene (Hong, et al., 2014). b) CRISPR interference system, the catalytically inactive Cas9 (dCas9) recognizes and binds the target DNA and the Cas9 protein physically interferes with the RNA polymerase blocking transcription (Qi, et al., 2013).
Table 7.2. Overview on genetic tools applied for gene knockout, knockdown, and genome editing strategies in *Clostridium*

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antisense RNA</strong></td>
<td>Gene knockdown</td>
<td>reversible and tunable</td>
</tr>
<tr>
<td><strong>Transposon mutagenesis</strong></td>
<td>Random gene disruption</td>
<td>mutant library construction</td>
</tr>
<tr>
<td><strong>ClosTron or ThermoTargetron</strong></td>
<td>Targeted gene disruption and construction</td>
<td>automated, streamlined targeting and construction</td>
</tr>
<tr>
<td><strong>Allele-coupled exchange</strong></td>
<td>Gene insertion and replacement</td>
<td>- broadly applicable</td>
</tr>
<tr>
<td><strong>Heterologous counter-selective integration</strong></td>
<td>Gene insertion and replacement</td>
<td>- broadly applicable</td>
</tr>
<tr>
<td><strong>Recombineering</strong></td>
<td>Point mutations</td>
<td>simple procedure</td>
</tr>
<tr>
<td><strong>CRISPR-Cas</strong></td>
<td>Gene editing and silencing</td>
<td>- broadly applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- efficient and scarless</td>
</tr>
</tbody>
</table>

*RAM: retro-transposition-activated marker*
APPLICATION OF GENETIC TOOLS TO MODULATE METABOLIC GENES

The development of genetic tools for thermophilic clostridia to silence metabolic genes have a substantial effect on product formation (chapter 3 and 4). Metabolism of the thermophiles apart from silencing are also affected by overexpression of genes. This section focuses on the influence of various aspects such as metabolic gene repression or plasmid-based gene expression on clostridial metabolism of two non-model thermophiles.

Lactate dehydrogenase and phosphotransacetylase knockdown in *H. thermocellum*

Fermentation end products of wild-type *H. thermocellum* DSM1313 on cellobiose are ethanol, acetate, lactate and formate. The strains have been genetically engineered to enhance ethanol production by disruption or elimination of the competing pathways for carbon and electron flux, respectively (Argyros, et al., 2011, Biswas, et al., 2014, Biswas, et al., 2015, Papanek, et al., 2015, Rydzak, et al., 2015, Kannuchamy, et al., 2016, Lo, et al., 2017).

For our attempts to exploit the CRISPRi tool in thermophilic clostridia, we targeted the metabolic genes *pta* and *ldh* in *H. thermocellum* as examples to modulate gene expression. It has been verified that the deletions of these genes are not detrimental for its growth and enhances ethanol production (Argyros, et al., 2011, Biswas, et al., 2014, Papanek, et al., 2015). The results show that when the *pta* was targeted for gene silencing, the organic acids production was affected. The acetate production reduced by around 27%, lactate increased to 30%, formate was improved and ethanol had no significant difference. Agyros *et.al.* have shown similar impact of improved lactate titer with *pta* knockout (Argyros, et al., 2011). Other studies with *pta* knockout mutants had decreased acetate and lactate production, respectively with increased ethanol (Mohr, et al., 2013). Regardless of the silencing, there is some residual gene expression ensuing enough enzyme activity for the cell to produce acetate and ATP instead of ethanol and NAD(P). Moreover, to balance the fluxes, the other metabolic pathways will be affected (Figure 7.3). Further engineering of pyruvate metabolism is required to enhance ethanol production (Tian, et al., 2016, Dash, et al., 2017). Similarly, for the *ldh* silencing transformants targeting the promoter had downregulation of gene expression with decreased lactate production. Earlier studies with *ldh* knockout depicted 4% decrease in lactate production in comparison to wild-type (Mohr, et al., 2013). In our case, we showed 24% decline in lactate and 15% improvement
in formate, however no increase in acetate or ethanol was observed. In another study, the \textit{ldh} mutants of \textit{H. thermocellum} showed no changes in acetate titer. The \textit{ldh} gene deletion therefore does not affect the flux through the pyruvate node if ethanol is produced (van der Veen, et al., 2013). Moreover, no significant differences were observed for ethanol when compared to the non-targeting control. Hence, we have repressive effect for both \textit{pta} and \textit{ldh} genes with reduction in acetate and lactate respectively. With the successful development of the CRISPRi tool, it would be worthy to apply it under an inducible promoter silencing system for pyruvate formate lyase, acetate kinase and other genes in the central metabolism. This can be combined with multiplex silencing to drive the flux towards ethanol production.

