UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FUNCTIONAL GENOMICS STUDIES OF CLOSTRIDIUM CELLULOLYTICUM FOR LIGNOCELLULOSE BIOCONVERSION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

TAO XU
Norman, Oklahoma
2016

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/46946

FUNCTIONAL GENOMICS STUDIES OF CLOSTRIDIUM CELLULOLYTICUM FOR LIGNOCELLULOSE BIOCONVERSION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Jizhong Zhou, Chair

_____
Dr. Charles Rice

_____
Dr. Laura Bartley

_____
Dr. Ralph Tanner

_____
Dr. Paul Lawson

© Copyright by TAO XU 2016
All Rights Reserved.

Acknowledgements

There are many people without whom this dissertation would not have been possible. I mention them here to express my sincere gratitude for their help in so many ways.

My deepest gratitude is to my advisor, Dr. Jizhong Zhou for his patience, diligence, insight, and motivation throughout my time at OU. Along the way, I really appreciated the financial support and the freedom he gave in my study such that I was equipped with cutting-edge skills and knowledge in both fundamental and applied microbiology. I would also like to thank my committee members, Dr. Laura Bartley, Dr. Charles Rice, Dr. Paul Lawson, and Dr. Ralph Tanner for their guidance and insightful comments at different stages of my research. I am thankful to Dr. Bartley for her enthusiasm and practical advice on my study, and for inviting me to contribute a book chapter. I am also grateful to Dr. Paul Lawson and Dr. Ralph Tanner for their advice on my career path.

I would also like to thank current and former members of the Zhou research group. They helped me in different ways and I learned so much from them. In particular, I want to thank Dr. Yongchao Li for teaching me anaerobic manipulation skills and helping with trouble shootings in my study, Dr. Zhili He and Dr. Joy Van Nostrand for their patient guidance on DNA microarray technology and great assistance in manuscript modification. I also want to thank Dr. Megan Kempher, Dr. Bo Wu, Zhou Shi, and Xuanyu Tao for their efforts in part of my research and helpful discussion. Additionally, I would like to thank all other lab members for being great colleagues and friends.

Special thanks go to my parents for their continued encouragement and quiet support throughout my intellectual journey. I am also grateful to my Mother and Father in-law for their support and interest in my success. Finally, and most importantly, I would like to thank my beautiful wife, Chengcheng, for her unwavering spirit and confidence in my abilities, and her endless devotion to this new family.

Table of Contents

Acknowledgements iv

List of Tables xi

List of Figures xii

Abstract xiv

Chapter 1 Introduction 1

1.1 Motivation for sustainable biofuels 1

1.2 Bioconversion of lignocellulosic biomass to biofuels 3

1.2.1 Naturally evolved lignocellulose degrading systems 4

1.2.2 Processing platforms for lignocellulose bioconversion 9

1.3 Consolidated bioprocessing by *Clostridium cellulolyticum* 12

1.3.1 General physiology 12

1.3.2 Cellulosome complex 14

1.3.3 Key genes and operons involved in cellulolysis 16

1.3.4 Regulatory mechanisms of carbohydrate utilization 18

1.4 CRISPR: A game-changing genetic engineering technique 20

1.4.1 Bacterial CRISPR/Cas system 21

1.4.2 Application of type II CRISPR/Cas system 25

1.4.3 Influential factors of CRISPR/Cas application 32

1.4.4 Comparison of targeted genetic engineering tools 36

1.5 Aim and focus of the study 40

Chapter 2 Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase 43

2.1 Abstract 43
2.2 Introduction 44
2.3 Materials and methods 46
2.3.1 Synthetic promoter design 46
2.3.2 Plasmid construction 47
2.3.3 Bacterial strains and culture conditions 48
2.3.4 Determination of editing efficiency and cargo capacity 50
2.3.5 RNA isolation, RT-PCR and quantitative real-time PCR 51
2.3.6 SDS-PAGE analysis 51
2.3.7 Fluorescence microscopy 51
2.3.8 Bioinformatic analysis of target sites 52
2.4 Results 52
2.4.1 Expression of CRISPR-Cas9 system in *C. cellulolyticum* 52
2.4.2 Lethality of Cas9-induced double-strand breaks 54
2.4.3 Precise genome editing via a single nick-triggered homologous recombination 56
2.4.4 Assessment of editing efficiency and genetic cargo capacity 58
2.4.5 Precise editing at non-specific target sites 61
2.5 Discussion 63
Chapter 3 Cas9 nickase-assisted RNA Repression Enables Stable and Efficient Manipulation of Essential Metabolic Genes in *Clostridium cellulolyticum* . 68
3.1 Abstract 68
3.2 Introduction 70

3.3 Materials and methods 72
3.3.1 Plasmid construction 72
3.3.2 Bacterial strains and culture conditions 75
3.3.3 *C. cellulolyticum* transformation 75
3.3.4 Enzyme activity assay 76
3.3.5 Measurement of cell growth, cellulose consumption and fermentation products 77
3.3.6 Quantitative real-time PCR 79
3.3.7 Microscopy and flow cytometry 80
3.4 Results and discussion 80
3.4.1 Plasmid-based antisense RNA expression 80
3.4.2 Metabolic changes in knock-down strains 83
3.4.3 Chromosomal integration and functional analyses 85
3.4.4 Evaluation of the gene-dosage effect between transformants and chromosomal integrants 88
3.4.5 Improved repression of acetic acid production by a tandem repetitive promoter 89
3.5 Conclusions 94
Chapter 4 Dockerin-containing Protease Inhibitor Protects Key Cellulosomal Cellulases from Proteolysis in *Clostridium cellulolyticum* 95
4.1 Abstract 95
4.2 Introduction 97
4.3 Results 100

4.3.1 *dpi* mutant construction and phenotypic analysis 100
4.3.2 Molecular analysis of the mutant cellulolytic system 102
4.3.3 Quantification of *cipC*, *cel48F* and *cel9E* transcripts, and cellulsome production 104
4.3.4 Characterization of *cel48F* and *cel9E* mutants 105
4.3.5 Expression and enzymatic activity assay of recombinant Dpi protein 107
4.4 Discussion 110
4.5 Materials and Methods 115
4.5.1 Bacterial strains and growth conditions 115
4.5.2 Plasmid construction and transformation of *C. cellulolyticum* H10 117
4.5.3 Growth and cellulose degradation measurement 118
4.5.4 RNA isolation and quantitative real-time PCR 118
4.5.5 Fractionation of extracellular proteins 119
4.5.6 Expression and purification of recombinant Dpi protein 119
4.5.7 SDS-PAGE analysis and MS identification 120
4.5.8 Inhibitory activity test of the recombinant Dpi 120
Chapter 5 Carbohydrate Utilization in *Clostridium cellulolyticum* Differentially Relies on Catabolite Regulation System 122
5.1 Abstract 122
5.2 Introduction 124
5.3 Results and discussion 128
5.3.1 In silico analysis of CCR components in *C. cellulolyticum* 128
5.3.2 Evaluation of carbon catabolite repression in *C. cellulolyticum* 130

5.3.3 Physiological characterization of the *hprK* knockdown mutant..... 132
5.3.4 Mutagenesis and characterization of the *crh* gene and three LacI regulator genes..... 135
5.3.5 Link differentially expressed genes to biological processes..... 139
5.4 Conclusion..... 150
5.5 Materials and methods..... 151
5.5.1 Plasmids and bacterial strains..... 151
5.5.2 Measurement of cell growth, sugar consumption, cellulose consumption, and fermentation products..... 153
5.5.3 Microarray hybridization and data analysis..... 154
5.5.4 Quantitative PCR..... 156
Chapter 6 Summary and Outlook..... 158
Appendix A: Supplementary Tables..... 165
Appendix B: Supplementary Figures..... 187
References..... 211

List of Tables

Table 2.1 Use of Cas9 nickase instead of wild-type Cas9 for genome editing in *C. cellulolyticum*. 56

Table 3.1 Plasmids and strains used to study acetate producing genes. 74

Table 4.1 Inhibitory activity of the recombinant Dpi against commercial proteases. . 109

Table 4.2 Bacterial strains and plasmids constructed to study the *dpi* gene..... 116

Table 5.1 Characteristics of cell growth and substrate consumption with a sole or dual sugar(s). 131

Table 5.2 The number of differentially expressed genes in sugar mixture-grown mutants at the exponential and stationary phases. 144

List of Figures

Figure 1.1 Model of cellulose degradation with complexed and non-complexed cellulase systems.. 7

Figure 1.2 Development of bioconversion platforms for biofuel production from lignocellulose biomass..... 10

Figure 1.3 Type II CRISPR/Cas systems in Streptococci. 23

Figure 1.4 Application of CRISPR/Cas9 in targeted genome editing..... 28

Figure 1.5 Application of engineered dCas9 and/or RNA components in transcriptional control..... 31

Figure 2.1 Generation and validation of Cas9 expression system..... 53

Figure 2.2 Precise deletion and insertion of a small fragment. 57

Figure 2.3 Evaluation of editing efficacy and cargo capacity. 59

Figure 2.4 Targeting specificity test.. 62

Figure 2.5 Bioinformatic analysis of targeting space in *C. cellulolyticum*..... 63

Figure 3.1 Pivotal metabolic pathways in *C. cellulolyticum* and RNA repression. 81

Figure 3.2 Comparison of acetate titers produced on 10 g/L Avicel cellulose and 10 g/L xylan 83

Figure 3.3 Chromosomal integration of functional modules via double-crossover recombination and the Cas9 nickase genome editing tool 86

Figure 3.4 Expression of anaerobic fluorescent protein in the P-AFP transformant and the G-AFP integrant. 89

Figure 3.5 Measurements of major end products and released sugars in the end-point fermentation broth. 91

Figure 3.6 Fermentation kinetics of LM and LM3P on 50 g/L cellulose. 93

Figure 4.1 Growth profiling of WT, *dpi* mutant, *dpi/over* and *dpi/zero* strains. 101

Figure 4.2 Composition of the cellulolytic system of the *dpi* mutant and WT. 103

Figure 4.3 Comparison of transcript levels and cellulosome productivity. 105

Figure 4.4 Growth curves for WT, *cel48F* and *cel9E* mutants grown on 10 g/L Avicel cellulose and 5 g/L cellobiose 106

Figure 4.5 Detection of the polar effect by quantitative real-time PCR. 107

Figure 4.6 Characterization of Dpi protein 109

Figure 4.7 A conceptual model of Dpi-mediated regulation of cellulosomal activity in *C. cellulolyticum* 115

Figure 5.1 Growth profiling of pRNAi and pRNAi-*hprK* strains 134

Figure 5.2 Measurements of residual sugars during cell growth 135

Figure 5.3 Effect of mutagenesis on colony development.. 137

Figure 5.4 Cell growth and sugar utilization in the defined medium with a sugar mix as carbon source 138

Figure 5.5 Microarray-based transcriptomic analyses in mutants at the exponential phase. 142

Figure 5.6 Microarray-based transcriptomic analyses in mutants at the stationary phase 147

Figure 5.7 A model of carbon catabolite regulation in *C. cellulolyticum*. 150

Abstract

The astonishing consumption of fossil fuels arouses serious concerns over energy security and environmental sustainability. Lignocellulosic biofuel, as a sustainable and carbon-neutral energy that can be produced from lignocellulosic biomass, has the potential to mitigate these pressures. However, the pace of microbial engineering towards efficient and cost-effective biofuel production has been hindered mainly due to the limited knowledge of biological systems in potential microbes and the lack of robust genome engineering tools for efficient functional genomics studies and engineering-oriented practices. Research in the model organism of mesophilic cellulolytic clostridia, *Clostridium cellulolyticum*, which can perform one-step lignocellulose bioconversion, is still facing the same challenges. The two major objectives of this dissertation are to: 1) develop genome editing tools that allow us to efficiently manipulate both essential and non-essential genes in a targeted manner; 2) conduct comprehensive studies on key metabolic genes, cellulose-degrading cellulosomes, and catabolite regulation systems to increase our understanding of carbohydrate assimilation and metabolism in *C. cellulolyticum*.

From the aspect of method development, the revolutionary Cas9 nuclease-mediated genome engineering tool was timely and successfully adapted to edit the genome of *C. cellulolyticum*. The established method employs a mutated Cas9 nickase to generate a single nick at the specific target site to trigger homologous recombination. It overcomes the toxicity of severe DNA damages that previously reported Cas9-based editing methods can cause. With intensive editing tests in *C. cellulolyticum*, this method presented the advantage of marker-independent gene delivery, versatile editing, and

multiplex editing in a single step at a very high editing efficiency and specificity. Besides, our combinatorial method using the Cas9 nickase editing tool to integrate gene repression modules into the chromosome was successfully applied to manipulate essential metabolic genes in this bacterium in a plasmid-independent way.

From the aspect of intellectual knowledge, this work firstly reduced acetate production via antisense RNA-mediated repression of the phosphotransacetylase gene. The effectiveness of both plasmid- and chromosome-based repression was compared; however, switching to chromosome-based expression dramatically decreased gene dosage and formed much less functional gene products, which resulted in a weak repression in chromosomal integrants. The challenge was overcome by integrating a tandem promoter-driven RNA expression module to enhance RNA expression.

Second, three cellulosomal components, Dpi, Cel48F and Cel9E, were identified to be important for cellulose degradation in *C. cellulolyticum*. Dpi was proven to be a cysteine protease inhibitor. Loss of the Dpi encoding gene dramatically decreased the abundance of major cellulosomal components, Cel48F endocellulase and Cel9E exoglucanase. These two cellulases were verified to be almost indispensable for cellulose degradation via mutagenesis. This study provides the first evidence to show the in vivo importance of cellulosomal protease inhibitors in protecting pivotal cellulosomal components from proteolysis.

Third, all predictable components of carbon catabolite regulation (CCR) in *C. cellulolyticum* were characterized at the physiological, genetic and transcriptional level. This bacterium lost the sugar-transporting phosphotransferase system in the genome and exhibited a very mild reverse catabolite repression. Mutagenesis of the predicted

regulatory system of CCR, including *hprK*, *crh* and *ccpA*, showed that cellobiose assimilation was independent of CCR under our test condition, but the utilization of monomers (both pentoses and hexoses) and insoluble cellulose were tightly associated with CCR. This study also provided the first genetic evidence to show the indispensability of the *crh* and *ccpA* genes in cellulose catabolism. Thus, carbohydrate utilization in this bacterium presented differential reliance on this regulation system. Transcriptomic analysis found that the *crh* gene played a significant regulatory role in gene expression; two other LacI member regulators *lfpC2* and *lfpC3*, which are similar to the *ccpA* gene, presented functional specificity and redundancy; the *ccpA* gene exerted minimal impacts when cells grow on soluble sugars.

Aforementioned functional genomics studies provide novel insights into the physiological and genetic importance of a series of genes in sugar assimilation, cellulose degradation and cellular metabolism in *C. cellulolyticum*. These discoveries will help microbial engineers to develop feasible strategies to improve lignocellulose bioconversion. The developed Cas9 nickase-based genome editing tool and its derivative, Cas9 nickase-assisted RNA repression, will facilitate microbial gene/genome modification for fundamental and applied research.

Keywords: **CRISPR/Cas; genome editing; RNA repression; cellulosome; carbon catabolite repression; biofuel; lignocellulose bioconversion**

Chapter 1 Introduction

1.1 Motivation for sustainable biofuels

With the fast development of human society and economy, energy consumption has increased dramatically in the past few decades. In 2014 fossil fuels (i.e., petroleum oil, coal and natural gas) accounted for 81% of world total primary energy supply (IEA, 2016). The astonishing reliance on fossil fuels confronts energy security, environment and even ecosystems with unprecedented pressures (IEA, 2009). Scientific researchers around the world are casting endless efforts to develop alternative energy as one of incentives to diminish fossil fuel usage.

Fossil fuels as nonrenewable energy sources have finite reserves on the earth. Most nations mainly use fossil fuels to make electricity, transportation fuels, and industrial chemicals for diverse uses (IEA, 2011). In the U.S. in 2015, 67% of electricity was generated from fossil fuels (EIA, 2016). Considering fossil fuel supplies are limited, a global energy shortage and national security will be of concerns in future. History has already indicated to us that largely importing oils from volatile regions of the world may cause political and geopolitical challenges that can then affect the economy (Mercier, 2009; Vandamme, 2009). As a nation, the solution to mitigate the concerns over national energy security is to change the energy supply mix towards renewables, such as solar, wind, and bioenergy (DOE, 2012).

Massive fossil fuel usage aggravates environmental issues. Increasing atmospheric greenhouse gasses, which are considerably attributed to the global CO_2 release (about 7 Gt of carbon per year) from fossil fuel usage, have been linked to global climate and environmental changes, such as rising sea levels, weakening of

thermohaline circulation and eradication of coral reef (O'Neill *et al.*, 2002; Pacala *et al.*, 2004; Lewis *et al.*, 2006). With the current upward trend of the CO_2 release rate, the global temperature by 2050 will increase approximately 2 ℃ above the level in 1900 (O'Neill *et al.*, 2002). Although the projected increase is numerically small, it would bring the risk of disintegration of the West Antarctic Ice Sheet and subsequently the increase of sea levels which could massively destroy our society, economy and our indispensable ecosystem (O'Neill *et al.*, 2002). Apart from these possible global disruptive effects, fossil fuel usage is causing air pollution in the form of air smog and acid rain in some countries. For example, fossil fuel combustion made the largest contribution to Beijing's air smog in 2013 (Zhang *et al.*, 2014). The World Health Organization reported that 3 million people die each year from air pollution-caused health problems, such as lung cancer, respiratory infections and heart disease (WHO, 2016). All these ongoing and potential impacts are too large to ignore such that all people should contribute to reducing fossil fuel usage and developing green, carbon-neutral renewable energy to replace fossil fuels.

Biofuels refer to fuels that are produced from bio-based materials such as biomass, or produced by biological systems. Liquid biofuels are superior to other renewable energy forms (e.g., solar/wind-based electricity, biogas) in terms of energy density and compatibility with current infrastructure (Liao *et al.*, 2016). There are various biomass resources that can be used for biofuel production, including food grade sources, non-edible lignocellulosic biomass, municipal solid waste, and algae. Although biodiesel and alcohols have been produced from food resources (e.g., soybean, corn grain, sugarcane, and oil) in an economically efficient way, this strategy competes with

land and water usage for food production leading to impacts on the global food market and food security especially in vulnerable regions (Naik *et al.*, 2010). In comparison, non-edible lignocellulosic materials are the most promising feedstock as natural and renewable resource. The global annual production of plant biomass is about 200 Gt, in which over 67% of dry mass is in the form of lignocellulose and from where 8-20 Gt of the primary biomass can be potentially obtained for biofuel production (Kuhad *et al.*, 1993). Since plant biomass is generated by photosynthetic CO_2 fixation, lignocellulosic biofuel usage is carbon neutral and eco-friendly. In practice, lignocellulosic biomass can be collected from dedicated energy crops, agricultural residues, forest harvesting residues or wood processing waste, rather than from food resources. Therefore, biofuels generated from lignocellulosic biomass are sustainable and environmental friendly.

1.2 Bioconversion of lignocellulosic biomass to biofuels

Lignocellulolytic microorganisms produce diverse enzymes to degrade cellulose, hemicellulose and even lignin, into soluble carbons to support cellular metabolisms (Lynd *et al.*, 2002; Doi, 2008). Extensive exploitation of these degraders and active enzymes has uncovered a wide variety of biological mechanisms in lignocellulose hydrolysis. However, only a very few of the biomass-degrading enzymes and microorganisms have been utilized for biofuels production but the cost and conversion efficiency still set significant challenges for large-scale industrial operation (Klein-Marcuschamer *et al.*, 2012; Balan, 2014; Liao *et al.*, 2016). To give a general overview of how lignocellulose is decomposed and converted into biofuels, here we will discuss lignocellulose decomposition and bioconversion processing strategies. Moreover, key

barriers of producing biomass-derived biofuels will be discussed along with potential addressing strategies.