**Malic enzyme knockdown in \textit{P. thermosuccinogenes}**

CRISPRi offers reduction of gene expression in thermophilic clostridia as demonstrated in \textit{H. thermocellum}, hence we applied the same system in \textit{P. thermosuccinogenes}. Recent studies have explored CRISPRi to control the metabolic pathways in \textit{E. coli} and other non-model organisms for improved production of numerous biotechnological products (Mougiakos, et al., 2018). We observed with downregulation of the malic enzyme gene expression that the lactate production is increased with a concomitant decline in acetate production. Due to less pyruvate accumulation, the cell directs towards lactate to balance the flux. We have expected effects in malate or pyruvate production, but no significant variations were observed. In addition, prominent fluctuations in other organic acids or ethanol were also not noticed. Though the reason behind the impact on lactate and acetate is unclear, it would be beneficial to knockdown these genes to understand the metabolism. We hypothesized that the malic enzyme gene knockdown would enhance succinic acid production by accumulation of malate, but there was no significant change observed. Similar studies in \textit{H. thermocellum} with deleted malic enzyme and lactate dehydrogenase genes improved ethanol and formate titers (Deng, et al., 2013). So, it would be beneficial to knockout or knockdown of lactate dehydrogenase or malate shunt pathway genes in \textit{P. thermosuccinogenes} to improve the succinate production.

**Heterologous expression of xylose metabolization genes in \textit{H. thermocellum}**

The capacity to co-ferment both C5 and C6 sugars by \textit{H. thermocellum} without additional strains can aid in developing it into a suitable CBP host. This can reduce feedstock costs, improve biomass consumption and allow the economic production of bio-based products. The key aspects of CBP relevant to thermophilic bacteria are: 1) Biomass solubilization with the unique capability of \textit{H. thermocellum}’s cellulosome to efficiently digest cellulose from plant
biomass and to hydrolyze both hemicellulose and cellulose compared to free cellulases. Extensive studies highlighted \textit{H. thermocellum} as an attractive candidate for CBP by achieving high solubilization yields from cellulosic biomass, ii) Bioenergetics of \textit{H. thermocellum} with secretion of biomass-solubilizing enzymes is energy intensive conserving more ATP or GTP than other fermenting organisms. Though quantitative understanding of the ATP or GTP yield is not complete, but the cellulose-specific bioenergetics benefits are higher than costs of cellulase synthesis and iii) Metabolic engineering with targeted manipulation strategies and optimization of fermentation conditions are needed for improving the feasibility of CBP as a production platform (Johnson, et al., 1982, Lu, et al., 2006, Lynd, 2016, Lynd, et al., 2017). Earlier studies have engineered \textit{H. thermocellum} by integration of xylose utilization genes from \textit{Thermoanaerobacter ethanolicus}. The engineered strain has doubled H\textsubscript{2} and ethanol production with xylose and avicel co-consumption than avicel only. This engineered strain also co-utilizes xylo-oligomers and xylose (Xiong, et al., 2018). Similarly, we created a \textit{H. thermocellum} strain that can consume xylose by plasmid-based xylose genes expression; for efficient breakdown of both C6 and C5 sugars from lignocellulosic feedstocks. So, the best option we had was the GTP-dependent xylose gene cluster xylose isomerase and xylulokinase present in \textit{P. thermosuccinogenes} (Koendjibiharie, et al., 2018). Compared to other studies, the xylose transformants were characterized based on their product profile both in anaerobic bottles and fermenter compared to cellobiose grown cultures. The xylose grown strains had an initial lag phase which might be due to the lack of essential factors compared to the native host. Ethanol and acetate production in both anaerobic bottles and fermenters were increased in \textit{H. thermocellum} (chapter 5). Lactate and formate production were reduced. The co-consumption of both pentose and hexose sugars for these xylose transformants are further required to improve ethanol production.
Figure 7.3. Fermentative pathways of thermophilic clostridia for conversion of cellobiose or glucose to ethanol and major organic acids production. (ACK - acetate kinase, AldH - aldehyde dehydrogenase; ADH - alcohol dehydrogenase; FH - fumarate hydratase; FR - fumarate reductase; LDH - lactate dehydrogenase; MDH - malate dehydrogenase; ME - malic enzyme; PEPCk - phosphoenolpyruvate carboxykinase; PFL - pyruvate-formate lyase; PTA - phosphotransacetylase. Adapted from (Poudel, et al., 2017, Koendjbiharie, et al., 2018, Xiong, et al., 2018).

**PERSPECTIVES**

To improve microbial fermentation for sustainable production of green chemicals from renewable resources, thermophiles *H. thermocellum* and *P. thermosuccinogenes* are being explored. The development of thermophiles as platform organisms require genetic accessibility and engineering tools as well as understanding of their metabolism to improve fermentation efficiency. The present research focuses on the strategies to genetically manipulate thermophilic clostridial strains. The major emphasis is on how to improve the transformation efficiencies of these organisms for high-throughput tool development. The idea would be to tackle the other RM systems with similar strategies applied for type I restriction system. Moreover, knockouts...
or knock-down of specific genes in RM systems can be done. Thorough methylome analysis is desirable and new in silico tools should be implemented. For example, SyngenicDNA, that identify precise targets of the RM systems within an intractable bacterium, remove these targets from the DNA sequence and synthesize a tailor-made version suited for that specific host. Also SyngenicDNA Minicircle Plasmid (SyMPL) tools, free from components essential for propagation in E. coli, but not needed in the specific host and thereby assisting less burden to the cell (Johnston, et al., 2019). These approaches can ease the process towards genetic tool development. Although substantial efforts have been made for developing CRISPR-Cas systems, more understanding of the basic biology and the endogenous systems of the CRISPR-Cas within clostridia are required to advance genome editing technologies. CRISPR-Cas systems as genetic tools using different variants (i.e., Cas9, Cas9n, dCas9, Cpf1/Cas12a, and endogenous systems) for genetic manipulation and regulatory control have been established in mesophilic clostridia. Nonetheless these tools require tight regulation under strong promoters to overcome any toxic side effects. Unfortunately, for clostridia, universal CRISPR-Cas system is not available as not all promoters or plasmid replicons will be functional in every organism. However, with the advent of “RiboCas” applicable to diverse clostridial species by regulating Cas9 expression at translation level, it can be expanded further as precise and high-throughput tools (Canadas, et al., 2019). Still, there is a large gap in understanding the physiology and availability of genetic tools compared to model organisms E. coli or S. cerevisiae. Moreover, optimizing their production capacities for a better platform organism is still not explored for these thermophiles. Hence, the necessary step is gaining further knowledge on the metabolic pathways/enzymes of the production hosts with in silico metabolic simulation, together with integration of omics for better engineering strategies. Nevertheless, the start has been made and with high-throughput engineering and automation approaches, these organisms can be considered as potential production platforms.
Bibliography


Bibliography


Bibliography


Stout, E., Klaenhammer, T., and Barrangou, R. (2017) CRISPR-Cas technologies and applications in food bacteria, Annual Review of Food Science and Technology 8: 413-437.