1.2.1 Naturally evolved lignocellulose degrading systems

Decomposition of most lignocellulose biomass requires the cleavage of O-glycosidic bonds, which link sugar units to form large sugar polymers, i.e., polysaccharides. Glycoside hydrolases (GHs) acting on these bonds are roughly classified into endo-acting and exo-acting enzymes (Naumoff, 2011). Endo-acting glycosidases cleave the internal glycosidic linkages of polymers; Exo-acting ones act on the bond between the sugar residue at the end of the chain and the rest of the polymer. Many GHs are modular enzymes consisting of glycosyl hydrolase catalytic domains (CD), carbohydrate-binding modules (CBM), and type I dockerin domains (DD) (Fontes *et al.*, 2010; Naumoff, 2011). The GHs have shown versatile enzymatic properties, in terms of substrate specificity, product diversity and catalytic efficiency.

Cellulose hydrolysis requires enzymatic cleavage of β-1, 4-glycosidic bonds between D-glucose units. GHs with this function are generally called cellulases, and can be mainly divided into three classes as follows (Lynd *et al.*, 2002). Endoglucanases randomly cleave interior glycosidic bonds in cellulose, releasing oligosaccharides of varied length with new reducing and non-reducing ends. This function greatly contributes to solubilizing the cellulose polymer by reducing molecular size and creating accessible chain ends for further attack. Cel48F, CelC, and Cel7B proteins are typical endoglucanases critical in cellulose degradation in *Clostridium cellulolyticum* (Perret *et al.*, 2004), *Clostridium thermocellum* (Wang *et al.*, 1993) and *Trichoderma reesei* (Kleman-Leyer *et al.*, 1996), respectively. In contrast, exoglucanases act from

chain ends of cellulose oligosaccharides to progressively chip off glucose or cellobiose (di-glucose) units (Lynd *et al.*, 2002). Glucose- and cellobiose- releasing exoglucanases are also called exo-1, 4-β-glucosidase and cellobiohydrolases, respectively. β-glucosidases (EC 4.2.1.21) typically split cellobiose dimers, or sometimes cellotrioses, into individual glucose units and then release the inhibitory effect of accumulated cellobiose on exo- and endo-glucanases activity (Gruno *et al.*, 2004; Yue *et al.*, 2004). These three classes of cellulases are critical to cellulose degradation and have been applied in different industries (Kuhad *et al.*, 2011). In addition, some bacteria produce cellobiose phosphorylases and cellobiose dehydrogenases to improve cellulose degradation (Alexander, 1968; Reichenbecher *et al.*, 1997; Sygmund *et al.*, 2012). Cellulolytic microorganisms produce a diversity of these enzymes for synergistic catalysis to speed up cellulose degradation (Doi, 2008; Fontes *et al.*, 2010).

Hemicelluloses, as the second most abundant polymer in nature, are heterogeneous polymers of pentoses (xylose, arabinose), hexoses (mannose, glucose, galactose), and sugar acids (Girio *et al.*, 2010). According to hemicellulose structure in the cell wall, it can be classified into xyloglucans (XGs), galactoglucomannans (GGMs) and glucuronoarabinoxylans (GAXs) (Girio *et al.*, 2010). Many microorganisms, such as *Penicillium capsulatum* and *Talaromyces emersonii*, possess complete degradation systems for the glucuronoarabinoxylans (GAXs) that are the most abundant hemicellulose in grasses (Filho *et al.*, 1991). Like cellulose biodecomposition, total biodegradation of GAXs also requires diverse enzymes for depolymerization and side-group cleavage. Endo-xylanases attack internal bonds in the main chains of xylans; exo-xylanases hydrolyze β-1, 4-xylose linkages at chain ends to release xylobiose; and then

β-xylosidase further hydrolyzes xylooligosaccharides and xylobiose to xylose (Gilbert *et al.*, 2008). Side chains on xylose units block the action of some xylanases, leading to the evolution of diverse accessory enzymes (e.g., α-arabinofuranosidase, α-glucuronidase, acetylxylan esterase, feruloylesterases and p-coumaric acid esterase) to remove the side chains and make the xylan backbone accessible for complete hydrolysis (Perez *et al.*, 2002; Gilbert *et al.*, 2008). Some xylanases have been identified in cellulosome complex of *C. cellulolyticum* and *C. thermocellum* (Raman *et al.*, 2009; Blouzard *et al.*, 2010). However, to degrade GGMs-rich biomass, additional endomananases and β-mannosidases are needed after the removal of side chains by some esterases (Gilbert *et al.*, 2008).

Lignin is a complex macromolecule (Boerjan *et al.*, 2003; Vanholme *et al.*, 2010). The crosslinking structure makes lignin the most recalcitrant substance for chemical or biological fermentation. However, due to its high energy content, it could be separated from lignocellulose for electricity and chemicals production. Unfortunately, lignin structure caused barriers for cellulose and hemicellulose digestion since enzymes could not get access to the wrapped substrates such that biomass pretreatment was applied prior to enzyme/cell-based hydrolysis (Chang *et al.*, 2000; Balan, 2014).

The individual classes of cellulases described above function within both non-complexed and complexed cellulase systems (Fontes *et al.*, 2010). The model for both systems has been shown in Figure 1.1. The non-complexed systems consist of individual enzymes that can have multiple catalytic and CBM domains, but that otherwise act without interacting with other classes of hydrolases. In contrast, the complexed systems, also known as cellulosomes, are superstructural, multi-polypeptide

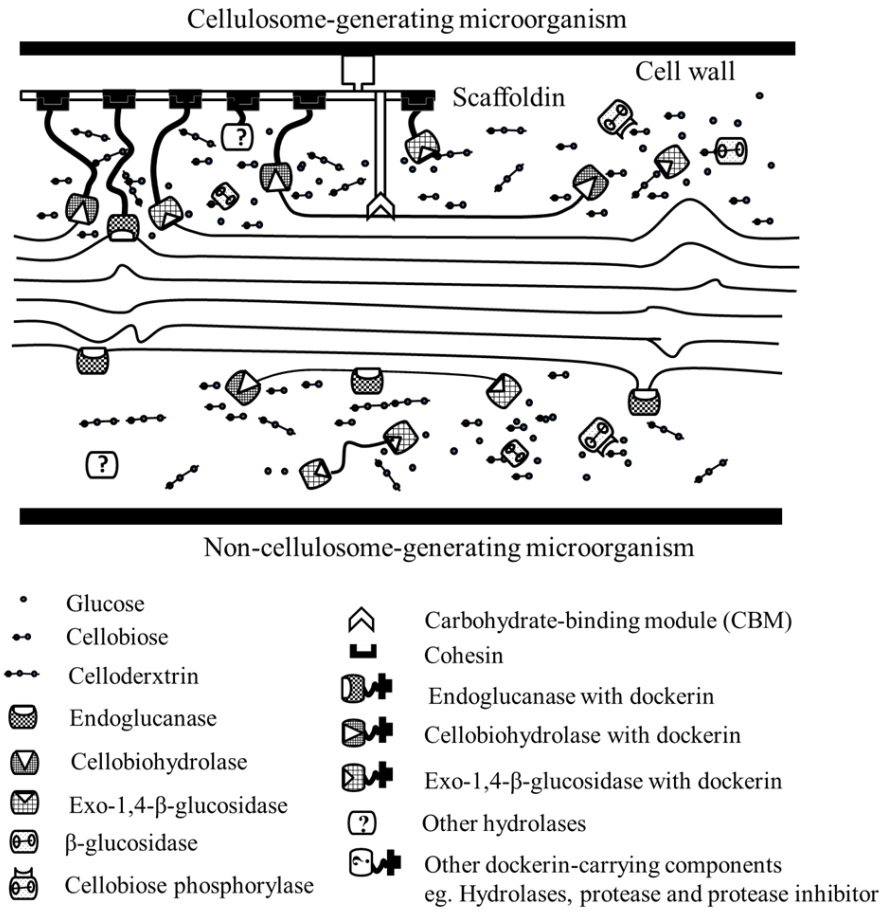


Figure 1.1 Model of cellulose degradation with complexed and non-complexed cellulase systems. The upper and lower parts demonstrate how cellulose is synergistically degraded by the two systems, respectively. All components in the model are not drawn to scale. Figure adapted from Lynd et al. (2002).

enzyme complexes that adhere to cell wall of lignocellulolytic bacteria or fungi (Fontes *et al.*, 2010). They consist of a multi-functional integrating subunit, called a scaffoldin, that is composed of multiple cohesion modules, and diverse enzymatic subunits with dockerin modules that interacts with the scaffoldin. For example, the cellulosomes of *C. cellulolyticum* could contain cellulases, xylanases, mananases, and even protease inhibitors (Blouzard *et al.*, 2010). Cellulosome composition is dynamic and heterogenous, depending on the bacteria and composition of extracellular

polysaccharides and even the relative amounts of the available dockerin-containing modules that can be incorporated into the complex (Raman *et al.*, 2009). Cellulosomes have higher cellulose degradation efficiency than non-complexed enzymes since their adhesion to cell surface prevents degrading substrates being lost to diffusion or uptake by neighboring bacteria and then facilitate the uptake of hydrolysis products (Schwarz, 2001). In vitro artificial construction of mini-cellulosomes and self-assembly of cellulosome on yeast surface has presented an efficient way to significantly enhance cellulose hydrolysis rates compared with free enzymes (Wen *et al.*, 2010; Fan *et al.*, 2012; You *et al.*, 2012).

Cellulosome-generating microorganisms have shown diversity in cellulosomal composition or architecture. A proteomic study on isolated cellulosomes from *C. cellulolyticum* confirmed the expression of 50 dockerin-containing proteins out of 62 predicted by bioinformatics (Blouzard *et al.*, 2010). The complexity of the cellulosome is highly related with the availability or abundance of cellulosomal components, whose expression were influenced by substrate induction or catabolite repression. When *C. cellulolyticum* grew on cellulose substrate, 36 enzymes were detected on the cellulosome, 30 on xylan, and 48 on hatched wheat straw (Blouzard *et al.*, 2010). Thus the cellulosome is heterogeneous with varied components and component abundance. Moreover, to some microbes, the diversity in cellulosome is beyond this level since the presence of multiple types of scaffoldins in a single genome has been reported (Fontes *et al.*, 2010). *C. thermocellum* contains four type II cohesion-containing anchoring scaffoldins (Bayer *et al.*, 1998). The cellulosomes assembled by type II dockerin domain of CipA could be further organized into a larger complex by interacting with

these type II cohesion-containing anchoring scaffoldins (Bayer *et al.*, 1986; Raman *et al.*, 2009).

1.2.2 Processing platforms for lignocellulose bioconversion

Generally, biological conversion of pretreated lignocellulose into biofuels contains four steps: glycoside hydrolase production, enzymatic hydrolysis of cellulose and hemicellulose, hexose fermentation and pentose fermentaion (Lynd *et al.*, 2002). So far, several processing platforms have been developped to accomplish all steps in several or only one units (Figure 1.2). Separate enzyme hydrolysis and fermentation (SHF) completes each step in an independent unit. The biggest advantage of this process is the ability to carry out enzymatic hydrolysis and microorgnism-based fermentations at their own optimum conditions. However, a major problem accompanying the separate hydrolysis is the inhibitory effects on cellulase activity caused by accumulated products, like glucose and cellobiose (Philippidis *et al.*, 1993; Gruno *et al.*, 2004). The combination of enzymatic hydrolysis and fermentation generated another platform, simultaneous saccharification and fermentation (SSF) (Olofsson *et al.*, 2008). This process has been successfully applied to convert lignocellulose to enthanol with higher yield, less enzyme dose and less equipment than SHF (Olsson *et al.*, 2006; Saha *et al.*, 2011; Zhu *et al.*, 2012). To maximize the fermentation efficiency of SSF, the key is to select hydrolases and fermenting microorgnisms with similar optimum temperature and pH. However, most microorgnisms need a lower optimum temperature than hydrolases, which makes saccharification a limiting factor to SSF. People trying to overcome this difficulty designed nonisothermal simultaneous saccharification and fermentation (NSSF), in which saccharification and fermentation occur simultaneously but in two

separate bioreactors at different temperatures, coupled with recirculation of fermentation broth between these two bioreactors (Wu *et al.*, 1998; Oh *et al.*, 2000). It presented several advantages compared with SSF, including higher ethanol yield, shorter residence time and less enzyme input.

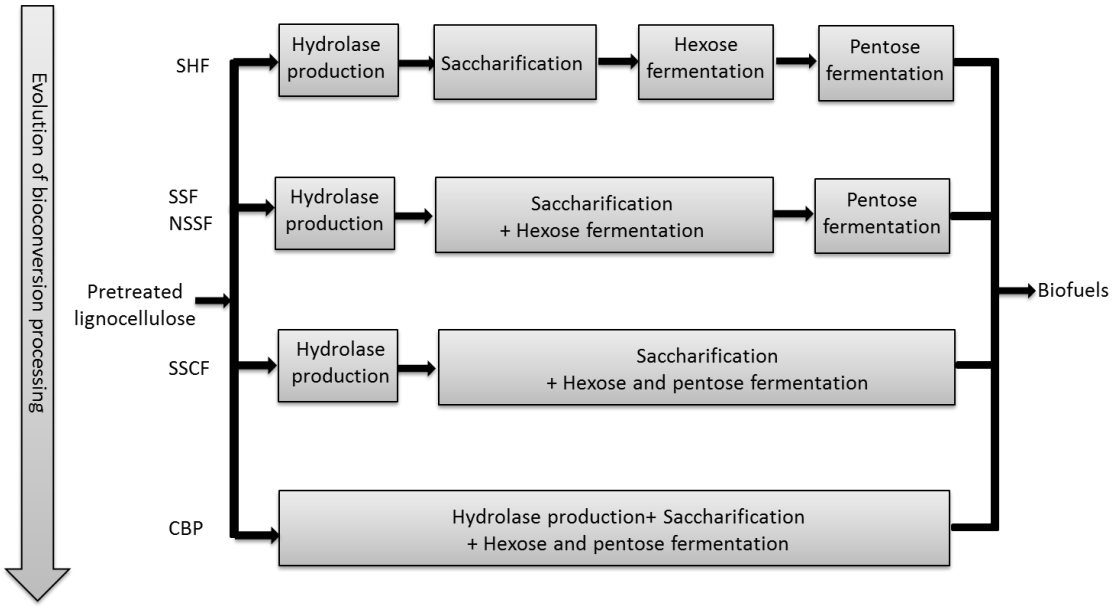


Figure 1.2 Development of bioconversion platforms for biofuel production from lignocellulose biomass. The function of each step in different platforms, separate enzyme hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), nonisothermal simultaneous saccharification and fermentation (NSSF), Simultaneous saccharification and cofermentation (SSCF), and consolidated bioprocessing (CBP), is indicated in the boxes. Figure adapted from Lynd et al. (2002).

For the above bioconversion processes, when using hemicellulose-rich biomass they all need additional separate pentose fermentation with different microorgnisms in another bioreactor. The more steps involved in the process, the more time is required to complete a feremention cycle and more money is used for equipment installation. Simultaneous saccharification and cofermentation (SSCF) as an improvement of SSF, sets out to ferment both pentose and hexose in a single bioreactor (Lynd *et al.*, 2002).

Therefore with this platform, microorgnisms or engineered ones are able to simultaneously ferment six- and five-carbon sugars to biofuels. Recently SSCF has been applied to ferment commercial furfural, corn kernels and pretreated wheat straw (McMillan *et al.*, 1999; Zhu *et al.*, 2007; Olofsson *et al.*, 2010; Tang *et al.*, 2011).

The biggest common barrier to all of the above platforms is the cost of cellulase. The newest concept of bioconverstion processes named consolidated bioprocessing (CBP), employs a single microorganism or microbial consortium for hydrolase production, saccharification and fermentation in a single step in one bioreactor (Lynd *et al.*, 2002). Obviously, CBP offers the potential of lower production costs due to simpler conversion processing, lower energy and money inputs, and potentially higher conversion efficiency than above platforms. However, the key challenge of CBP is that there is no ideal CBP-enbaling microorganism capable of efficient cellulose hydrolysis and biofuels production at the same time. Two strategies were proposed to enable consolidated bioprocessing (Lynd *et al.*, 2005): (i) engineering naturally occurring cellulolytic microorganisms to improve the formation of interesting products, and (ii) engineering non-cellulolytic organisms that exhibit superior fermentation ability to express cellulolytic systems. So far, *E.coli* and yeast have been engineered to directly convert cellulose and xylan to ethanol and biodiesel (Steen *et al.*, 2010; Bokinsky *et al.*, 2011; Goyal *et al.*, 2011; Fan *et al.*, 2012); *Clostridium* species with native lignocellulose-degrading ability have been metabolically engineered to synthesize a variety of biofuels, such as hydrogen, isopropanol, butanol and ethanol (Higashide *et al.*, 2011; Lutke-Eversloh *et al.*, 2011; Lee *et al.*, 2012). In addition, researchers have proposed the idea of co-culture, in which non-biofuel products generated by one

microorgnism could be further converted to biofuels by a second one, or in which the metabolism of one microorgnism could be boosted by the existence of another one (Masset *et al.*, 2012; Park *et al.*, 2012). However, we still have a long way to go before industrial application since problems in CBP (Balan, 2014; Liao *et al.*, 2016), like low-efficiency substrate utilization, microbial growth inhibition, low microbial tolerance to products and low product yield, require our endless efforts to be addressed.

1.3 Consolidated bioprocessing by *Clostridium cellulolyticum*

Cellulolytic microorganisms play an important role in cellulose decomposition, which is a key process in carbon cycling in nature. About 80% of cellulolytic bacteria isolated previously are Gram-positive, belonging to only two phyla, Actinobacteria and Firmicutes (Desvaux, 2005). Most Gram-positive cellulolytic anaerobes are found in the Firmicutes and more particularly in the genus *Clostridium*. These isolates vary a lot in detailed mechanisms of lignocellulose degradation, carbon metabolism regulation, and other physiological features (Lynd *et al.*, 2002).

1.3.1 General physiology

Clostridium cellulolyticum strain H10 (ATCC 35319) is a non-ruminal cellulolytic mesophilic bacterium isolated from decayed grass in France (Petitdemange *et al.*, 1984). It is an anaerobic bacillus, straight to slightly curved rod that is 3-6 μm long and 0.6-1 μm wide, with peritricheous flagella. Under unfavorable or harsh growth conditions, spherical terminal spores can be generated with a 1.5 μm diameter, which can resist 100°C for 30 min (Petitdemange *et al.*, 1984; Li *et al.*, 2014). This mesophilic cellulolytic bacterium can grow at 25-45°C with an optimum growth temperature at 34°C. There are a variety of carbohydrates *C. cellulolyticum* can use for growth

(Petitdemange *et al.*, 1984): 1) moderate growth was observed with cellulose, cellobiose, glucose, xylose, arabinose, and fructose; 2) poor growth was observed with galactose, mannose, or ribose; 3) no growth was observed with sucrose, lactose, glycerol, glycogen, or sugar alcohols. Interestingly, this bacterium can grow faster on cellobiose than on glucose and cellulose (Petitdemange *et al.*, 1984). At a deeper level, how sugars can be metabolized, especially when multiple sugars are present, needs to be systematically studied in *C. cellulolyticum* since previous reports revealed distinct sugar preference between *Clostridium thermohydrosulfuricum* and *C. thermocellum* (Ng *et al.*, 1982). It is quite common to see bacteria consume glucose in preference to other carbohydrates (Stulke *et al.*, 1999; Singh *et al.*, 2008). For *C. cellulolyticum*, growth also occurs utilizing the most complex lignocelluloses including switchgrass, wheat straw, corn stover, and xylan (Blouzard *et al.*, 2010; Li *et al.*, 2012; Xu *et al.*, 2013). With more and more studies on cellulose hydrolysis and metabolism in *C. cellulolyticum*, this bacterium has been considered as a model of mesophilic cellulolytic Clostridia.

The most attractive feature of *C. cellulolyticum* is its capability of anaerobic fermentation of lignocellulosic biomass to produce lactate, acetate, ethanol, and H_2 (Petitdemange *et al.*, 1984; Lynd *et al.*, 2002; Li *et al.*, 2012), which can be used as biofuels and commodity chemicals. In other words, this bacterium can simultaneously accomplish both jobs that lignocellulose-degrading enzymes/bacteria and sugar-fermenting bacteria can do separately. Therefore, it is a consolidated bioprocessing-enabling bacterium. Compared with thermophilic bacteria capable of consolidating bioprocessing, *C. cellulolyticum* saves energy by avoiding the high temperature demand

to maintain the fermentation condition. From the aspect of industrial production, *C. cellulolyticum* has a few advantages over cellulolytic aerobes, including no need of air agitation and a lower chance of contamination. To produce advanced isobutanol in this bacterium or enhance the production of more valuable ethanol from cellulose, researchers conducted metabolic engineering via overexpressing a series of foreign genes of intended pathways (Guedon *et al.*, 2002; Higashide *et al.*, 2011; Li *et al.*, 2014; Lin *et al.*, 2015), or eliminating competing and promiscuous lactate and malate pathways (Shaw *et al.*, 2008; Li *et al.*, 2012; Papanek *et al.*, 2015). Since acetate production acidifies the growth medium and consumes intermediates that can be used by more useful pathways, the attempt of eliminating acetate formation has been attempted but failed to isolate knock-out mutants of acetate producing genes, phosphotransacetylase and acetate kinase (Li *et al.*, 2012). To stably manipulate essential metabolic pathways within a biologically allowable range, we need to develop other genetic modification approaches instead of just relying on traditional gene knockout.

1.3.2 Cellulosome complex

Like many other cellulolytic bacteria, this bacterium secretes abundant sets of lignocellulose degrading enzymes that can work synergistically as described above. A total of 148 putative carbohydrate-active enzymes were identified in the genome, among which there are 90 putative glycoside hydrolases, 4 putative polysaccharide lyases, 15 putative carbohydrate esterases (Blouzard *et al.*, 2010). Studies on cellulosomal composition showed that regardless of lignocellulose sources (e.g., cellulose, xylan, and wheat straw) for cell growth (Blouzard *et al.*, 2007; Blouzard *et*

al., 2010), the majority of the proteins encoded by the *cip-cel* operon were detected in all cellulosome preparations; cellulosomal composition varied depending on the growth substrate and the availability of dockerin-containing proteins present in the extracellular matrix (Perret *et al.*, 2004; Blouzard *et al.*, 2010). In *C. cellulolyticum*, enzymes with dockerin domains can physically interact with cohesion domains of a large non-catalytic scaffolding protein CipC to form stable multi-enzymatic cellulosome complex with a molecular weight of about 600 kDa. In vitro, purified cellulosomes from *C. cellulolyticum* formed aggregates with the size of 16 MDa (Gal *et al.*, 1997). As aforementioned, cellulosomes have an efficient cellulolytic activity mainly due to synergistic effects of diverse enzymes in the complex and close proximity of enzymes and cells to substrates. In a few well-studied cellulosome-producing bacteria (e.g., *C. thermocellum* and *Clostridium cellulovorans*), cellulosomes are tethered to cell surface and then appear as protuberances at bacterial cell surface under scanning electron microscope (Bayer *et al.*, 1986; Blair *et al.*, 1998; Carvalho *et al.*, 2003); however, no protuberances have been observed on the cell surface of cellulose-grown *C. cellulolyticum* (Ferdinand *et al.*, 2013). Both cell-bound and cell-free cellulosomes have been reported in *C. cellulolyticum* and *C. thermocellum* (Mohand-Oussaid *et al.*, 1999; Xu *et al.*, 2016). Recently, the cell-free cellulosomal system of *C. thermocellum* has been proven to be involved in cellulose degradation remotely from bacterial cells (Xu et al., 2016). Apart from producing multi-enzymatic cellulosomes, *C. cellulolyticum* secrets free hydrolytic enzymes without dockerin domains or any other cell surface anchoring domains. Even though these free enzymes are not enzymatically superior to

cellulosomes, they presumably contribute to long-distance hydrolysis and help with substrate supply.

1.3.3 Key genes and operons involved in cellulolysis

C. cellulolyticum has a single circular chromosome with the size of 4,068,724 bp and a GC content of 37.4% (GenBank Accession: NC_011898.1) (Hemme *et al.*, 2010). It has 3390 protein-encoding genes. In comparison with other mesophilic cellulosome-producing Clostridia, this bacterium has the least number of carbohydrate active genes but statistically with the largest proportion of cellulosomal genes including 62 putative dockerin-containing enzyme genes and three putative cohesion-encoding genes (cipC/Ccel_0728, orfX/Ccel_0733 and Ccel_1543) (Blouzard *et al.*, 2010; Xu *et al.*, 2013). Expression of 50 cellulosomal proteins has been identified in isolated cellulosome by proteomics (Blouzard *et al.*, 2010). Interestingly, many cellulosomal genes are organized in gene clusters across the genome.

The first gene cluster is called *cip-cel* operon (Ccel_0728-0740) with the size of 24 kb, consisting of 12 genes **(***cipC, cel48F, cel8C, cel9G, cel9E, orfX, cel9H, cel9J, man5K, cel9M, rgl11Y,* and *cel5N***)** (Maamar *et al.*, 2006). The whole transcriptional activity was controlled by a sole promoter upstream of the first encoding gene (*cipC*) without any internal active promoters experimentally identified; however, RNA processing occurred on the primary transcripts, resulting in a highly skewed transcript ratio and then a large variation in cellulosome stoichiometry due to RNA stabilization (Xu *et al.*, 2015). RNA sequencing and qPCR analysis consistently revealed that the upper genes in the operon (Ccel_0728-0732), particularly including *cipC*, *cel48F*, and *cel9E*, possessed much higher transcript abundance than those genes located in the 3′

part of the cluster (Maamar *et al.*, 2006; Xu *et al.*, 2015). Most gene products of this operon are cellulases; CipC, Cel48F, and Cel9E are three major cellulosomal components in *C. cellulolyticum* (Maamar *et al.*, 2004; Perret *et al.*, 2004). The second gene cluster is called *xyl-doc* (Ccel_1229-1242), with the size of 32 kb, consisting of 14 cellulosomal genes encoding exclusively enzymes which are probably involved in hemicellulose degradation (Blouzard et al., 2010; Celik et al., 2013; Xu et al., 2015). All genes in the cluster were co-transcribed and their expression was regulated by a two-component system (XydS/R) in response to straw (Celik *et al.*, 2013). Apart from these aforementioned two big operons, there are a few biocistronic operons encoding cellulosomal enzymes (Man 26A/Ccel_0752-Cel9P/Ccel_0753, PL10/Ccel_1245-CE8/ Ccel_1246, Ccel_1655-1656, and Ccel_1549-1550) (Xu *et al.*, 2015). This clustering organization of cellulosomal genes in *C. cellulolyticum* is not found in *C. thermocellum*, a well-studied thermophilic cellulosome-producing anaerobe (Guglielmi *et al.*, 1998).

CipC is a modular scaffolding protein without catalytic activities, consisting of a cellulose-binding domain, two hydrophilic X-modules with unknown functions (hereafter called X2 modules), and eight type I cohesion domains, all of which are separated by short linker sequences. The cohesions physically interact with type I dockerin domains borne by diverse enzymes to finally build up cellulosome complexes. Of note, CipC does not contain a type II dockerin domain which is responsible for cellulosome anchorage at the cell surface of *C. thermocellum* and then mediating cell binding to cellulose. In *C. thermocellum*, cellulosomes are tethered on cell surface by the physical interaction between the type II dockerin in cellulosomal scaffolding proteins and type II cohesins of several surface layer proteins (Hong *et al.*, 2014). In *C.*

cellulovorans, cellulosome anchorage is attributed to both hydrophilic modules of the scaffolding protein, which is capable of binding to bacterial cell wall fractions, and the cellulosomal enzyme Eng5E, which contains both type I dockerin and surface layer homology domains (Doi *et al.*, 2004). However, Eng5E homologs and type II cohesion/dockerin surface layer protein encoding genes are not identified in *C. cellulolyticum* genome. In comparison with the important role of cellulosomal scaffolding proteins in mediating cell adhesion to cellulose in both *C. thermocellum* and *C. cellulovorans*, CipC of *C. cellulolyticum* has been proven to be only partly involved in binding of cells to cellulose (Ferdinand *et al.*, 2013). Yet we know little about how CipC mediates cell adhesion, what other mechanisms also contribute to cell adhesion to cellulose, and why cellulosome protuberance does not appear at the cell surface.

1.3.4 Regulatory mechanisms of carbohydrate utilization

Previous studies have mainly focused on cloning, expression and in vitro enzymatic characterization of many cellulosomal enzymes, but rarely on the physiological/genetic importance of these enzymes on cell-based cellulolysis and barely on the regulatory mechanism of key operons especially in response to changing environmental factors. At the beginning of this project, the only genetic studies on cellulosomal genes just showed *cipC* disruption and *cel48F* repression severely impaired cellulolysis (Maamar *et al.*, 2004; Perret *et al.*, 2004) such that we do not know the contribution of other cellulosomal components to the well-known synergistic hydrolysis. The *cip-cel* operon promoter has been studied using a transcriptional fusion approach, showing that the promoter activity was enhanced by switching growth substrates from cellobiose to cellulose; more interestingly, a catabolite-responsive element (*cre*) downstream from

the transcriptional start site was proven to be functional in regulating the promoter activity probably via interacting with transcriptional regulators of carbon catabolite repression (Abdou *et al.*, 2008). Recent studies found that a low concentration of glucose unexpectedly stimulated cellulose degradation (Xu *et al.*, 2013).

With lignocellulosic biomass as a carbon source, lignocellulose hydrolysis is a prerequisite of carbon catabolism to support cell growth and fermentative activities. Usually, diverse cellodexins and simple sugars (pentose and hexose) will be released during the degradation of cheap but complex lignocellulosic feedstock (Li *et al.*, 2012). To make biofuels competitive with petroleum–based products, we must make use of fermentable hydrolysates (Liao *et al.*, 2016). Although bacteria evolutionarily obtain metabolic versatility and flexibility in response to diverse substrates, these features usually cause diauxic cell growth and stepwise utilization of fermentable sugars (Stulke *et al.*, 1999; Singh *et al.*, 2008), which will result in lower substrate utilization efficiency and a longer fermentation time in industry. *C. cellulolyticum* can utilize diverse sugars as mentioned above; however, we know little about how sugars are sensed and transported, and how sugar catabolism is regulated in this model bacterium, particularly when multiple sugars, hexoses and/or pentoses, are present simultaneously. Carbon catabolite regulation, which is regarded as an important regulatory system in bacteria (Goerke *et al.*, 2008), may be associated with the regulation of lignocellulose hydrolysis and the assimilation of available sugars in *C. cellulolyticum*. This assumption is supported by the verification of a functional *cre* operator in the *cip-cel* promoter (Abdou *et al.*, 2008) and the negative correlation between the transcriptional levels of

the *cip-cel* operon and LacI family regulators (Xu *et al.*, 2013); however, direct biochemical/genetic evidence is yet to be explored.

1.4 CRISPR: A game-changing genetic engineering technique

In the post-genomic era, researchers are often overwhelmed by the enormous amount of genomic information available as a result of high-throughput sequencing technologies (Liu *et al.*, 2012). Deciphering gene function and connecting genotype to phenotype has become a primary challenge in utilizing these resources to engineer biological systems to relieve and address global challenges such as environmental clean-up, clean energy production and human disease treatment. To date, a variety of available tools have been applied to create genetic modifications in many organisms (Esvelt *et al.*, 2013; Gaj *et al.*, 2013). However, the demand for genetic engineering is transforming from low-efficiency and time-consuming methods to efficient and fast ones, from targeting one site to multiple sites in a single genome for efficient genome-scale engineering (Esvelt *et al.*, 2013). The clustered, regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated proteins (Cas) system is an adaptive RNA-mediated immune system in approximately 40% of bacteria and ~90% of archaea (Marraffini *et al.*, 2010). The CRISPR/Cas system can be reprogrammed to reject invading bacteriophages and conjugative plasmids (Carroll, 2012; Jinek *et al.*, 2012). Continued improvement in understanding of the mechanisms of the type II CRISPR/Cas system launched the birth of novel programmable CRISPR/Cas9-based platforms, native Cas9 nuclease (Cas9) or Cas9 nickase (Cas9n)-based targeted genome editing (Cho *et al.*, 2013; Cong *et al.*, 2013; Dicarlo *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013; Mali *et al.*, 2013; Ran *et al.*, 2013; Wang *et al.*, 2013; Yang *et al.*, 2013) and inactivated or

dead Cas9 (dCas9)-based transcriptional control (Bikard *et al.*, 2013; Gilbert *et al.*, 2013; Qi *et al.*, 2013). Cas9-based tools, thus far, have been successfully applied in diverse organisms, showing a great promise to realize multiplex and efficient genome editing and regulation of gene expression without host dependence. Here, we review the molecular basis of the type II CIRSPR/Cas system, application of Cas9-based tools, and factors influencing their utilization. We also compare the advantages and limitations of Cas9-based tools with several widely-used targeted tools, such as Zinc-finger nucleases (ZFNs) (Pabo *et al.*, 2001; Gaj *et al.*, 2013), and transcription activator-like effector nucleases (TALEs) (Mussolino *et al.*, 2011; Gaj *et al.*, 2013).

1.4.1 Bacterial CRISPR/Cas system

The CRISPR/Cas system as an adaptive immune system (Horvath *et al.*, 2010), employs CRISPR RNA (crRNA)-guided Cas proteins to recognize target sites (known as protospacers) within the invader genome via base-pairing complementarity and then cleaves DNA within the protospacer sequences. It is classified into three types (I, II and III) based on the sequence and structure of the Cas protein (Makarova *et al.*, 2011; Makarova *et al.*, 2011). The crRNA-guided surveillance complexes in types I and III need multiple Cas subunits (Zhang *et al.*, 2012; Sinkunas *et al.*, 2013); however, type II only requires Cas9 (Deltcheva *et al.*, 2011; Sapranauskas *et al.*, 2011). The type II system as a reduced system has been primarily studied in *Streptococci* (Figure 1.3) (Deltcheva *et al.*, 2011; Gasiunas *et al.*, 2012) and *Neisseria* (Zhang *et al.*, 2013), and the former has been developed as a robust programmable tool (Figure 1.4). The native type II system requires at least three crucial components: RNA-guided Cas9 nuclease, crRNA and a partially complementary trans-acting CRISPR RNA (tracrRNA)

(Deltcheva *et al.*, 2011; Gasiunas *et al.*, 2012). Each of these components is discussed below.

Cas9 nuclease. Cas9 is the first indispensable component of type II CRISPR/Cas systems and is able to cleave double-stranded DNA (dsDNA) in a sequence-specific manner (Deltcheva *et al.*, 2011; Makarova *et al.*, 2011; Esvelt *et al.*, 2013). Although there are other *cas* genes (eg. *cas1*, *cas2*, and *csn2*) present in a single genome, disruption of these other genes did not impair crRNA biogenesis (Deltcheva *et al.*, 2011; Sapranauskas *et al.*, 2011; Jiang *et al.*, 2013). Cas9 is a large multi-domain protein with two nuclease domains, a RuvC-like nuclease domain near the amino terminus and an HNH (or McrA-like) nuclease domain in the middle (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). *In vitro* tests indicate that the endonuclease activity of the *S. pyogenes* Cas9 creates blunt dsDNA breaks (DSBs) that are 3 bp upstream of the 3' terminal complementarity region formed between the crRNA recognition sequence and the genomic protospacer (Figure 1.4A) (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). Mutagenesis of each catalytic site in the RuvC and HNH motifs abolished the ability to create DSBs, leaving only nickase activity. Biochemically, the RuvC (D10A for *S. pyogenes* Cas9, D31A for *S. thermophilus* Cas9) and HNH mutants (H840A for *S. pyogenes* Cas9, N891A for *S. thermophilus* Cas9) cut the non-complementary and complementary strands, respectively, of the protospacer at the same position as the intact Cas9–crRNA complex (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012; Cong *et al.*, 2013), indicating that each active site acts on the opposite DNA strand to generate DSBs. Intriguingly, mutations in these active sites did not alter the affinity of the CRISPR/Cas complex for binding the protospacer (Gasiunas *et al.*, 2012). Importantly,

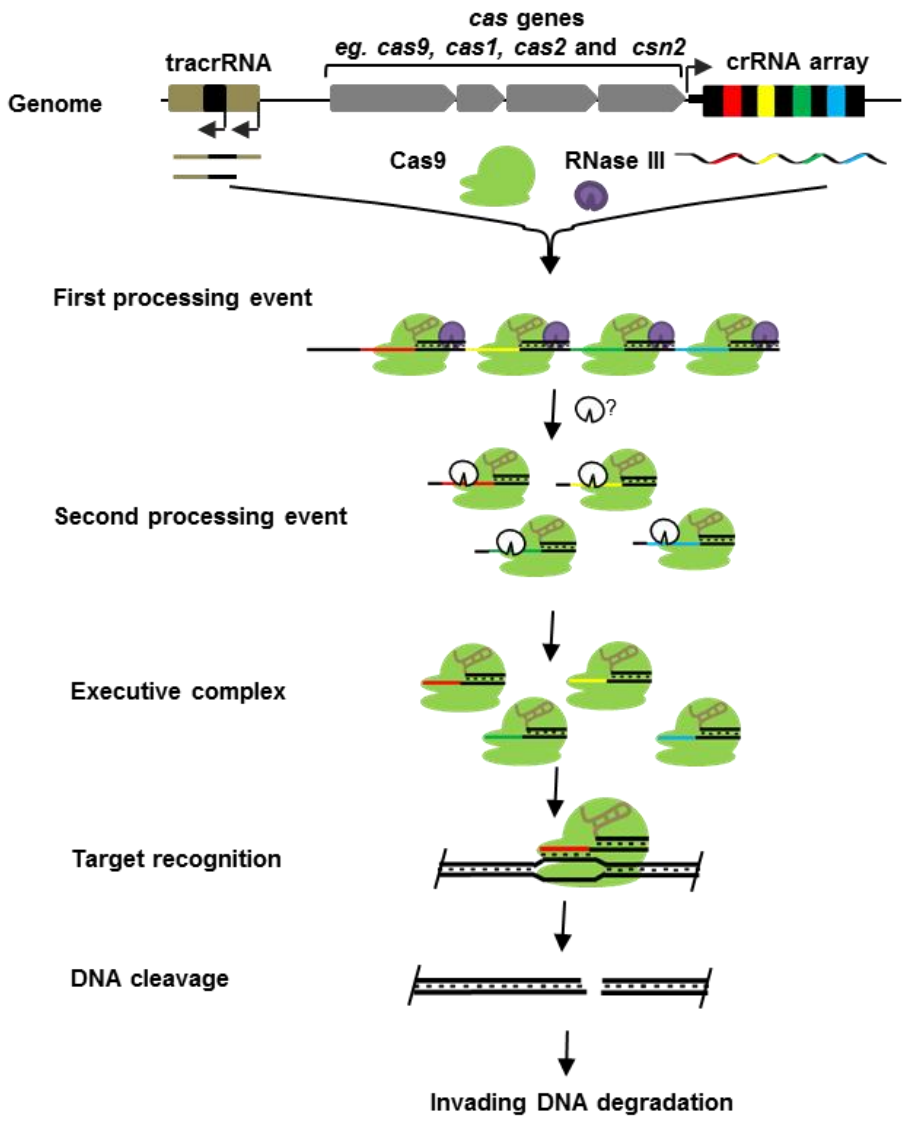


Figure 1.3 Type II CRISPR/Cas systems in Streptococci. The type II system needs three major steps to accomplish target DNA cleavage. First, tracrRNA precursor and pre-crRNA transcripts are processed by RNase III in the presence of Cas9 to split the crRNA array and shorten the tracrRNA precursor within the complementation region. Second, the spacer region of crRNA is further trimmed by unknown RNases to produce mature crRNA with a 20 nt target recognition region. Third, tracrRNA-crRNA duplex is incorporated into Cas9 forming an executive complex to specify protospacers and create DSBs to degrade invading DNA.