Acknowledgments

As the saying goes ‘There's always a light at the end of the tunnel’, after 4.5 years of the PhD journey I did see the light! This journey was challenging from day one but eventually I learned enormously during the course and it became a life-changing experience. Most importantly, during this journey I crossed paths with some amazing people who have played significant roles in my life. So, I would like to express my gratitude to each one of you in this dedicated section.

Richard, I still remember the first conversation we had, you mentioned about the challenges in this project and advised me that ‘patience’ is required for these four years. I am glad that I could keep my patience till the end and we managed the project so well. Thank you for motivating me during the tough times as well as rejoicing on our successes. You were constantly there to listen out and gave me the freedom to explore my ideas. I always enjoyed our conversations and brainstorming sessions. It has been fun to spend time with you not only in trips or events but also the car rides to Corbion where we discussed on both scientific and non-scientific topics. Just to add, many research ideas formulated in those car rides have culminated as chapters in my thesis. Thanks for those incredible four and half years where I learned a lot from you and helped me evolve as a researcher over time.

John, thanks for the opportunity to be a guest PhD student in Bacterial Genetics group and glad that you made me and Jeroen part of all the activities. I am grateful for the suggestions or ideas you gave on the project and as well, checking on me over time including the stressful writing phase. I could always drop by your office to discuss and ask advice about anything. It has been a great experience to work in this lab and being part of such a nice group.

Serve, thank you for the guidance and tips on handling clostridia. It was great talking to you and have always helped me whenever needed. Raymond, thanks for the coffee corner scientific discussions and suggestions. Though my interactions with you have been limited but I enjoyed our good conversations.

Claudine and Paty (Patricia), my two pillars that stood beside me through thick and thin. Its feels surreal that VLAG PhD week gave me two friends for life, which I could have never imagined. Thanks girls for all the care and love. Claudine, my ‘Special tea’ friend, I am amazed how well you understand me and always been there for me at every step of my PhD journey. Thank you for being there during my happy times but also being a shoulder to lean on in difficult phases. I am glad to have you as my paranymph in the last stretch of the journey. Paty, my ‘Brigadeiro’ friend, I connected with you instantly when we met and enjoyed our nice chats. Thanks for listening and taking care of me even when not in Wageningen. I adored that you always made me part of all the parties, activities and events organized by you. You both are my constants and can always count on you for anything and everything.

Prarthana, the first time we met in Dreijen, hardly did I know that will connect with you so well. Thank you for being the friend with whom I can discuss on various topics and bringing new insights into it. It has been
always fun hanging out with you and thanks that you agreed so easily to my random plans. I am glad that we could share so many great lunches, coffee-breaks, late-evening lab work and of course the endless laughter sessions. Ashwin, thanks for the amazing jokes and keep inspiring her with your fantastic cooking. Prarthana, I am happy to finish my PhD with having you beside me as a paranymp.

Jeroen, it’s hard to believe that we have accomplished 4 years of our PhD journey which looked impossible in the beginning. We have learned a lot in the process and helped us to connect better with time. I am grateful to you for all the suggestions and help. Thanks for the support and best wishes for your future endeavours.

This PhD study would not have been possible without the efforts and immense dedication of all my students. Alkan, Bram, Cairn, Ivo, Maria, Mark, Richard and Teun. I enjoyed our working together in the anaerobic tent, brainstorming sessions, fruitful discussions and team spirit that lead to successful completion of this project. Thanks for your inputs and hard work in this challenging project. I am delighted in mentoring you all and learned through the process. I am happy to know that many of you now are pursuing your careers either in academia or industry and I wish you success in life.

It was a great experience to be part of Corbion and grateful to all my colleagues for their immense support. Bastienne and Rudy thanks for the bioinformatics analyses and valuable discussions. Mariska and Judith, I appreciate all the help and co-operation received when I was conducting experiments there. Koendert, Miranda and Zhen, thanks for the nice conversations. Remco and Victoria, thanks for the assistance with the flow cytometry. I feel privileged that my research work was financially supported by Corbion.