protospacer adjacent motifs (PAMs) that are short conserved nucleotide stretches next to the protospacers, such as NGG (van der Ploeg, 2009), NGGNG (Horvath *et al.*, 2010), NAAR (van der Ploeg, 2009) or NNAGAAW (Deveau *et al.*, 2008), are absolutely necessary for Cas9 binding and cleavage (Gasiunas *et al.*, 2012). Orthogonal Cas9 nucleases from different microorganisms require different PAM sequences (Bhaya *et al.*, 2011; Esvelt *et al.*, 2013).

tracrRNA. The tracrRNA is the second indispensable component of the type II CRISPR/Cas system and is a non-protein coding RNA for crRNA maturation and subsequent DNA cleavage (Karvelis *et al.*, 2013). In *S. pyogenes*, the tracrRNA gene is transcribed from two start sites producing two primary species of 171 nt and 89 nt, both of which are processed into ~75 nt RNA species (Figure 1.3) (Deltcheva *et al.*, 2011). The resulting tracrRNA precursors have a stretch of almost perfect (one mismatch) complementarity with each of the pre-crRNA repeats. The base-pairing RNA duplex is important for tracrRNA precursor trimming and crRNA maturation as mentioned below (Deltcheva *et al.*, 2011; Chylinski *et al.*, 2013).

crRNA biogenesis in type II systems. Recent studies uncovered different crRNA maturation processes in type II systems (Bhaya *et al.*, 2011; Sorek *et al.*, 2013). *S. pyogenes* produces only one form of the full-length primary pre-crRNA with 511 nt, consisting of a leader region and a number of repeat-spacer-repeat units (Deltcheva *et al.*, 2011). Then, a two-step crRNA biogenesis is used (Figure 1.3), with a first cleavage within the repeat regions and a second cleavage within the spacers. During the first cleavage, the base-pairing RNA duplex formed by the tracrRNA precursor and the pre-crRNA is attacked by the housekeeping RNase III within the repeats, generating a

75 nt tracrRNA and a 66 nt intermediate crRNA species. The coordinated action of RNase III and Cas9 is necessary to process the duplex and the complementarity of the duplex is a prerequisite for the RNase III-mediated co-processing (Deltcheva *et al.*, 2011; Karvelis *et al.*, 2013). The second cleavage is assumed to depend on the Cas9-mediated ruler-type mechanism whereby the spacers are cleaved at a fixed distance using the first processing site as an anchor, generating 39-42 nt mature species carrying a unique 20 nt spacer sequence and a 19-22 nt repeat sequence (Deltcheva *et al.*, 2011). These processed RNA components are assembled with Cas9, forming executive nucleoprotein complexes that target and cleave the protospacer recognized by 20 nt spacer sequences in crRNAs.

1.4.2 Application of type II CRISPR/Cas system

Due to the simplicity and customizability of type II CRISPR systems, host-independent gene-targeting platforms has been developed for genome editing and transcriptional control in both eukaryotes and prokaryotes (Bikard *et al.*, 2013; Cho *et al.*, 2013; Dicarlo *et al.*, 2013; Dickinson *et al.*, 2013; Hwang *et al.*, 2013; Jiang *et al.*, 2013; Jiang *et al.*, 2013; Karvelis *et al.*, 2013; Mali *et al.*, 2013; Nakayama *et al.*, 2013; Yang *et al.*, 2013). In general, current applications of type II systems can be classified into three categories: native Cas9-mediated genome editing; Cas9 nickase-mediated genome editing; and inactivated Cas9-mediated transcriptional control. Promisingly, type II systems can also be engineered for high-throughput genome editing and silencing.

Native Cas9-mediated genome editing. Cas9-mediated genome editing depends on two sequential steps occurring in the cells (Figure 1.4B). First, genomic DNA is cleaved by Cas9 at a specific site determined by the 20 nt target recognition

sequence in crRNA (Deltcheva *et al.*, 2011; Jinek *et al.*, 2012). Second, the resulting double-strand DNA breaks (DSB) are ligated by native DNA repair systems (Wyman *et al.*, 2006), native non-homologous end joining (NHEJ) (Shuman *et al.*, 2007), or template-dependent homology-directed repair (HDR) (Smith, 2001). NHEJ, as an error-prone process, often generates undefined small insertions and deletions (indels) during the repair process (Cong *et al.*, 2013; Dicarlo *et al.*, 2013; Li *et al.*, 2013), presumably resulting in malfunction of targeted genes. When an editing template with homologous flanking arms was used, the DSB could be precisely repaired by HDR, generating defined deletions, insertions, and nucleotide substitutions (Cong *et al.*, 2013; Gratz *et al.*, 2013; Jiang *et al.*, 2013; Yang *et al.*, 2013).

To utilize type II CRISPR/Cas systems, three components, including the Cas9 protein, tracrRNA and customized crRNA, need to be expressed in foreign hosts. Even though *S. pyogenes* RNase III has been reported to be an indispensable component involved in crRNA maturation (Deltcheva *et al.*, 2011), reports showed it was not necessary in a number of diverse heterogeneous systems (Cong *et al.*, 2013; Dicarlo *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013). A plasmid-based CRISPR/Cas system was established to edit the *E. coli* genome using two plasmids: pCas9 expressing tracrRNA and Cas9, and pCRISPR expressing the crRNA array (Jiang *et al.*, 2013). Through using customized 20 nt target recognition sequences in a crRNA array, double deletion and/or multiplexed editing has been achieved in a single step (Cong *et al.*, 2013; Jiang *et al.*, 2013). Thus, three-component CRISPR/Cas9 systems are convenient to realize targeted multiplexed editing by only programming the crRNA array.

The tracrRNA:crRNA duplex has been engineered into one molecule, called a chimeric guide RNA (gRNA), with a length of 39-40 nt containing a 20 nt target recognition sequence at the 5' end followed by a hairpin structure (or gRNA scaffold) retaining the base-pairing interactions within the tracrRNA:crRNA duplex (Figure 1.4A) (Jinek *et al.*, 2012; Sinkunas *et al.*, 2013; Sorek *et al.*, 2013). This progress further simplified the application of type II CRISPR systems in genome editing. Researchers have successfully edited the genomes of many organisms (e.g., human cells, mice, zebrafish, yeast, Arabidopsis, rice, tobacco, *E. coli,* and many others) by co-expressing Cas9 nuclease and customized gRNAs from expression vectors or by delivering RNA transcripts (Figure 1.4B) (Dicarlo *et al.*, 2013; Jiang *et al.*, 2013; Jiang *et al.*, 2013; Mali *et al.*, 2013; Wang *et al.*, 2013). By designing DNA donor templates, multiple point mutations (Dickinson *et al.*, 2013), site-specific recombination sites (loxP and attP) (Chang *et al.*, 2013; Dickinson *et al.*, 2013; Gratz *et al.*, 2013), endogenous protein tagging (Dickinson *et al.*, 2013) and expression cassettes of green fluorescent protein (Mali *et al.*, 2013) have been successfully introduced into the targeted genome loci. The Cas9-gRNA complex has been used to simultaneously disrupt five genes in a single genome (Wang *et al.*, 2013). Therefore, the CRISPR/Cas9 system is an efficient tool to edit genomes with wide applications in a broad range of hosts.

Cas9 nickase-mediated genome editing. gRNA-guided Cas9n with a RuvC or HNH mutation has the ability to create a nick instead of a DSB at the target site (Gasiunas *et al.*, 2012; Cong *et al.*, 2013; Ran *et al.*, 2013). Although individual nicks are predominantly repaired by the high-fidelity base excision pathway (Dianov *et al.*, 2013), the combination of nick generation and HDR has successfully edited genomes at

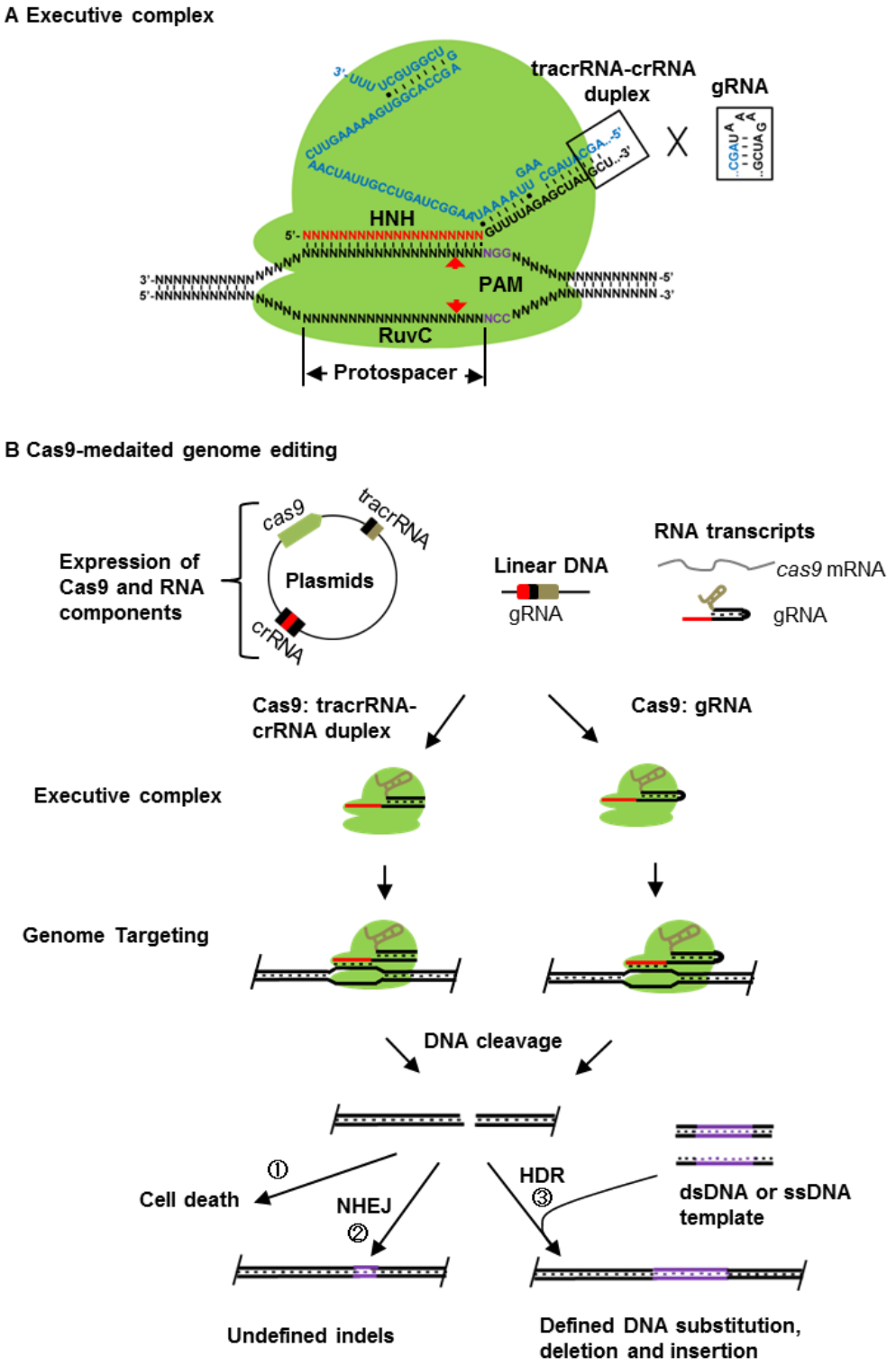


Figure 1.4 Application of CRISPR/Cas9 in targeted genome editing. The widely-used S. pyogenes Cas9 with HNH and RuvC domains are directed by tracrRNA-crRNA duplexes or gRNA (A) to cut the complementary and non-complementary strands, respectively. Cuts are made at the positions (indicated by red arrows) that are 3 bp

upstream of PAM sites (purple characters). All components required for RNA-guided genome editing in foreign hosts are expressed by delivering co-expression plasmids, DNA expression cassette fragments, or sole RNA transcripts (B). Expressed tracrRNA-crRNA duplexes or gRNA are assembled with Cas9, generating executive complexes. These complexes generate breaks in the genome that may lead to cell death if the DSBs are not removed (①), or induce error-prone nonhomologous end joining (NHEJ) to rejoin the ends and introduce undefined small deletions and additions (indels) (②), or trigger homology-directed repair (HDR) when homology-containing dsDNA or ssDNA templates are given (③), to confer precise DNA substitution, deletion or insertion.

the intended site (Cong *et al.*, 2013). Introduction of a double nick using a pair of gRNA-guided Cas9n's targeting the opposite strands of the target site has been successfully applied to generate DSBs and NHEJ-induced mutations (Mali *et al.*, 2013; Ran *et al.*, 2013). A paired nicking strategy was reported to facilitate high-efficiency HDR at levels comparable to those of native Cas9-mediated HDR and at significantly higher rates than single Cas9n-mediated HDR (Ran *et al.*, 2013). Interestingly, this paired nicking significantly reduced off-target cleavages by 50- to 1,500-fold in human cells, but without sacrificing on-target cleavage efficiency (Ran *et al.*, 2013). Additionally, creating a pair of double nicks at two sites by four customized gRNAs successfully deleted genomic fragments up to 6 kb (Ran *et al.*, 2013). Thus, multiplex nicking created by Cas9n has the ability to create high-precision genome editing.

Inactivated Cas9-based transcriptional control. CRISPR/Cas systems have also been developed as an innovative facile and multiplex approach for transcriptional control without altering the target gene sequence, this is called CRISPR interference (CRISPRi) (Figure 1.5) (Qi *et al.*, 2013). It consists of a completely inactive dCas9 and a custom gRNA (or tracrRNA:crRNA duplex). As mentioned before, dCas9 loses its endonuclease activity, but its ability to incorporate gRNA and bind to targets is not affected. Like RNA interference (RNAi), CRISPRi also depends on base-pairing

complementarity to recognize target sites. However, they apply different mechanisms to control gene expression. RNAi mainly causes transcript degradation and/or translation blocking (Wilson *et al.*, 2013), but CRISPRi blocks transcription initiation and elongation. Qi et al reported the mechanism of CRISPRi and its initial applications in efficiently repressing the expression of targeted genes in *E. coli* and human cells (Qi *et al.*, 2013). Through co-customizing several gRNAs, simultaneous regulation of multiple genes became possible. dCas9-mediated transcriptional control also has been tested in *S. pneumonia* (Bikard et al., 2013), and silencing effects can be induced and reversed using an anhydrotetracycline-inducible promoter to drive dCas9 and gRNA expression (Qi *et al.*, 2013). The repression efficiency varied (10-300 fold) depending on several major factors, which will be discussed below. Combining two gRNAs targeting the same gene could produce up to 1,000-fold repression (Qi *et al.*, 2013). Therefore, the CRISPRi targeting platform holds promise as a general approach for modulating gene expression at the transcriptional level.

Like a variety of ZFNs and TALENs that were generated by coupling specific DNA binding domains with different, non-specific effectors (Minczuk *et al.*, 2006; Li *et al.*, 2007; Miller *et al.*, 2011), dCas9 also has been fused with transcription effectors, generating chimeric dCas9-effector proteins (Figure 1.5) (Gilbert *et al.*, 2013; Mali *et al.*, 2013). The consequence caused by the chimera depends on effector functions since the major role of gRNA-guided dCas9 is just to recognize and localize the chimera. KRAB, a repressive chromatin modifier domain, was grafted onto dCas9 and presented significantly higher repression efficiency than dCas9 by itself in HEK293 cells (Gilbert *et al.*, 2013). In addition, dCas9-activator proteins, like dCas9-VP64 and dCas9-p65AD,

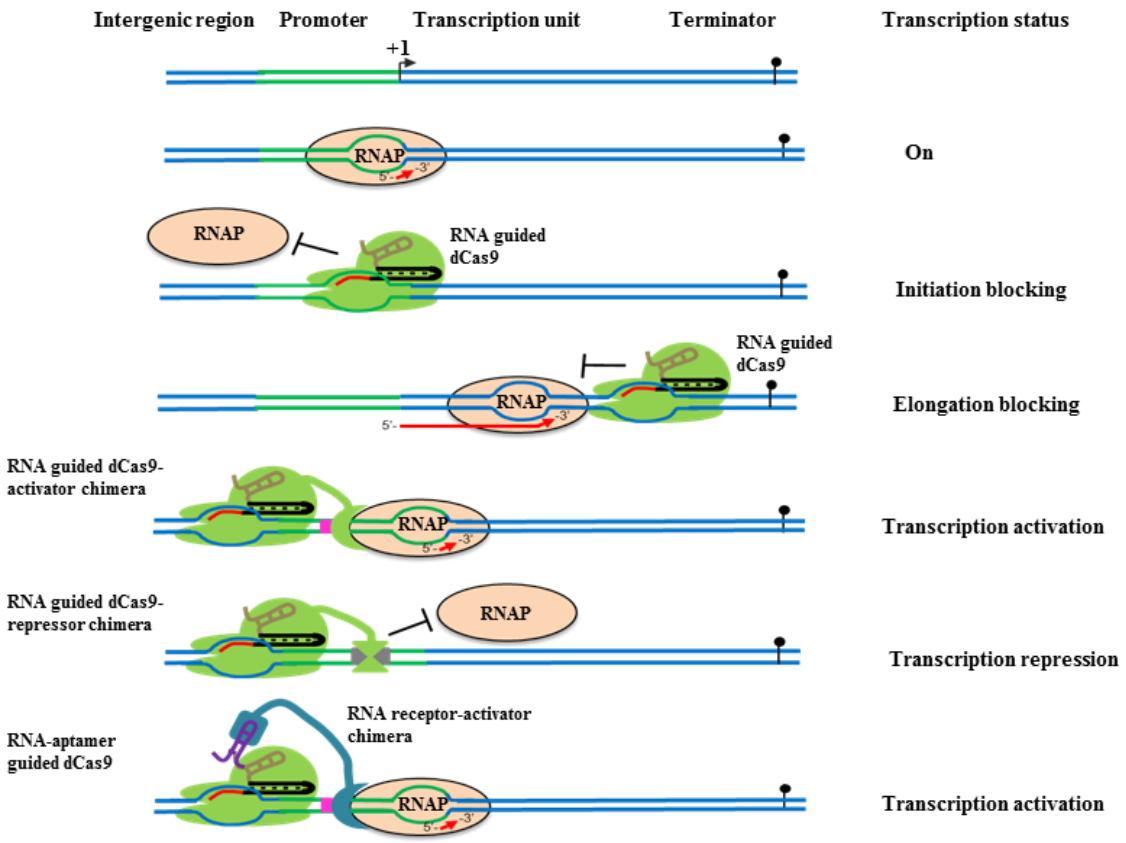


Figure 1.5 Application of engineered dCas9 and/or RNA components in transcriptional control. RNA polymerase (RNAP) initiates transcription within the promoter region; however, the binding of RNA-guided dCas9 to the promoter region and the encoding region could block transcription initiation and transcription elongation, respectively, leading to the repression of gene expression at the transcriptional level. Through fusing dCas9 with transcriptional activators or repressors, the positioning function of gRNA or crRNA molecules will direct the dCas9-effector chimera to bind to the promoter vicinity and then the effector modules will stimulate or repress gene transcription by interacting with DNA motifs or RNAP. Also, gRNA or crRNA could be fused with RNA aptamers generating chimeric RNA that will direct dCas9 to bind to specific sites, allowing localization of specific RNA receptors. Generation of RNA receptor-activator/-repressor chimera will lead to activator or repressor localization, followed by expression activation or repression of neighboring genes.

exhibited up to 25-fold increase in gene expression (Gilbert *et al.*, 2013). In *E. coli*, activation of gene expression was realized by fusing dCas9 to the ω subunit of RNA polymerase (Bikard *et al.*, 2013). Also, by tethering customized gRNA with the MS2

bacteriophage coat protein-binding RNA stem loop, a MS2-VP64 fusion protein was localized to the target site by the dCas9 complex and then stimulated gene expression (Figure 1.5) (Mali *et al.*, 2013). Thus, the dCas9-gRNA complex has a large potential for the design of sequence-specific transcriptional regulation in different organisms, and potentially for diverse epigenetic investigation.