Athina and Diego from EMC Rotterdam, it was a great experience to collaborate with you on the MinIon methylation project and thanks for your endless support. Tiny from WEMC facility, thanks for the amazing TEM pictures of Pseudoclostridium thermosuccinogenes. Tjakko and Marcel from Food Microbiology, it was a nice collaboration with you on the sporulation heterogeneity project and thanks for the valuable inputs.

Anja and Heide, thanks a lot for your timely help on administrative stuffs even during corona crisis. The technicians team Guus, Jame, Ineke, Merlijn, Monika, Philippe, Rob, Sjon, Steven, Tom, Tom and Ton, thanks for the organization and support to ensure that the labs are functioning smoothly and efficiently.

Thanks, Adriano, Amaury, Ana, Carolina, Daphne, Eleonora, Elodie, Inès, Isabelle, Jeroen, Kristian, Liz, Mamou, Maria, Marjon, Matt, Miia, Nigel, Raquel, Richard, Rubab, Sasha and Tom, for all the exciting times during CLOSPORE meetings and workshops. Aleksandr, Erika, Hugo, Ioannis, Jeroen, Johanna, Prarthana and Wen, it was an amazing experience with you all organizing the MIB SSB PhD trip (2017). Thanks Lab trip organization buddies Emmy, Linde and Martijn for the teamwork and fun times.

I am grateful to all my colleagues working at MIB and SSB for the great times in the last 4.5 years, that made my PhD experience more exciting and valuable. Thanks, Becca and Jochem, it has been great fun catching up with you both and looking forward for our visit to the Indian Crepes place in Amsterdam. Costas, Eric and James, the Clostridium team, I always enjoyed our conversations and happy that you all are managing
well with stubborn clostridia. Diana, thanks for the coffee corner conversations and suggestions. Hanne, thank you for the scientific discussions and enjoyable chats. Jame, Monika, Steven and Ton thanks for all the help with the anaerobic stuffs. Linde, it has been pleasant talking to you in the corridors about both scientific and non-scientific topics and thanks for making me part of your wedding party. Maria, my first MSc student, I really admired your motivation and perseverance. It was so easy working with you and always enjoyed our conversations. I am glad we continue to share such an amazing bond. I am extremely happy for you for starting your PhD in SSB and wish you all the luck! Mamou, you related the most to my clostridium challenges. Thanks for the scientific conversations and suggestions that helped me in the project. All the very best for the last phase of finishing up your PhD. Nam, thanks for the suggestions about dealing with the writing phase. Thanks Nikolas, Micphyser friend for the support and we still need to plan the Bollywood movie night for both the Greek and Indian MIBers. Ran, thanks for the nice chats and good wishes for finishing up your PhD. Rob, thanks for all the help and arrangements in the lab. Thank you, Tom for the fun-filled conversations.

To my current and former office mates, Costas, Eric, Gerben, Ioannis, Jeroen, Mamou, Mihris, Romy, Tijn and Tom, thanks for the good times we shared together. Many thanks to Adini, Belen, Brenda, Carina, Catarina, Despoina, Elleke, Enrico, Franklin, Ismael, Janneke, Joep, Jorrit, Jurrie, Lonne, Lorenzo, Max, Marnix, Melvin, Sebastian, Sjoerd, Stan, Stijn, Thijs, Thomas, Tim, Yifan and all the other past and present BacGen colleagues for creating such a great atmosphere in the group. I would also like to thank, Caifang, Caroline, Daan, Detmer, Emma, Enrique, Giannis, Irene, Ivette, Lot, Luis, Lyon, Menia, Nico, Nong, Peer, Sudarshan, Yuan and everyone else from MIB and SSB.

I would also like to extend my gratitude to all the I-scream group members for the lovely dinners and exciting activities. Alex and Carrie, thanks for all the nice conversations. Dorett, it’s been pleasure talking to you and enjoyed our discussions about Heidelberg. Emmy, I am glad to find a good friend during PhD trip for listening out and motivating me towards a sustainable lifestyle. Mark, thanks for the interesting chats and I am amazed by your collection of plants. Thanks, Nico for the nice discussions and I am happy that you are back in MIB. Prarthana and Ashwin, thanks for the entertaining chats. Ruben, thanks for the great discussions on patents and informing me about the Wageningen board games club. Teunke, thanks for being so thoughtful and understanding. It was great fun making Prarthana’s video together. Wen, I am inspired by your enthusiasm and hope you are one day going to cook for me the special Chinese dish with eggs.