Cas9-based high-throughput genetic screen. People are also interested in developing CRISPR/Cas9-derived platforms for genetic studies. It is highly possible to use multifunctional Cas9 variants to create mutant libraries for screening and identifying genome-scale phenotype-related genetic elements (Larson *et al.*, 2013; Mali *et al.*, 2013). For high-throughput targeting, the key is to construct high-specificity gRNA libraries. The rules applied to select genome-wide targetable sites have been discussed (Dicarlo *et al.*, 2013; Larson *et al.*, 2013; Xie *et al.*, 2013). This strategy was successfully applied in genetic screening in human cells. If using dCas9 or dCas9-effector chimera, knock-down or activation mutant libraries will be generated. Compared to loss-of-function mutant libraries, the knock-down or activation mutant libraries have an unmatched advantage for studying lethal genes.

1.4.3 Influential factors of CRISPR/Cas application

Thus far, numerous studies have examined the diverse factors that impact the efficiency and/or specificity of Cas9-based tools, such as Cas9 activity, the length and structure of RNA components, Cas9:gRNA ratio, and RNA-target complementarity extent and complementary position. Discussion of these factors will help direct future experiments using CRISPR and improve performance.

Cas9 is a pivotal component. Mutation of catalytic sites, incorrect subcellular localization or inappropriate Cas9 dosage all affects genome editing. In eukaryotic cells, prokaryote-derived Cas9 is generally fused with a nuclear location signal (NLS) at the N- or C-terminus, or both, to direct protein translocation into the nucleus (Dicarlo *et al.*, 2013; Gratz *et al.*, 2013; Hwang *et al.*, 2013; Mali *et al.*, 2013; Qi *et al.*, 2013; Wang *et al.*, 2013). Codon optimization is also necessary for producing functional Cas9 in heterogeneous expression systems (Li *et al.*, 2013; Mali *et al.*, 2013; Nakayama *et al.*, 2013). The ratio of Cas9 to gRNA greatly affected mutagenesis efficiency (Li *et al.*, 2013; Nakayama *et al.*, 2013; Wang *et al.*, 2013). Theoretically, the more complexes are formed, the higher editing efficiency is expected to be. However, a potential risk that accompanies excessive executive complex availability is the off-target effect due to the unavoidably low complementarity of non-specific regions in the genome (Fu *et al.*, 2013). To overcome these issues, we need to control component expression, improve target selection criteria and engineer the Cas9 protein to provide higher specificity.

Another major class of determinants is the RNA components. The gRNA chimera exhibits comparable efficiency to the tracrRNA:crRNA duplex in *in vitro* plasmid cleavage assays (Jinek *et al.*, 2012). gRNAs presented higher efficiency than RNA duplexes in rice plant but conversely in human and mouse cells (Cong *et al.*, 2013; Esvelt *et al.*, 2013; Miao *et al.*, 2013). Some undetermined cellular factors or RNA features might influence editing efficiency. In addition, base-pairing is critical to the folding structure of gRNAs. Elongation of the self-complementation region in gRNAs enhanced site-specific NHEJ-mediated mutagenesis (Jinek *et al.*, 2013).

crRNA and gRNA molecules harboring target recognition sequences determine target specificity, as such, the selection of target protospacers is a critical issue. A protospacer within an N(21)GG format (or N20+ NGG) is widely used for *S. pyogenes* Cas9 targeting. This protospacer contains a 20 nt base-pairing region immediately followed by a PAM (NGG). The amount of base-pair complementarity between target recognition sequences and protospacers is of importance to Cas9-based editing efficiency and dCas9-based transcriptional control. Extension of the 5' end of the gRNA target recognition region to increase base-pairing complementarity with a protospacer did not improve either editing efficiency or targeting specificity (Ran *et al.*, 2013). Several studies reported that mismatches occurring in the 3' half of the gRNA severely affected Cas9-mediated cleavage (Semenova *et al.*, 2011; Jinek *et al.*, 2012; Jiang *et al.*, 2013; Qi *et al.*, 2013). The same position within different targeting sequences presented varying importance, and not all mismatches in the 5′ half of the gRNA were well tolerated (Fu *et al.*, 2013). For double nicking strategy, the relative positions of the gRNA pairs with offsets from -4 to 20 bp were most efficient to induce NHEJ (Gratz *et al.*, 2013) and introduction of 5' overhangs created by offset nicks stimulated more robust NHEJ and HDR events than that of 3' overhangs (Mali *et al.*, 2013; Ran *et al.*, 2013). For CRISPRi, dCas9 also presented similar rules to maintain silencing efficiency (Qi *et al.*, 2013).

The above discussion focuses on the determinants of DNA cleavage, which is the most critical step in introducing frameshift mutations to a specific genome site by the error-prone NHEJ. Another way to resolve DSBs is to stimulate HDR by providing editing templates, which are single-stranded DNAs (ssDNAs) or dsDNA fragments with

homologous flanking arms. DSB generation can increase homologous recombination rates of ssDNA and dsDNA donors by 5-fold and 130-fold, respectively (Dicarlo *et al.*, 2013). During recombination, editing templates should not overlap with crRNA/gRNA target recognition sequences, which might decrease editing efficiency (Dicarlo *et al.*, 2013). If multiple template DNAs are co-transformed with plasmids expressing Cas9 and gRNAs targeting multiple sites, a single-step double or more deletions could be generated as desired (Jiang *et al.*, 2013). However, some factors potentially affecting HDR remain to be evaluated, including the size and position of the homologous flanking arms and the stability of the given templates before HDR occurs.

CRISPRi has been systematically studied (Qi *et al.*, 2013) and several factors, in addition to the ones aforementioned, have been identified as influencing the dCas9-based transcriptional control. First, CRISPRi-mediated blocking of transcriptional elongation presents strand specificity (Bikard *et al.*, 2013; Qi *et al.*, 2013). gRNAs targeting a non-template DNA strand presented much higher repression efficiency than those targeting the template strand. Second, the silencing efficiency is inversely correlated with the distance of the target from the translation start codon. Third, an augmentative silencing effect may be observed when two or more gRNAs bind to separate target sites on the same gene (Mali *et al.*, 2013; Qi *et al.*, 2013); however, if they bind to overlapping regions, repression is suppressed. To block transcriptional initiation in *E. coli*, the -35 box-containing regions chosen as gRNA targets are more efficient than other adjacent regions. For dCas9-effector dependent transcriptional regulation, performance also presented positional and accumulative effects (Esvelt *et al.*, 2013; Mali *et al.*, 2013).

1.4.4 Comparison of targeted genetic engineering tools

A wide variety of tools are available for editing targeted genomes and regulating gene expression. Based on target recognition mechanisms, they can be grouped into two major classes, protein-directed or nucleotide-directed specificity (Esvelt *et al.*, 2013). Recombinases, integrases, ZFNs and TALENs, are well-known approaches that depend on protein-directed specificity; RNA interference (RNAi), group II intron retrotransposition and the innovative Cas9-based platforms, rely on nucleotide-directed specificity. All of these have been widely used in prokaryotes and/or eukaryotes. This section will discuss the advantages and limitations of these widely-used tools in terms of their flexibility, multiplex targeting potential, and targeting efficiency and specificity.

Generally, protein-directed specificity is comparatively harder to customize than nucleotide-directed specificity. Recombinases and integrases require suitable pre-existing recognition sites in the genome and often have some inherent application limitations (Groth *et al.*, 2004; Wang *et al.*, 2011; Esvelt *et al.*, 2013). Both ZFNs and TALENs are generated by coupling a customized DNA binding domain with a non-specific nuclease domain (Mussolino *et al.*, 2012; Gaj *et al.*, 2013). The DNA binding domain of ZFNs and TALENs is a tandem array of zinc finger (ZF) motifs and transcription activator-like (TAL) repeats, respectively (Urnov *et al.*, 2010; Mussolino *et al.*, 2012). However, it is difficult and expensive to customize ZFs or TALs by protein engineering, and if using *Fok*I nuclease domain, two ZFNs or TALENs must be customized for each new target site (Pabo *et al.*, 2001; Gaj *et al.*, 2013). Also, ZFN and TALEN activity is affected by numbers of factors. Even though ZFNs and TALENs

have already generated extensive modifications, they are difficult to apply to the creation of multiple mutations in a single genome via step-wise mutagenesis.

As tools based on nucleotide-directed specificity, RNA-directed RNAi, group II intron retrotransposition and Cas9-based methods only require DNA synthesis or PCR amplification to retarget, so obviously these methods are more convenient and economical. RNAi is mostly used to repress gene expression in both prokaryotes and eukaryotes instead of knocking of them out. Although RNAi also can be used to target multiple genes, sometimes the need for long target sites and amplification of small interference RNAs can sometimes result in severe off-target effects (Jackson *et al.*, 2003; Maida *et al.*, 2013). Group II intron retrotransposition is widely applied to inactivate genes in bacterial genomes (Enyeart *et al.*, 2013; Esvelt *et al.*, 2013). Cas9-based tools can be used in diverse applications, as mentioned above. All of the essential components required by these tools can be expressed by delivering plasmids (Cong *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013; Shan *et al.*, 2013), linear DNA expression cassettes (Dicarlo *et al.*, 2013) or RNA transcripts (Hwang *et al.*, 2013; Waaijers *et al.*, 2013; Wang *et al.*, 2013; Yang *et al.*, 2013). In addition, bioinformatic analysis of genome-wide target sites (N21GG) revealed that most genes or exons can be targeted specifically in Arabidopsis (Li *et al.*, 2013), rice (Xie *et al.*, 2013), and yeast (Dicarlo *et al.*, 2013). Therefore, Cas9-based genome editing provides a highly flexible and programmable method.

The ability to multiplex targeting is another notable advantage that Cas9-based tools have. Efficient methods enabling multiplex genome editing are urgently needed for genome-scale engineering. Several reports demonstrated the creation of

simultaneous multiple mutations with Cas9-based tools (Cong *et al.*, 2013; Wang *et al.*, 2013). To realize multiplexed editing, the only things required are the construction of crRNA arrays that produce various crRNAs, or constructing several different chimeric gRNAs to direct Cas9 to edit multiple targets at the same time. In this way, as many as five gene mutations have been generated simultaneously in mouse embryonic stem cells with high efficiency (Wang *et al.*, 2013). In addition, using gRNAs to direct mutated dCas9 to specifically target transcriptional regions of two different genes, the expression of both targeted genes was simultaneously decreased (Qi *et al.*, 2013). Then, multiple genes were activated or repressed at the transcriptional level by coupling dCas9 with transcriptional effectors, or fusing gRNA with recognizable RNA aptamers (Bikard *et al.*, 2013; Mali *et al.*, 2013; Qi *et al.*, 2013). Thus, versatile Cas9-based tools hold promise to realize both multiplexed genome editing and transcriptional control, avoiding tedious step-wise genetic manipulations.

Targeting efficiency and specificity greatly impacts the application potential of targeted tools. The editing efficiency of Cas9-based tools varies greatly among different organisms, cell types and mutation types and even target sites. HDR-mediated insertion occurred at an efficiency of 100% in *S. pneumonia*, 64% in *E. coli* (Jiang *et al.*, 2013) and 100% in *S. cerevisiae (Dicarlo et al., 2013)*. Cas9-mediated genome editing in human cells and zebrafish embryos produced efficiencies similar to those obtained using ZFNs and/or TALENs (Cong *et al.*, 2013; Hwang *et al.*, 2013). To date, Cas9-based tools have presented the ability to delete 6 kb genomic fragments (Ran *et al.*, 2013) and insert up to 3 kb of DNA into the intended genomic locus (Yang *et al.*, 2013). However, for application in synthetic biology, the potential of delivering larger DNA

fragments still needs to be evaluated. Off-target activity, which potentially produces misleading conclusions, is a big challenge to all targeted tools. Cas9-based tools face this same problem (Fu *et al.*, 2013; Xie *et al.*, 2013). TALENs appear to have lower off-target activity than ZFNs (Mussolino *et al.*, 2011). Cas9-gRNA complexes and 18-mer TAL effectors can potentially tolerate 1-3 and 1-2 target mismatches, respectively (Fu *et al.*, 2013; Mali *et al.*, 2013). Further studies with Cas9-gRNA complexes revealed that the frequency of off-target cleavage was sometimes the same as for on-target frequency (Fu *et al.*, 2013). Cas9n was reported to greatly reduce off-target effects without sacrificing the efficiency of HDR induction (Cong *et al.*, 2013). To improve the efficiency and specificity of Cas9-based tools, much effort needs to be made on Cas9 engineering, optimizing gRNA selection rules, and further elucidating Cas9-gRNA recognition features.

In summary, Cas9-based tools possess notable advantages that current, widely-used targeted tools cannot match. These tools will greatly enhance our ability to engineer and edit genomes and regulate gene expression in diverse organisms. These technologies also pave the way to easily dissect individual gene functions and are expected to accelerate the *in vivo* study of functionally redundant genes and epigenetic investigations, and will enable a broad range of research and applications in diverse biological fields, biotechnology, metabolic engineering and medicine. The ability to do multiplex targeting will revolutionize genome-scale engineering by providing a method for multiple disruptions, insertions and deletions at high efficiency and low cost.

1.5 Aim and focus of the study

As aforementioned, the production and utilization of lignocellulosic biofuels will bring a far-reaching positive impact on energy sustainability, environment protection, and even human health. To make biofuels competitive with conventional fuels, it is imperative for microbiologists to understand and then engineer biological processes in microorganisms that are capable of consolidated bioprocessing. Studies on *C. cellulolyticum*, which is a model organism of mesophilic cellulolytic Clostridia and a consolidated bioprocessing-enabling candidate, will potentially bring more instructional significance and application values than just studying lignocellulose degrading or sugar fermenting bacteria. However, the functional characterization of interesting genes in many lesser-studied microbes, including *C. cellulolyticum*, has been widely hindered due to the lack of efficient and precise genome editing tools. Moreover, while many lignocellulose degrading enzymes of *C. cellulolyticum* have been characterized in vitro, we have insufficient knowledge of the physiological importance of many cellulosomal components and the regulatory mechanism of key genes associated with extracellular lignocellulose hydrolysis, sugar assimilation, and intracellular metabolism. This study aimed to: 1) adapt the bacterial CRISPR/Cas9 system to edit the genome of *C. cellulolyticum* in a sequence-specific manner; 2) develop a Cas9 nickase-based platform to stably manipulate essential metabolic genes; 3) characterize the role of a cellulosomal protease inhibitor in cellulose degradation; 4) examine the role of carbon catabolite regulation in carbohydrate utilization, including cellulose and a variety of simple sugars. Major results of this study are presented in the following four chapters (2-5).

Chapter 2 presents the successful use of *S. pyogenes* Cas9 nickase, instead of native Cas9, to edit the *C. cellulolyticum* genome. First, a synthetic promoter was created, evaluated and then employed to drive the expression of the Cas9 system in a single plasmid. Second, all Cas9-based strategies (i.e., Cas9-NHEJ, Cas9-HR, Cas9 nickase-NHEJ and Cas9 nickase-HR) were experimentally tested but only Cas9 nickase-HR succeeded in genome editing. Third, the editing efficiency, accuracy, and versatility were evaluated systematically.

Chapter 3 presents the development of Cas9 nickase-assisted RNA repression for stable genetic manipulation on essential acetate-producing genes in *C. cellulolyticum*. First, plasmid-based expression of antisense RNAs was employed to knockdown gene expression, the effectiveness of which was evaluated at the enzymatic and metabolic levels. Second, we compared the repression efficacy between plasmid transformants and chromosomal integrants, and also experimentally evaluated the dramatic difference in gene expression between plasmid-based and chromosome-based expression. Then, chromosome-based repression was improved by a tandem promoter which was integrated by the Cas9-nickase genome editing tool in a single step.

Chapter 4 presents the functional characterization of a dockerin-containing protease inhibitor gene (*dpi*) (Ccel_1809) in *C. cellulolyticum* H10. First, the *dpi* mutant was generated and characterized at the phenotypic, physiological and protein levels. Then, we evaluated the in vivo importance of two cellulosomal components (Cel48F and Cel9E), which were highly associated with the functioning of Dpi, by mutagenesis and growth profiling on cellulose. Finally, the purified His-tagged Dpi was characterized in terms of inhibitory specificity and efficiency.

Chapter 5 presents systematic investigation into carbon catabolite regulation (CCR) in *C. cellulolyticum*. To begin with, bioinformatic prediction of all CCR components was performed, followed with the experimental evaluation of carbon catabolite repression. Then, mutants of the CCR components were generated and characterized mainly at the physiological level. Finally, microarray-based transcriptomic analysis was carried out in all knockout mutants in order to decipher how carbohydrate utilization is regulated by CCR in this bioenergy-related bacterium.

Chapter 2 Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase

2.1 Abstract

The CRISPR-Cas9 system is a powerful and revolutionary genome-editing tool for eukaryotic genomes but its use in bacterial genomes is very limited. Here we investigated the use of the *Streptococcus pyogenes* CRISPR-Cas9 system in editing the genome of *Clostridium cellulolyticum*, a model microorganism for bioenergy research. Wildtype Cas9-induced double-strand breaks were lethal to *C. cellulolyticum* due to the minimal expression of non-homologous end joining (NHEJ) components in this strain. To circumvent this lethality, Cas9 nickase was applied to develop a single nick-triggered homologous recombination strategy, which allows precise one-step editing at intended genomic loci by transforming a single vector. This strategy has a high editing efficiency (>95%) even using short homologous arms (0.2 kb), is able to markerlessly deliver foreign genes into the genome in a single step, enables precise editing even at two very similar target sites differing by two bases preceding the seed region, and has a very high target site density (median interval distance of 9 bp and 95.7% gene coverage in *C. cellulolyticum*). Together, these results establish a simple and robust methodology for genome editing in NHEJ-ineffective prokaryotes.

Keywords: CRISPR; Cas9; genome editing; *Clostridium cellulolyticum*

2.2 Introduction

Targeted genome editing is critical for both fundamental molecular biology and applied genetic engineering. Even though current methods (i.e., allele exchange, group II intron retrotransposition and recombineering) can be used for genome modification in many microbes (Esvelt *et al.*, 2013; Xu *et al.*, 2014), they have some limitations: i) traditional stepwise recombination-dependent allele exchange is time-consuming and low-efficiency (Heap *et al.*, 2012), which can be worse when host transformation efficiency is low and/or usable selection markers are limited; ii) insertion/deletion-based mutagenesis of large DNA fragments can potentially cause polar effects on downstream genes (Maamar *et al.*, 2004; Xu *et al.*, 2014); and iii) insertion of large DNA fragments, such as metabolic pathway transfer, are difficult with current genome engineering tools, which require existing recombination sites and/or recombinases (Enyeart *et al.*, 2013; Esvelt *et al.*, 2013). Thus, a facile and efficient method capable of performing precise, markerless and versatile genome manipulations is needed to expedite microbial studies.

The clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein (Cas) system is an RNA-guided immune system in many bacteria, that is able to recognize and cleave invasive DNAs (Barrangou *et al.*, 2007). The type II-A CRISPR-Cas system of *Streptococcus pyogenes*, which requires a mature CRISPR RNA (crRNA), a trans-activating crRNA (tracrRNA) and a DNA endonuclease Cas9, has been harnessed for targeted genome editing in many organisms (Jinek *et al.*, 2012; Cong *et al.*, 2013; Dicarlo *et al.*, 2013; Friedland *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013; Mali *et al.*, 2013). Mechanistically, under the guidance of the tracrRNA-crRNA duplex or latterly engineered single guide RNA (gRNA), *S.*

pyogenes Cas9 or Cas9 nickase (Cas9n) can cut any target DNA having a 5'-N20NGG-3' region (Figure S1.1A), where N represents any nucleotide and N20 represents the protospacer appended with a protospacer-adjacent motif (PAM) (NGG) at the 3' end (Jinek *et al.*, 2012). The cleavage site will then be repaired by non-homologous end-joining (NHEJ) or homologous recombination (HR) (Xu *et al.*, 2014; Selle *et al.*, 2015). Thus far, Cas9-based tools have shown their versatility for foreign gene knock-in and gene inactivation by DNA deletion or insertion, with attractive features such as ease of use, high efficiency, strong adaptability, and multiplex targeting ability (Xu *et al.*, 2014; Selle *et al.*, 2015). However, reports of their application in bacterial genome editing are quite limited (Jiang *et al.*, 2013; Cobb *et al.*, 2014; Oh *et al.*, 2014; Huang *et al.*, 2015; Jiang *et al.*, 2015; Tong *et al.*, 2015). By coupling Cas9-mediated cleavage with HR repair, the genomes of *Escherichia coli* (Jiang *et al.*, 2015), *Streptococcus pneumoniae* (Jiang *et al.*, 2013), four *Streptomyces* species (Cobb *et al.*, 2014; Huang *et al.*, 2015; Tong *et al.*, 2015) and *Tatumella citrea* (Jiang *et al.*, 2015), were edited at a high efficiency. Cas9-assisted elimination of unmutated cells, after single-stranded DNA recombineering, significantly improved the editing efficacy in *E. coli* and *Lactobacillus reuteri* (Jiang *et al.*, 2013; Oh *et al.*, 2014). Using the inefficient repair of double-stranded breaks (DSB) in some microbes, reprogrammed Cas9 has been applied as an antimicrobial to selectively kill some strains (Bikard *et al.*, 2014; Citorik *et al.*, 2014; Gomaa *et al.*, 2014). Naturally, the lethal effect of Cas9-induced DSB does not allow genome editing in repair-defective microbes, however, exploiting a strategy to circumvent this lethality will theoretically allow genome editing in many microbes.

As a model system of mesophilic cellulolytic bacterium, Clostridium *cellulolyticum* can directly convert lignocellulosic biomass to valuable end products (i.e., lactate, acetate, ethanol, hydrogen) (Desvaux, 2005). It holds promise of producing renewable green chemicals from cellulose to replace petroleum-based products (Lan *et al.*, 2013). However, genome editing of *C. cellulolyticum* for metabolic engineering is still challenging due to the lack of efficient editing tools. Despite the predicted presence of the type II-C CRISPR-Cas system in *C. cellulolyticum* (Chylinski *et al.*, 2014), without a basic understanding of this system (e.g., protospacer length, PAM and gRNA features), we cannot immediately examine its use in genome editing. Here we tested the use of the single gRNA-directed *S. pyogenes* Cas9 to edit the *C. cellulolyticum* genome and found an inefficiency of host NHEJ and HR in repairing Cas9-induced DSB. Then, we developed a single nick-assisted HR strategy using a Cas9 nickase and a plasmid-borne donor template to efficiently modify targeted genomic loci by DNA deletion and insertion. This strategy also presented the ability of markerlessly integrating foreign genes in a single step, making this a promising step in facilitating genome-level metabolic engineering coupled with synthetic biology in the future.

2.3 Materials and methods

2.3.1 Synthetic promoter design

Promoter sequences in the *C. cellulolyticum* genome were predicted by PePPER (de Jong *et al.*, 2012). Then over 100 predicted sigma^{A} promoters were aligned to create 39-nt long DNA logos using WebLogo (Crooks *et al.*, 2004). Based on the alignment result, at each position the nucleotide with the highest usage frequency was selected to build a mini P4 promoter (5'-

TTGACAAATTTATTTTTTAAAGT**TAAAAT**TAAGTTG-3’). To test promoter activity, P4 was used to drive an anaerobic fluorescent protein -encoding gene (*afp*). Between the P4 promoter and the *afp* open reading frame is a short sequence containing a ribosome RNA binding site (RBS) (5’-TT**AGGAGGT**ACCCCG-3’).

2.3.2 Plasmid construction

The P4 promoter generated by anneal extension PCR using P4F and P4R, was ligated into the pCR8/GW/TOPO TA vector (Invitrogen). The RBS-containing promoter fragment amplified by using PromF and PromR was assembled with a EcoRI- and BamHI-linearized pLyc017 backbone (Li *et al.*, 2014) using a Gibson assembly kit (NEB), generating pP4-AFP.

The *cas9* gene from *S. pyogenes* SF370 was codon-optimized and synthesized with a His tag-encoding sequence at the C terminal (Invitrogen). The adapted *cas9* fragment was ligated with the modified pLyc017 (empty vector) to generate an Fd::*cas9* cassette in the resultant pCas9. The gRNA scaffoldin was also synthesized (Invitrogen) (Figure S2.1B). All gRNA cassettes were constructed by splicing the RBS-free P4 promoter and the gRNA fragment using Splicing by Overlap Extension (SOEing). The P4::non-customized gRNA cassette was generated using primers, P4gRF and P4gRR for the promoter, gRCKF and gRNAR for the gRNA region, and then assembled with the modified pLyc017, generating pGRNA. To target *pyrF*, *mspI*, *β-gal*, 3198D, X21 and X22, one target site in each gene or site was selected (Table S2.1) and P4gRR and gRCKF were replaced by corresponding primers (Table S2.2). Customized gRNA cassettes were assembled with linearized pCas9, generating pCas9-*pyrF*, pCas9-*mspI* and pCas9-*β-gal*. The wild-type Cas9 endonuclease was mutated to Cas9 nickase

(D10A) via site-directed mutagenesis by using mutagenic primers, Cas9nF and Cas9nR. The *cas9* in the above plasmids was replaced by the *cas9n*, generating pCas9n, pCas9n-*pyrF*, pCas9n-*mspI* and pCas9n-*β-gal*. gRNA cassettes targeting 3198D, X21 and X22, were assembled with linearized pCas9n, generating pCas9n-3198D, pCas9n-X21 and pCas9n-X22, respectively.

To generate all-in-one vectors, user-defined donor templates were constructed by SOEing and then inserted into co-expression vectors. To construct a 2-kb donor template for a 23-bp deletion in the *pyrF* gene, 1-kb left (LH) and right (RH) homologous arms were firstly amplified separately using primer pairs, 0614LF and 0614LR and 0614RF and 0614RR, respectively, and then both fragments were spliced to produce the 2-kb donor for assembly with linearized pCas9n-*pyrF*, generating pCas9n-*pyrF*-donor. Similarly, pCas9n-*mspI*-donor, pCas9n-X21-donor, pCas9n-X22-donor, and pCas9n-*β*-gal-donor vectors with 1-kb, 0.5-kb, 0.2-kb and 0.1-kb arm sizes were constructed with designed primers (Table S2.2). A series of pCas9-*pyrF*-donors with the 0.71-kb Fd::*afp* expression cassette, 3-kb λ DNA and 6-kb λ DNA between 1-kb homologous arms were constructed by three-piece SOEing or sequential cloning using pBR322 (NEB) as intermediate plasmid. The pCas9n-3198D-donor with 1.72-kb promoterless α-acetolactate synthase (*alsS*) between 1-kb arms was constructed by sequential cloning. The promoterless *alsS* fragment was amplified from pLyc025. All constructs were verified by DNA sequencing for further studies.

2.3.3 Bacterial strains and culture conditions

E. coli Top10 (Invitrogen) was used for all cloning. *E. coli* transformants were grown at 37℃ in Luria-Bertani medium with chloramphenicol (15 μg/ml) for the pLyc017-

derived series, or ampicillin (50 μg/ml) for the pBR322-derived series. *C. cellulolyticum* H10 (ATCC 35319) and its developed strains were cultured anaerobically at 34 ℃ in VM medium with yeast extract (2.0 g/l) and cellobiose (5 g/l). If not otherwise specified, methylated plasmids were used for *C. cellulolyticum* electroporation (Li *et al.*, 2014) and then transformants were normally selected by thiamphenicol (TMP) (15 μg/ml). For Δ*pyrF* mutant identification, selective medium was additionally supplemented with 5-fluoroorotic acid (5-FOA) (500 μg/ml). Single colonies were anaerobically developed on VM plates at 34 ℃. Serial transfer was conducted by transferring a cell culture (OD600>0.4) to a new medium (1:10 v/v) and TMP was added if required. Cell growth was determined with three replicates by monitoring OD_{600}.

Δ*pyrF* mutants created by the pCas9n-*pyrF*-donor were initially screened with 5-FOA and then were identified individually by PCR amplicon sequencing. The Δ*pyrF* mutant created by Group II retrotransposition (Li *et al.*, 2014) was used as a positive control for phenotype identification. Similarly, ΔX21 and ΔX22 mutants, created by pCas9n-X21-donor and pCas9n-X22-donor, respectively, were identified by PCR amplicon sequencing. The TMP-resistant population generated from pCas9n-*mspI*-donor, containing Δ*mspI* mutants, was serially transferred and then the population genomic DNA was extracted for PCR identification and sequencing. The Δ*β*-gal mutant population generated by pCas9n-*β*-gal-donor was additionally identified with amplicon digestion by *Eco*RV.

To generate plasmid-cured strains, pure Δ*mspI* and Δ*pyrF/afp*$^{+}$ mutants were serially transferred in TMP-free medium. Then, cells were streaked on TMP-free plates

for colony development. Plasmid-cured colonies were screened by both PCR amplification of the plasmid-born region and TMP selection, and then verified by transforming unmethylated pGRNA.

2.3.4 Determination of editing efficiency and cargo capacity

Transformants of each construct (pCas9n-*β*-gal-donor with varying arm size) were generated by electro-transforming 0.25 pmol methylated plasmids with two replicates. Each recovered culture (T0) was equally inoculated into the selection medium (T1). Then, two more serial transfers (T2 and T3) were conducted sequentially when the OD600 was 0.4-0.5. At each transfer point, cell culture was collected for genomic DNA extraction. The extracted genomic DNA was used as PCR template to specifically amplify a 2-kb genomic region covering the entire donor, using primers, p3 and p4. A portion (1 μg) of each purified amplicon was digested with 10 U *Eco*RV in NEBuffer 4.1 at 37 ℃ for 3 h, for the purpose of distinguishing the edited and unedited amplicon by gel electrophoresis. Gel images were subject to densitometry analysis using Thermo MYImage. The editing efficiency (%) was calculated by dividing the intensity of the 2-kb bands from the selected culture by the initial intensity of the bands from the corresponding T0 control, and then multiplying by 100.

To examine the genetic cargo capacity, a series of vectors (pCas9-*pyrF*-donor with 0.71-kb Fd::*afp* expression cassette, 3-kb λ DNA and 6-kb λ DNA, and pCas9-3198D-donor with 1.72-kb *alsS*) were transformed. During three serial transfers only under TMP selection, resistant populations were subjected to genomic DNA extraction and then the edited genomes in the population were distinguished from WT by PCR amplification and gel electrophoresis.

2.3.5 RNA isolation, RT-PCR and quantitative real-time PCR

Total RNA was extracted from cellobiose (5 g/l)-grown *C. cellulolyticum* (OD_{600}=~0.45) by TRIzol (Invitrogen) and then reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen). The cDNA product was diluted as appropriate and used as a template. gRNA expression was examined by RT-PCR using *recA* as an internal calibrator (98 ℃ for 30 s, 22 cycles of 98 ℃ for 10 s, 56 ℃ for 10 s and 72 ℃ for 10 s). Quantitative PCR was performed using iTaq SYBR Green Supermix with ROX (Bio-Rad) on a Bio-Rad iQ5 thermal cycler. Gene-specific primers for each transcript are listed in Table S2. Thermal cycling conditions were as follows: 95 ℃ for 3 min, 40 cycles each of 95 ℃ for 15 s, 56 ℃ for 15 s and 72 ℃ for 45 s. The relative expression level of target genes compared to *recA* was calculated with the Pfaffl Method (Pfaffl, 2001).

2.3.6 SDS-PAGE analysis

To examine the expression of full-length His-tagged Cas9 in *C. cellulolyticum*, single colonies of pCas9 or CK (empty vector) transformants were cultured. Cells were lysed in the SDS loading buffer and then supernatant cell lysates were subjected to SDS-PAGE using 9% resolving gels (Bio-Rad). Additionally, His-tagged Cas9 protein in the gel was detected by Pierce 6xHis Protein Tag Stain Reagent Set (Thermo Scientific).

2.3.7 Fluorescence microscopy

Fresh cultures of wild-type *C. cellulolyticum*, P4::*afp* and Fd::*afp* strains and plasmid-cured Δ*pyrF/afp*$^{+}$ mutants at mid-log phase were analyzed using an Olympus BX51 fluorescence microscope equipped with optical filter sets with excitation at 490 nm and

emission at 525 nm for green fluorescence. The images were collected by an Olympus DP71 digital camera.

2.3.8 Bioinformatic analysis of target sites

All N20NGG sites in *C. cellulolyticum* genome (NC_011898.1) and their locations were extracted from both strands. Then, unique and transcribable target sites were selected by filtering out those with >2 identical sites across the genome, a string of six or more Ts in the 23-mer sequence (Dicarlo *et al.*, 2013) and T3 in the 6-mer region upstream of NGG (Wu *et al.*, 2014), or an extreme GC content (<25% and >80%) (Wang *et al.*, 2014). Usable target sites that had at least two base-pair mismatches with the rest of that region of the genome were used for targeting space analyses including calculating the distances between all adjacent usable sites and histogram plotting. The number of usable sites in all predicted genes was determined for histogram plotting. Gene coverage percentage was calculated by dividing the number of genes that had at least one usable target site by the total gene number. The genome-wide distribution was drawn by GenomeDiagram (Pritchard *et al.*, 2006). Following similar procedures, we analyzed the genomes of *Clostridium acetobutylicum* ATCC 824 (NC_003030.1), *E. coli* K-12 (NC_000914.3), *Bacillus subtilis* 168 (NC_000964.3) and *L. reuteri* DSM 20016 (NC_009514.1).

2.4 Results

2.4.1 Expression of CRISPR-Cas9 system in C. cellulolyticum

To establish Cas9-based genome editing in *C. cellulolyticum*, functional promoters are needed to drive the expression of Cas9 and gRNA. To quickly expand the promoter library, synthetic promoter design was applied. Since σ^A is the primary sigma factor

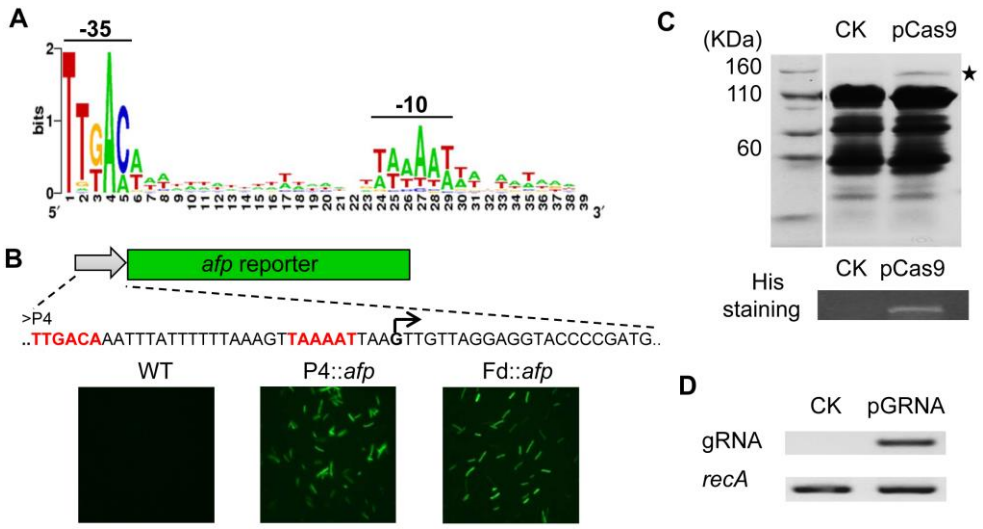


Figure 2.1 Generation and validation of Cas9 expression system. (A) Alignment of predicted σ^A-dependent promoters from C. cellulolyticum. Two highly conserved regions (-35 and -10) are separated by a 17-nt T/A rich spacer. (B) Promoter activity test, in which synthetic promoter P4 drives an anaerobic florescent protein-encoding gene (*afp*). The right angle arrow indicates the potential transcriptional start site. The -35 and -10 regions are in red. Fluorescent microscopy of *C. cellulolyticum* wild-type (WT) and transformants carrying P4::*afp* or Fd::*afp* constructs are shown below. (C) SDS-PAGE analysis of whole cell proteins from transformants with empty vector (CK) and pCas9. The asterisk denotes the estimated Cas9 band. The full-length His-tagged Cas9 is further verified by His protein staining. (D) RT-PCR analysis of gRNA in both CK and pGRNA strains, using the *recA* as an internal calibrator.

responsible for transcribing most genes in microbial cells (Osterberg *et al.*, 2011), *in silico* analysis of genome-wide σ^A-dependent promoters was conducted for *C. cellulolyticum*. An alignment of predicted promoters showed two characteristically conserved regions (-35 and -10) that were separated by a 17-nt T/A rich spacer (Figure 2.1A). A synthetic promoter (P4) comprised of nucleotides with the highest usage frequency at each position was chemically synthesized (length, 36 bp). The activity of P4 was tested in *C. cellulolyticum* by driving the expression of a reporter gene (*afp*) encoding the anaerobic fluorescent protein (Figure 2.1B). Under fluorescence

microscopy, the P4::*afp* construct presented a fluorescent signal in *C. cellulolyticum* (Figure 2.1B); the fluorescence intensity was comparable to the positive control in which a ferredoxin promoter (Fd) from *Clostridium pasteurianum* was used to control *afp* expression, generating the Fd::*afp* construct (Li *et al.*, 2014).

Next, we chose the P4 and Fd promoters to drive *gRNA* and *cas9* gene expression, respectively. The *cas9* gene of *S. pyogenes* was codon-adapted to *C. cellulolyticum* and fused with a His-tag at the C terminal. To examine Cas9 expression, we constructed a pCas9 shuttle vector carrying an Fd::*cas9* expression cassette. The full-length His-tagged Cas9 protein was successfully expressed as evidenced by SDS-PAGE analysis and His protein staining (Figure 2.1C). Additionally, we constructed a pGRNA vector harboring a P4::*gRNA* expression cassette. This construct was able to generate non-customized gRNA transcripts as shown by RT-PCR (Figure 2.1D). Then both expression cassettes (Fd::*cas9* and P4::*gRNA*) were combined into a single vector, pCas9-gRNA (Figure S2.1C). Once the gRNA is customized, the resultant vector is able to co-express Cas9 and gRNA to edit targeted genomic loci in a single step.