A good support system is vital to survive and stay sane during the PhD journey. I was lucky to have great friends in Wageningen that were always encouraging and motivating me. Elvira, thanks for the life discussions we had somewhere always in Utrecht and agreeing to most of my impromptu plans. Enrico, it’s been always fun with you to have those endless conversations on science and our struggles in research. Francesca (Fra), my Italian ex housemate, thanks for all the Italian nice food you cooked for me and a listening ear to my whining about life. Fra and Jurij, thanks for enlightening me about the history and politics of Italy and Slovenia. Juveli, it has been pleasant talking to you and appreciate your presence in my life. Koko (Konstantina) and Leire, it’s been great talking to you girls about life and most importantly life after
PhD. Thanks, Lari (Andrea), my housemate, for being so supportive during my writing phase and helping me with R scripting. Lerato, I am inspired with your candor and positive outlook. Marco, I am amazed by your calmness and enjoyed every bit of making Claudine’s video together. I am happy for you and Elisa for embracing the parenthood phase. Raisa, my DJ ‘Chole’ friend, you stun me with your management skills by handling so many events and doing a PhD as well. I have enjoyed to the core all the parties you have invited me. Ruth, thank you for listening out to me always and giving me advice on different situations. Sema, I see you like my little sister and glad that we connect on so many ways. Souvik, thanks for making all the nice chicken curries and to our discussions plus endless laughter sessions!

I feel so privileged to have a small bong family in Eindhoven away from home. Thanks Keya di and Rajesh da for hosting me at your place every year during Durga Puja. It has been great spending time with you both. Thanks, Abanti di, Aditi di, Nandini di, Rani di, Rita di, Sharmistha di, Sriparna di, Sudipta di and Trishita di, it was fun organizing the event with you all. I am grateful to each one of you for the entertaining conversations and making me part of the family.

It has been quite a journey moving from one country to another in Europe and making good friends through the course. I am fortunate to have a set of friends in Germany that I can continually rely on. Anshu di, thanks for all the help and guidance you provided me. I am glad that we shared together so many fun conversations and trips. Karan, ‘Halle baba’, thanks for being so supportive and being there for me. I am inspired by your positivity towards life and enjoyed all our long discussions, outings and reunions. I am so happy for you and Heta (your perfect addition in life), thank you for making me part of your beautiful Gujrati wedding. Shah (Abhijeet), ‘Thane friend’ thanks for providing me with your immense support and listening out to my whining sessions. I am so grateful to you for continuously motivating and cheering me up. Karan and Shah, my support system, I can always count on you both in good as well as bad times.

Ankita, thanks for meeting up in Amsterdam and reminiscing the golden days of Vizag. Apu (Aparupa), it feels great that our friendship and connection has evolved so beautifully with time. Thanks for your love and endless support. Harika and Dhanush thanks for such a warm welcome in San Jose and enjoyed spending time with you both. Thanks, Ipi (Ipsita), Geet, Paboni (Pavani) and Ponda (Jyoti) for the lifetime memories in Hyderabad that I will cherish forever. Madhuri, it was great fun meeting you in Amsterdam after seven years, we should plan these reunions more often. Manisha di, thanks for hosting me in New York and catching up after so many years from where we left. Rajiv, it’s so strange that we studied together but became such good friends after we moved to Europe. It’s been amazing discussing with you about the fun as well as serious experiences in life. Ridhi, my doctor pal, I realized we are friends for fifteen years and it’s hard to believe. I connect with you the most and sometimes you understand me better than myself. Thanks for being there for me. Sharan (Abhishek), thanks for the constant support and understanding me always!

Didibhai (Pial), I appreciate you for being my strict sister and protecting me always. Thanks for taking care of me and cooking all the amazing egg curry and ‘dhoka dalna’. Subrata jiju, you will be always in our thoughts. Subho (Partha), you are an awesome nephew and I enjoy teasing you a lot.
Chordibhai (Bratati), thanks for pampering me the most. I am glad that can discuss life with you as a friend rather than sister. Shrikant jijis, thanks for your support and a great addition to the family. Aadrika, you are a cutie pie and I hope to meet you soon.

বাপি, তেমার সমর্ভন এবং অবিরাম পরিপ্রমেয়া জন্ম ধনবাদ যা আমাকে আমার বংশীলি অর্জনে অনুপ্রাণিত করেছে। আমি বর্ধিত যে আমি তেমার মত চ্যালেঞ্জ প্রহর এবং সেগুলি পরামর্শ করতে লিখেছি। মা, আমার মা, তেমার বহু হওয়ার জন্ম ধনবাদ। আমি তেমার কাছ থেকে জীবনে খোঁজ এবং অধ্যয়ন করতে লিখেছি। ধনবাদ, মা এবং বাপি তেমার নিশ্চিত ভালবাসা এবং সমর্ভনের জন্ম আমার PhD বিষয়ের সকল সমর্শিতে সহায়তা করেছে।
Overview of completed training activities

Discipline specific activities

- Gene Road Map Training, University of Nottingham, Nottingham, UK (2015)
- RNAseq Workshop, Institut Pasteur, Paris, FR (2016)
- CLOSPORE Workshop, Corbion, Gorinchem, NL (2016)
- Clostridia workshop on Cell and Developmental Biology, NOVA University Lisbon, Oeiras, PT (2017)
- Workshop on Biorefineries, WUR-FBR, Wageningen, NL (2018)
- Training in Translational Skills, NIZO, Ede, NL (2018)
- Microbial Physiology and Fermentation, TU-Delft, Delft, NL (2019)

Meetings and Conferences

- Prokagenomics conference, Gottingen University, Gottingen, DE (2017)
- Microbiology Centennial Symposium, WUR-Laboratory of Microbiology, Wageningen, NL (2017)
- Netherlands Biotechnology Congress 19, Ede, NL (2019)
- Genome Engineering: Frontiers of CRISPR/Cas, Cold Harbor Spring laboratory, New York, USA (2019)

*Poster presentation, **Oral presentation

General Courses

- VLAG PhD week, VLAG, Soest, NL (2016)
- Competence assessment, WGS, Wageningen, NL (2016)
- Basic IP workshop, Corbion, Gorinchem, NL (2016)
- Mid-term progress review, European Commission, Brussels, BE (2016)
- Essentials of Scientific Writing and Presenting, WGS, Wageningen, NL (2016)
- 7-Habits of Effective people, Corbion, Gorinchem, NL (2017)
- Project and Time Management, WGS, Wageningen, NL (2018)
- Critical thinking and argumentation, WGS, Wageningen, NL (2019)
- Career Perspectives, WGS, Wageningen, NL (2019)

Optionals

- Research proposal, WUR, Wageningen, NL (2015)
- Bacterial Genetics group meetings (weekly)
- Microbiology PhD meetings (monthly)
- Organizing committee, PhD study trip to Hamburg, Copenhagen and Lund
Overview of completed training activities

Discipline specific activities

Courses
- Gene Road Map Training, University of Nottingham, Nottingham, UK (2015)
- RNAseq Workshop, Institut Pasteur, Paris, FR (2016)
- CLOSPORE Workshop, Corbion, Gorinchem, NL (2016)
- Clostridia workshop on Cell and Developmental Biology, NOVA University Lisbon, Oeiras, PT (2017) **
- Workshop on Biorefineries, WUR-FBR, Wageningen, NL (2018) **
- Training in Translational Skills, NIZO, Ede, NL (2018)
- Microbial Physiology and Fermentation, TU-Delft, Delft, NL (2019)