2.4.2 Lethality of Cas9-induced double-strand breaks

To demonstrate genome editing by gRNA-guided Cas9, a *pyrF* gene encoding orotidine-5'-phosphate decarboxylase (Ccel_0614) in *C. cellulolyticum* was chosen as our first target gene since inactivation of this gene would generate uracil auxotrophic and 5-fluoroorotic acid (5-FOA)-resistant phenotypes, which are easily monitored (Tripathi *et al.*, 2010). The pCas9-*pyrF* vector co-expressing Cas9 and the customized gRNA targeting the *pyrF* gene was electroporated into *C. cellulolyticum* in parallel with pCas9 and pGRNA-*pyrF*, both of which served as negative controls only expressing

either Cas9 or customized gRNA. Transformation tests revealed that both controls generated antibiotic-resistant transformants but not 5-FOA-resistant transformants (Table 2.1 and Figure S2.2A and S2.2B); the co-expression vector did not produce cells with both antibiotic and 5-FOA resistance. These results suggested that co-expressing Cas9 and gRNA was toxic at least at the selected target site. Then, we tested two more target sites, one in *β-galactosidase* (β-*gal*, Ccel_0374) and the other in *mspI endonuclease* (*mspI*, Ccel_2866), and determined that co-expression vectors were unable to produce antibiotic-resistant cells (Figure S2.2C). We suspected that the problem might be in the unsuccessful repair of DSBs created by the Cas9-gRNA complex since DSBs can interrupt chromosome replication and cell reproduction. To verify this hypothesis, the wild-type Cas9 was replaced with Cas9n (D10A) (Jinek *et al.*, 2012), generating pCas9n-*pyrF*. Interestingly, after transformation we observed the propagation of antibiotic-resistant cells but these cells were not 5-FOA-resistant (Table 2.1 and Figure S2.2A and S2.2B), suggesting that the Cas9-induced lethality can be voided by Cas9n and that the single nick created by Cas9n did not enable genome editing via NHEJ. Afterwards, we investigated the expression of major NHEJ components (Bowater *et al.*, 2006; Pitcher *et al.*, 2007), including Ku (Ccel_0364), ATP dependent DNA ligase (Ccel_0365) and DNA polymerase LigD (Ccel_0366). Strikingly, all three genes were expressed at a very low level in comparison to the *recA* housekeeping gene (Xu *et al.*, 2014) (Figure S2.3). Taken together, these results indicate that *C. cellulolyticum* NHEJ is inefficient in repairing DSBs, which restricts the use of Cas9 in editing the *C. cellulolyticum* genome.

Table 2.1 Use of Cas9 nickase instead of wild-type Cas9 for genome editing in *C. cellulolyticum*.

Plasmid	Component	Cell growth by resistance type[a]:	
		TMP^r	$5\text{-}FOA^r$
pCas9	Cas9	Y	N
pGRNA-pyrF	gRNA	Y	N
pCas9-pyrF	Cas9 + gRNA	N	N
pCas9n-pyrF	Cas9n + gRNA	Y	N
pCas9-pyrF-donor	Cas9 + gRNA + donor template	N	N
pCas9n-pyrF-donor	Cas9n + gRNA + donor template	Y	Y

[a]TMP^r, thiamphenicol resistant; $5\text{-}FOA^r$, 5-fluoroorotic acid-resistant; Y, cell growth; N, no cell growth. Growth profiles are shown in Figure S2.2.

2.4.3 Precise genome editing via a single nick-triggered homologous recombination

Homology-directed repair is another way to fix DNA lesions when a homologous template is present (Dillingham *et al.*, 2008). To mutate the *pyrF* gene by small DNA deletion, we designed a homologous donor template with a length of 2 kb carrying a 23-bp deletion in the middle and cloned it into pCas9-*pyrF* and pCas9n-*pyrF*, generating all-in-one pCas9-*pyrF*-donor and pCas9n-*pyrF*-donor plasmids (Figure 2.2A and Figure S2.1C). In this way, editing templates can be maintained during plasmid replication. Transformation tests showed that even though the editing templates were present, Cas9-induced DSBs did not produce any resistant cells; however, Cas9n-induced single nicks, coupled with HR, produced resistant cells under antibiotic and 5-FOA selection (Table 2.1 and Figure S2.2A and S2.2B), suggesting Δ*pyrF* mutants may be generated. After spread plating, we randomly picked 12 colonies for sequencing and found that all were Δ*pyrF* mutants containing a precise deletion of the 23-bp target sequence in the gene (Figure 2.2B). Using the same strategy, we targeted the *mspI* gene (Figure S2.4A), which encodes an endonuclease of the restriction-modification system in *C.*

cellulolyticum (Cui *et al.*, 2012). After constructing and transforming the pCas9n-*mspI*-*donor* carrying a 2-kb donor template with a 23-bp deletion inside, we examined the Δ*mspI* mutants in the antibiotic-resistant population. PCR amplification revealed that the wild-type was specifically detected in the control using an empty vector but was not detected in the resistant population (Figure S2.4B), indicating the deletion of the 23-bp

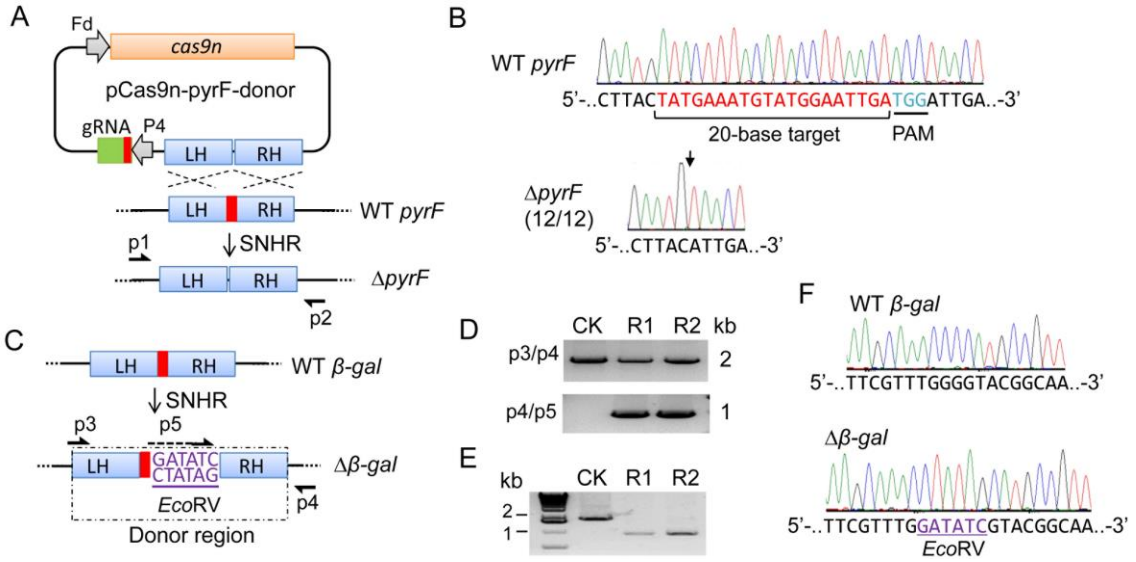


Figure 2.2 Precise deletion and insertion of a small fragment. (A) Schematic all-in-one vector for *pyrF* disruption by a single nick-triggered homologous recombination (SNHR). The vector consists of an Fd-driven *cas9n* gene, P4-driven gRNA targeting *pyrF* gene and donor template with a 23-bp deletion flanked by 1-kb left homologous (LH) and right homologous (RH) arms. (B) DNA sequence chromatograms showing the deletion of a 23-bp target site in the *pyrF* gene. The 23-bp region carries 20-base gRNA sequence and 3-base protospacer adjacent motif (PAM). Twelve colonies all present precise deletion at the position indicated by a downward black arrow. Amplicon for sequencing was generated using primers, p1 and p2, as schematized in A. (C) SNHR-mediated insertion of an *Eco*RV site at a target cut site in the *β-gal* gene. The donor template shown in the dashed box carries the *Eco*RV site flanked by 1-kb LH and RH starting from the Cas9n cleavage site. (D) PCR identification of Δ*gal* mutants. Transformant population of empty vector (CK) and pCas9n-*β-gal*-donor (R1 and R2, two replicates) is identified by two primer pairs as drawn in C. (E) *Eco*RV digestion of p3/p4 PCR products. (F) DNA sequence chromatograms verifying the precise insertion of *Eco*RV (underlined) in the Δ*β-gal* mutant.

target fragment in that population. Then, DNA sequencing further confirmed a precise deletion in the Δ*mspI* mutant (Figure S2.4C). After plasmid curing, the Δ*mspI* mutant was further shown to be transformable with unmethylated plasmids (Figure S2.4D). Taken together, these results demonstrate that this single nick-triggered HR (SNHR) allows a one-step precise DNA deletion in *C. cellulolyticum*.

In genetic engineering, small DNA insertions are useful for integrating short functional elements and introducing frameshift mutations. To test the potential of small DNA insertions, we tried to introduce an *Eco*RV site (5'-GATATC-3') into the target site of the β-*gal* gene (Figure 2.2C). A donor template harboring an *Eco*RV site in the middle flanked by 1-kb homologous arms starting from the cleavage site was constructed and used to generate a pCas9n-β-*gal*-donor for transformation. *Eco*RV insertion was initially indicated by differential PCR amplification (Figure 2.2D), which generated the intended amplicon only when edited genomes were present. Then, amplicon digestion by *Eco*RV and amplicon sequencing both confirmed the insertion of *Eco*RV at the anticipated locus (Figure 2.2E and 2.2F). Thus, small insertion is also operable using this strategy.

2.4.4 Assessment of editing efficiency and genetic cargo capacity

A powerful genome editing tool should have a high efficiency allowing for marker-independent editing. Here we evaluated the editing efficacy of this SNHR strategy and the effect of arm size on editing since the length of homologous arms affects recombination frequency (Khasanov *et al.*, 1992; Bertolla *et al.*, 1997; Kung *et al.*, 2013). We constructed a series of donor templates, all of which harbor an *Eco*RV site in the middle flanked by homologous arms of varied length (0.1 kb, 0.2 kb, 0.5 kb and 1

kb), and then constructed pCas9n-*β-gal*-donor vectors (Figure 2.3A). Since co-existence of the Cas9n-gRNA complex and the donor template may continuously trigger editing, extending the reaction time and possibly increasing the mutant population abundance, cell cultures from post-transformation recovery (T0) and three serial transfers (T1, T2 and T3) under antibiotic selection were collected for genomic DNA composition

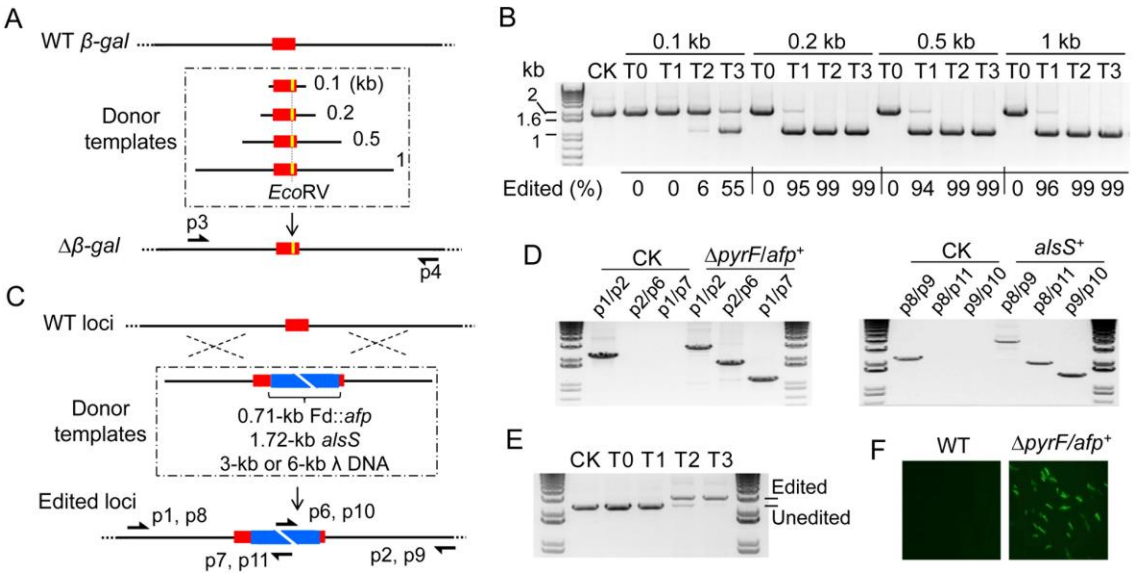


Figure 2.3 Evaluation of editing efficacy and cargo capacity. (A, B) Effect of arm size on editing efficacy. (A) Design of donor templates with varying arm size (0.1-1 kb), in which the target site (red) is modified to contain an *Eco*RV site (yellow). The all-in-one vectors with these templates introduce *Eco*RV into the *β-gal* gene via SNHR. (B) Editing efficacy evaluation by *Eco*RV digestion of p3/p4 PCR product. The percentage of edited genome in the whole population of control with donor-free vector (CK), recovered cells (T0) and TMP-resistant cells from three serial transfers (T1-3) is calculated by densitometry analysis. (C-F) Genetic cargo capacity evaluation by delivering foreign DNA fragments with varying size into the genome. (C) Design of four donor templates with 0.71-kb Fd::*afp*, 1.72-kb promoterless *alsS*, 3-kb and 6-kb λ DNA (blue) in between 1-kb arms. Using SNHR, the *alsS* fragment and the remaining are inserted into two different sites, 3198D and *pyrF*, respectively. (D) PCR identification of Δ*pyrF/afp*$^+$ and *alsS*$^+$ mutants generated by the insertion of Fd::*afp* and *alsS* fragments, using wild type (CK) as control. Primer pairs are indicated and drawn in C. (E) Enrichment of Δ*pyrF/afp*$^+$ mutant in the population during serial transfer (T0-3) using wild type (CK) as control. (F) Fluorescence microscopy of plasmid-cured Δ*pyrF/afp*$^+$ mutant.

analysis. Amplicon digestion by *Eco*RV reflected the relative abundance of the edited genomes across the whole population (Figure 2.3B), demonstrating that i) the control group using donor-free pCas9n-*β-gal* never produced any detectable genome editing (unedited = 2 kb, edited = 1 kb); ii) the 0.1 kb arm group did not produce edited genomes in T0 or T1, but 6% of the population of T2 carried edited genomes and 55% of T3 carried edited genomes; and iii) in the 0.2 kb, 0.5 kb and 1 kb groups, editing was not detected in T0 samples but strikingly jumped to over 95% in all T1 samples and then to nearly 100% in T2 and T3. Obviously, the length of the homologous arms exerts an important effect on editing efficiency, and the abundance of edited genomes can be significantly enriched with serial transfers. Once the arm length is greater than 0.2 kb, the editing efficiency of this SNHR strategy was very high (> 95%), indicating the ease of marker-independent genome editing.

We then examined the genetic cargo capacity of this strategy in delivering foreign DNA into the genome, which is of critical importance for future genome-level metabolic engineering. We constructed a series of all-in-one vectors in which donor templates contained 1-kb homologous arms and foreign fragments of varying size (0.71-kb Fd::*afp* expression cassette, 1.72-kb promoterless α-acetolactate synthase (*alsS*), 3-kb and 6-kb λ DNA) (Figure 2.3C). After conducting transformation and serial transfer, we successfully integrated the Fd::*afp* construct and *alsS* fragment into the targeted loci (Figure 2.3D), but not the larger λ DNA fragments. Meanwhile, we examined enrichment during serial transfer for the Fd::*afp* construct. The edited cells (Δ*pyrF*$^{-}$*/afp*$^{+}$ mutant) quickly accumulated to nearly 100% after three serial transfers (Figure 2.3E). The inserted *afp* gene in the plasmid-cured Δ*pyrF*$^{-}$*/afp*$^{+}$ mutant was well-expressed as

shown by fluorescence signal (Figure 2.3F). Therefore, the SNHR method can efficiently and markerlessly deliver foreign genes in a single step.

2.4.5 Precise editing at non-specific target sites

The specificity of the 23-bp target sites greatly affects the precision of the Cas9-based editing tools; without this specificity, unwanted off-target mutations will occur (Fu *et al.*, 2013; Ran *et al.*, 2013; Lin *et al.*, 2014). Since the four target sites tested above are highly specific, they are not ideal for examining editing specificity of this SNHR method. Instead, two highly similar target sites, X21 and X22, were selected from a *cipC* scaffoldin gene (Ccel_0728). These sites differ by only two bases in the 5' region preceding the same 12-bp seed region (Figure 2.4A and Figure S2.5). Loss of specificity in the seed region will dramatically decrease editing precision such that off-target mutations would occur (Fu *et al.*, 2013; Lin *et al.*, 2014). For each target site, a corresponding donor template was constructed to introduce a deletion of a 12-bp DNA fragment spanning the protospacer adjacent motif. After transformation and plating, we picked individual colonies for site-specific amplification and sequencing. Results showed that (Figure 2.4B and 2.4C): i) the editing system targeting X21 exhibited a 100% on-target editing ratio (12/12) for introducing a deletion there, and no off-target mutations (0/12) were detected at X22; and ii) the editing system targeting X22 also presented a 100% on-target editing ratio (10/10) and no off-target mutations (0/10) occurred at X21. Obviously, this method presented an extraordinary editing precision at non-specific target sites. This feature does not need the high-specificity target sites for precise genome editing required by other Cas9-based methods (Fu *et al.*, 2013; Ran *et al.*, 2013; Lin *et al.*, 2014).

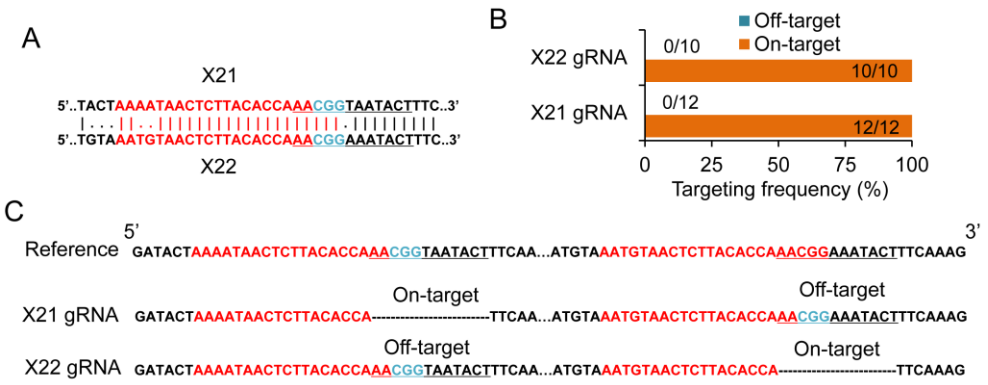


Figure 2.4 Targeting specificity test. (A) Pair-wise alignment of two target sites, X21 and X22 (colorized region). The 12-bp deletion regions are underlined. (B) On-target and off-target frequency in mutants generated by X21 and X22 gRNAs. There are twelve and ten individual colonies analyzed for X21 and X22, respectively. (C) Results of amplicon sequencing at both sites in each mutant.

To further assess the potential use of this method for genome editing, we analyzed the targeting space in the genome of *C. cellulolyticum*. After screening for usable target sites, those N20NGG sites (N is any base) that are unique, transcribable and have a certain degree of specificity, 75% of all extracted N20NGG sites met these criteria (Table S2.3). The sites were spread across the genome, but there are 91 regions (>1 kb) without any usable target sites with a maximal non-targetable region length of 21.9-kb (Figure 2.5A, in the outer two tracks of the map). Further statistical analysis indicated that the median interval distance between target sites was 9 bp (Figure 2.5B) and that almost all genes (95.7%) had at least one usable site and the median number of usable sites per gene was 35, without considering fragment length (Figure 2.5C). This high targeting coverage was also observed in other bacteria, including *E. coli* K-12, *B. subtilis* 168, *C. acetobutylicum* ATCC 824 and *L. reuteri* DSM 20016 (Table S2.3). Thus, this repurposed CRISPR-Cas9 tool is applicable for editing nearly all encoding genes despite some inaccessible non-coding genomic regions.