Meetings and Conferences
- 7th European Spore conference, University of London, London, UK (2016) *
- Prokagenomics conference, Gottingen University, Gottingen, DE (2017) *
- Microbiology Centennial Symposium, WUR-Laboratory of Microbiology, Wageningen, NL (2017) *
- Netherlands Biotechnology Congress 19, Ede, NL (2019) *
- Genome Engineering: Frontiers of CRISPR/Cas, Cold Harbor Spring laboratory, New York, USA (2019) *

*Poster presentation, **Oral presentation

General Courses
- VLAG PhD week, VLAG, Soest, NL (2016)
- Competence assessment, WGS, Wageningen, NL (2016)
- Basic IP workshop, Corbion, Gorinchem, NL (2016)
- Mid-term progress review, European Commission, Brussels, BE (2016)
- Essentials of Scientific Writing and Presenting, WGS, Wageningen, NL (2016)
- 7-Habits of Effective people, Corbion, Gorinchem, NL (2017)
- Project and Time Management, WGS, Wageningen, NL (2018)
- Critical thinking and argumentation, WGS, Wageningen, NL (2019)
- Career Perspectives, WGS, Wageningen, NL (2019)

Optionals
- Research proposal, WUR, Wageningen, NL (2015)
- Bacterial Genetics group meetings (weekly)
- Microbiology PhD meetings (monthly)
- Organizing committee, PhD study trip to Hamburg, Copenhagen and Lund
List of Publications


* Authors contributed equally
About the author

Joyshree Ganguly was born on the 12th of October 1988 in Asansol, West Bengal, India. In 2007, she started her BTech Biotechnology at ICFAI University, Dehradun. Her BTech thesis was performed at JK Seeds, Hyderabad, where she looked into phenotypic and genotypic segregation pattern of a plant selectable marker in transgenic rice.

In 2012, after completion of her Bachelor’s degree, she moved to Halle, Germany to study MSc Pharmaceutical Biotechnology at Martin Luther University Halle-Wittenberg. Her MSc thesis was performed in the Downstream Processing group at Institute of Pharmazie, under the supervision of Prof. Markus Pietzsch and Ramona Konrad. Her thesis was focused on metabolic engineering approaches to increase the production of recombinant proteins in engineered Escherichia coli. She continued there as a Research assistant for a year to work on Fed-batch fermentation to improve the yield of products of interest.

In 2015, she moved to Wageningen, The Netherlands to start her PhD project at Corbion under the supervision of Prof. Richard van Kranenburg. The project was part of the EU funded Marie Skłodowska-Curie Innovative Training Network called CLOSPOR. This research was conducted in the Bacterial Genetics group of Prof. John van der Oost, at the laboratory of Microbiology at Wageningen University and Research. During this four year project, her work focused on the the genetic accessibility and CRISPRi-mediated silencing of central metabolism for thermophilic clostridia. Most of the results has been described in this thesis.

About the cover

The background of the cover is made up of spores and vegetative cells of the thermophilic clostridia, Pseudoclostridium thermosuccinogenes. Spores are metabolically inactive structures, that are formed in response to adverse conditions and have been studied in this thesis. On the front, inside of a clostridial cell illustration represents the genetic parts with the brown random pattern that signifies genomic DNA, the black circle depicts CRISPRi plasmid DNA and the yellow plus purple structures indicate native restriction and methylation systems, respectively. These genetic parts are vital to overcome the restriction barriers by in vivo methylation of plasmid DNA to create a genetically accessible strain. After escaping the clostridial restriction systems, the CRISPRi plasmid as a silencing tool was applied to modulate the expression of metabolic genes to study the impact on the production of succinic acid and ethanol. These products are represented as the two molecules coming outside the cell.
The research described in this thesis was financially supported by the European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068] and Corbion, The Netherlands.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Iliana Boshoven-Gkini.

Printed by GVO-drukkers & vormgevers B.V Ede.
The research described in this thesis was financially supported by the European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068] and Corbion, The Netherlands.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Iliana Boshoven-Gkini.

Printed by GVO-drukkers & vormgevers B.V Ede.