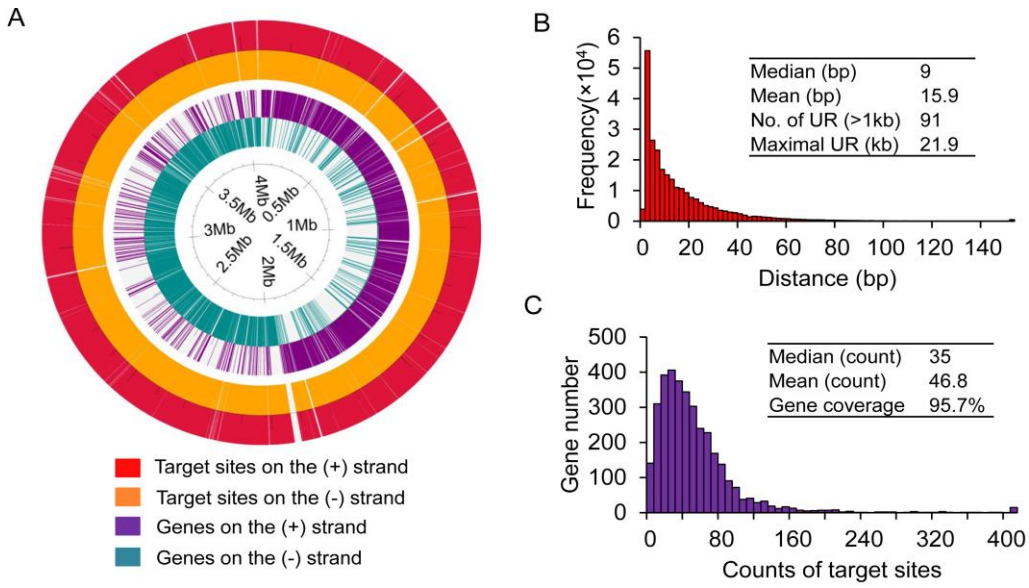


Figure 2.5 Bioinformatic analysis of targeting space in *C. cellulolyticum*. (A) Genome-wide distribution of genes and target sites on both DNA strands. White areas in each track indicate gaps between adjacent genes or target sites. Color code is given below the map. (B) Histogram of distance between adjacent usable target sites. Values of mean and median, the number of untouchable regions (UR) with the length of >1 kb and the length of the maximal UR are inset within the plot. (C) Histogram of the number of usable target sites in genes. The values of mean, median and gene coverage are inset.

2.5 Discussion

We have developed a highly efficient strategy for genome editing in *C. cellulolyticum* using Cas9n-mediated single nick generation and HR. This SNHR strategy is capable of circumventing the DSB lethality to allow versatile editing in hosts with inefficient DSB repair systems. Although NHEJ and HR assist Cas9-mediated genome editing in diverse eukaryotes (Xu *et al.*, 2014), our study demonstrated the NHEJ components of *C. cellulolyticum* were minimally expressed, which resulted in ineffective rejoining of DSB created by Cas9. Since key components of the NHEJ system, specifically the signature protein Ku, are present in only 27.5% of sequenced microbes (Figure S2.6) (Bowater *et al.*, 2006) and even those genomes harboring these genes may not encode

functional proteins, as is the case for *C. cellulolyticum*, the Cas9-/double nicking-triggered NHEJ system will not work in a majority of prokaryotes. The alternative to NHEJ is template-directed HR, which is a ubiquitous housekeeping process involved in the maintenance of chromosome integrity and the generation of genetic variability, although the exact mechanism of HR may vary (Rocha *et al.*, 2005; Dillingham *et al.*, 2008). Our plasmid-borne homologous donor successfully triggered HR at the nick created by Cas9n but not at the break induced by Cas9. Recent studies showed that single nick-triggered HR may undergo a distinct mechanism without proceeding through a DSB intermediate of DSB-induced HR (Davis *et al.*, 2011; Metzger *et al.*, 2011). It is also possible that DSBs created by Cas9 are more toxic than the single-strand nicks or nick-induced one-end DSBs occurring during DNA replication and may be beyond the host's ability to repair (Dillingham *et al.*, 2008). Although little is known about the molecular basis of the *C. cellulolyticum* type II-C CRISPR-Cas system, our study suggests that the native system did not affect the *S. pyogenes* type II-A system and might use separate mechanisms (e.g., different PAM, gRNA structure as well as protospacer length) since the gRNA-expressing strain was not able to direct the native Cas9 to accomplish targeted editing. The Cas9 orthogonality demonstrated in *E. coli* and human cells also supports this point (Esvelt *et al.*, 2013).

The SNHR strategy presents unmatched advantages over mainstream bacterial genome editing tools. Compared with the widely-used double cross-over recombination method, it is much faster, more efficient and more versatile. As we demonstrated, the SNHR strategy allows a one-step generation of an edited genome using a single vector. The high efficiency of this strategy enables markerless editing so that difficulties

associated with low transformation efficiency, tedious step-wise screening and the need for multiple positive-/negative-selection markers can be avoided, unlike in the double cross-over recombination method (Heap *et al.*, 2012; Esvelt *et al.*, 2013; Xu *et al.*, 2014). Studies have shown that a low spontaneous recombination frequency in bacteria, which is the basis of double cross-over recombination, can decrease exponentially when reducing the size of homologous arms or increasing the non-homologous insert between the flanking homologous arms because these changes can affect the efficiency of recombination pathways and RecA binding (Shen *et al.*, 1986; Khasanov *et al.*, 1992; Bertolla *et al.*, 1997; Kung *et al.*, 2013). While both SNHR and double cross-over recombination can generate defined mutations (deletion, insertion and replacement) via HR, the SNHR strategy exhibited a strong ability to use homologous arms as short as 0.2 kb to trigger recombination and deliver DNA fragments within a single step, so the SNHR method is a more robust method for small gene insertion within a short time-frame. However, the genetic cargo capacity is relatively low and needs to be improved in order to integrate the large DNA fragments required for massive metabolic engineering. Group II intron retrotransposition is also widely used for gene disruption in many bacteria (Esvelt *et al.*, 2013), yet this method has some targeting limitations including an obvious bias for intron insertion near the replication origin (Zhong *et al.*, 2003), a relatively sparse targeting space (every few hundred bases on average) and no guarantee of efficiency depending on the insertion site and species (Perutka *et al.*, 2004). In contrast, the SNHR strategy has a very wide targeting space with a median interval distance of 6 to 14-bp in the multiple bacterial genomes analyzed in this study. It also allows editing of over 95% of genes in multiple genomes, demonstrating the

great versatility of this editing system. In addition, the customizability of the SNHR strategy, which enables the generation of precise micro-deletion, micro-insertion or codon change to inactivate gene function, can minimize the polar effect on downstream genes that can be exerted by intron insertion or insertion/deletion of other large DNA fragment (Maamar *et al.*, 2004; Xu *et al.*, 2014). With these demonstrated strengths, the SNHR strategy can overcome the limitations of currently available genetic approaches to engineering bacterial genomes. This new Cas9 technology can be used for *in vivo* and *in situ* characterizing and altering biological functions of interest (e.g., DNA sequence motif, gene, protein domain and protein localization), in addition to genetic engineering of Clostridia and other industrial microorganisms for metabolic and physiologic improvement.

In addition, compared with reported Cas9-based strategies (i.e., Cas9-NHEJ/HR, double nicking-NHEJ/HR) (Cong *et al.*, 2013; Li *et al.*, 2013; Ran *et al.*, 2013), this strategy can enable precise editing at target sites with low specificity. For instance, Cas9n guided by X21 gRNA probably induces at least two nicks in the *C. cellulolyticum* genome, including at the on-target X21 and the off-target X22, but the donor template of X21 will specifically choose the nick in X21 to repair through HR and then other nicks will be faithfully religated without introducing any unwanted mutations, as usually occurs during NHEJ-dependent DSB repair. That means the SNHR strategy not only improves editing accuracy, but also expands our editing target space. However, strategies still need to be developed to target those genomic regions lacking targeting sites and to increase targeting resolution across genomes, which is problematic for all Cas9-based methods, including SNHR. Considering that different

Cas9s have varied PAM preferences (e.g., NGG in *S. pyogenes*, NNNNGANN in *Neisseria meningitidis*, NAAAAN in *Treponema denticola*) (Esvelt *et al.*, 2013), exploiting or engineering Cas9 to have an expanded ability to use multiple short protospacer adjacent motifs, and to decrease the length requirement of protospacers without sacrificing targeting specificity, may offer solutions for allowing accurate editing anywhere.

In conclusion, the single nick-triggered HR strategy described here allows for marker-independent gene delivery and versatile editing in a single step with a high editing efficiency and precision. This method provides an exemplary strategy for precise genome editing in prokaryotes that are sensitive to DSB toxicity. This approach will facilitate microbial genome editing for fundamental and applied research.

Chapter 3 Cas9 nickase-assisted RNA Repression Enables Stable and Efficient Manipulation of Essential Metabolic Genes in *Clostridium cellulolyticum*

3.1 Abstract

The lack of simple methodologies for stably manipulating essential genes hinders their functional characterization and engineering-oriented studies in bacteria. *Clostridium cellulolyticum* is a promising candidate for consolidated bioprocessing to convert lignocellulose into value-added chemicals. Eliminating the formation of less-valuable lactate and acetate will significantly improve its value to industry. However, reducing acetate formation or co-manipulating it with other pathways is challenging due to the essentiality of acetate-producing genes. Here we developed a Cas9 nickase-assisted chromosome-based RNA repression to stably manipulate essential genes, allowing for advanced metabolic engineering in *Clostridium cellulolyticum*. Plasmid-based expression of antisense RNA (asRNA) molecules targeting the phosphotransacetylase (*pta*) gene successfully reduced the enzymatic activity by 35% in cellobiose-grown *C. cellulolyticum*, metabolically decreased the acetate titer by 15% and 52% in wildtype transformants on cellulose and xylan, respectively. Transformants of the double mutant of lactate dehydrogenase and malate dehydrogenase reduced acetate titer by more than 33%, concomitant with negligible lactate formation. The strains with *pta* gene repression diverted more carbon into ethanol. However, further testing on chromosomal integrants that were created by double-crossover recombination exhibited only very weak repression because DNA integration dramatically lessened gene dosage. With the design of a tandem repetitive promoter-driven asRNA module and the use of a new

Cas9 nickase genome editing tool, a powerful chromosomal integrant (LM3P) was generated in a single step and successfully enhanced RNA repression, with a 27% decrease in acetate titer on cellulose in antibiotic-free medium. The LM3P integrant exhibited additional changes in cell growth, cellulose utilization, and other fermentation products especially at higher cellulose loading. Gene repression dramatically reduced acetate formation and enhanced carbon flux to produce ethanol. The tandem promoter-driven RNA repression module in chromosome overcame the weakened repression caused by chromosomal integration. This combinatorial method using a Cas9 nickase genome editing tool to integrate the gene repression module demonstrates easy-to-use and high-efficiency advantages, paving the way for stably manipulating genes, even essential ones, for functional characterization and microbial engineering.

Keywords: metabolic engineering; consolidated bioprocessing; *Clostridium cellulolyticum*; genome editing; gene repression; essential genes

3.2 Introduction

Essential genes are indispensable for building up the chassis of living organisms (Glass *et al.*, 2006), and accounts for 5-80% of bacterial genomes (Gao *et al.*, 2011). Investigation into these genes will provide insights on basic biological functions and allow for the discovery of cellular activities that could be used in industrial or biomedical applications, although many of these would require subsequent engineering for improved utilization (Lee *et al.*, 2009; Juhas *et al.*, 2012). If genes are essential it becomes more technically challenging because genetic knock-outs of essential genes are lethal, making mutants unobtainable (Glass *et al.*, 2006); and attempting to modify the gene expression, instead of knocking it out completely, can result in unpredictable changes in the magnitude of gene expression (Ji *et al.*, 1999).

There are three major approaches available for targeted gene repression in bacteria, including antisense RNA (asRNA)-mediated repression (Desai *et al.*, 1999; Perret *et al.*, 2004; Thomason *et al.*, 2010), Hfp-dependent RNA repression (Man *et al.*, 2011; Na *et al.*, 2013) and nuclease-null Cas9-mediated repression (which is named CRISPRi) (Bikard *et al.*, 2013; Qi *et al.*, 2013). The latter two require an RNA binding protein, Hfp chaperone and non-catalytic Cas9 endonuclease, respectively, which need to be consistently co-expressed with RNA molecules that recognize target transcripts. Plasmid-based expression of these components has been widely applied in diverse bacteria (Desai *et al.*, 1999; Perret *et al.*, 2004; Thomason *et al.*, 2010; Man *et al.*, 2011; Bikard *et al.*, 2013; Na *et al.*, 2013); however, concerns are raised about the stability and antibiotic dependence of plasmid-based expression (Lee *et al.*, 1993; Friehs, 2004), especially in industrial microorganisms, and potential side effects caused by the

specificity of RNA binding proteins (Martinez-Alonso *et al.*, 2010; Bikard *et al.*, 2013; Qi *et al.*, 2013). Development of a relatively clean, easy and efficient approach allowing for rapidly generating stable knock-down mutants would increase our ability to study and manipulate essential genes. Considering the easy-to-use and highly efficient advantages of CRISPR/Cas9-based genome editing tools (Xu *et al.*, 2014; Xu *et al.*, 2015) and the simplicity and universality of antisense RNA-mediated repression (Thomason *et al.*, 2010), here we propose a combination of these two methods using Cas9 technology to integrate antisense RNA modules into the genome. By doing so, knock-down mutants can be created in a single step with features that are plasmid-independent and can be sustained without using antibiotics.

Clostridium cellulolyticum H10, a model organism of mesophilic cellulolytic Clostridia, is an excellent consolidated bioprocessing host (Desvaux, 2005; Lynd *et al.*, 2005). It can hydrolyze lignocellulose without adding commercial cellulases and simultaneously ferment a variety of C5 and C6 sugars to end products (lactate, acetate and ethanol) (Desvaux, 2005). Metabolic engineering significantly improved microbial characteristics via overexpressing foreign genes of intended pathways (Guedon *et al.*, 2002; Higashide *et al.*, 2011; Li *et al.*, 2014; Lin *et al.*, 2015), or eliminating competing and promiscuous pathways (Shaw *et al.*, 2008; Li *et al.*, 2012; Papanek *et al.*, 2015). In *C. cellulolyticum*, a double mutation of lactate and malate dehydrogenase genes *(*Δ*ldh* Δ*mdh,* hereafter LM mutant) abolished lactate production, accompanied with carbon flux redistribution (Li *et al.*, 2012). However, no knock-out mutants of acetate producing genes, phosphotransacetylase (*pta*) and acetate kinase (*ack*), were isolated to abolish acetate formation, suggesting that these two genes are essential in *C.*

cellulolyticum under the condition tested (Li *et al.*, 2012). The difficulty hampered combined metabolic engineering to maximize the elimination of less useful products (acetate and lactate) as demonstrated in the triple mutant of *T. saccharolyticum* (Δ*ldh* Δ*pta* Δ*ack*) (Shaw *et al.*, 2008) and the quintuple mutant of *C. thermocellum* (Δ*hpt*, Δ*ldh*, Δ*hydG*, Δ*pfl*, and Δ*pta-ack*) (Papanek *et al.*, 2015). With the aim of reducing acetate formation by manipulating these essential metabolic genes, both the traditional double-crossover recombination (Heap *et al.*, 2012) and the newly developed Cas9 nickase-triggered homologous recombination, which has been proven in *C. cellulolyticum* (Xu *et al.*, 2015), were employed to deliver the cassettes of antisense RNA expressing modules into a targeted genomic locus. The RNA repression effect in plasmid transformants and chromosomal integrants was determined and compared. Then, we improved the repression effect in chromosomal integrants by using a synthetic tandem promoter. The genetic regulatory strategies established in this study will greatly expand our ability to stably tune the expression of genes for genetic and metabolic engineering of bacteria.

3.3 Materials and methods

3.3.1 Plasmid construction

To construct plasmids expressing asRNAs, a partial transcriptional region of either the *pta* or *ack* gene, spanning from the predicted transcriptional start site to the downstream site approximately 120 bp away from the start codon, was amplified with specific primer sets (Table S3.1). Then, qualified PCR products were fused with the *Clostridium pasteurianum* ferredoxin promoter in an inverted orientation by assembling with BamHI-linearized pRNAi control plasmid (Gibson assembly kit, NEB), generating

pRNAi-pta and pRNAi-ack harboring Fd::pta asRNA module and Fd::ack asRNA module respectively.

To conduct chromosomal integration of asRNA modules via double-crossover recombination (Heap *et al.*, 2012), plasmids containing these asRNA modules flanked by homologous arms were constructed as follows. First, DNA fragments of interest were amplified and purified separately: promoterless *mlsR* gene amplified from pLyc1217Er (Li *et al.*, 2012); asRNA cassettes from pRNAi and pRNAi-pta; upper and lower homologous arms from the wildtype genome; and linear backbone from pRNAi. These fragments were then mixed and assembled together using a Gibson assembly kit and the resulting reaction product was transformed into *E. coli* for colony screening. Consequently, plasmids pLyc045 and pLyc046 were constructed with 3198up-mlsR-empty asRNA-3198down and 3198up-mlsR-pta asRNA-3198down for the integration of Fd::empty and Fd::pta asRNA cassette at the selected locus. Similarly, to integrate the Fd::afp cassette there, pLyc048 was constructed with 3198up-mlsR-Fd::afp-3198down.

To increase asRNA expression, a tandem promoter cluster consisting of three P4 promoters was synthesized and then fused with the same asRNA region by overlapping PCR, generating a 3P4::pta asRNA cassette. Since Cas9 nickase-based chromosomal integration is simpler and much more efficient (Xu *et al.*, 2015), it was applied to deliver 3P4::pta asRNA into the genome. The 23-bp target site (5'-AAGTAAGAAACATTTGGTTCCGG-3') was located in the downstream intergenic region of Ccel_3198. pCas9n-3198D with a customized donor was constructed in two steps. First, pCas9n-3198D reported previously was linearized by BamHI (Xu *et al.*,

2015) and then assembled with both left and right homologous arms amplified from the wildtype genome, generating pCas9n-3198D with NcoI-containing donor. Second, the resulting plasmid was linearized by NcoI for the assembly with the 3P4::pta asRNA cassette, generating pCas9n-3198D-donor. Descriptions of all plasmids used in this study were listed in Table 3.1.

Table 3.1 Plasmids and strains used to study acetate producing genes.

Name	Description	Reference
Strain		
E. coli TOP10	Host cells for plasmid construct	Invitrogen
WT	Wildtype *C. cellulolyticum* H10	ATCC
WT-P	WT with pRNAi control plasmids	This study
WT-P-pta	WT with pRNAi-pta plasmids	This study
WT-P-ack	WT with pRNAi-ack plasmids	This study
WT-P-afp	WT with pFd-AFP plasmids	(Xu *et al.*, 2015)
WT-G	WT with a chromosomal Fd::empty cassette	This study
WT-G-afp	WT with a chromosomal Fd::afp cassette	This study
LM	Δ*ldh* Δ*mdh*	(Li *et al.*, 2012)
LM-P	LM with pRNAi plasmids	This study
LM-P-pta	LM with pRNAi-pta plasmids	This study
LM-G	LM with a chromosomal RNAi control	This study
LM-G-pta	LM with a chromosomal Fd::pta asRNA cassette	This study
LM3P	LM with a chromosomal 3P4::pta asRNA cassette	This study
Plasmid		
pRNAi	CMP^r in E. coli; TMP^r in H10; Fd::empty cassette	(Xu *et al.*, 2015)
pRNAi-pta	pRNAi derivative with a Fd::pta asRNA cassette	This study
pRNAi-ack	pRNAi derivative with a Fd::ack asRNA cassette	This study
pLyc045	pRNAi derivative with 3198up-mlsR-Fd::empty-3198down	This study
pLyc046	pRNAi derivative with 3198up-mlsR-Fd::pta asRNA-3198down	This study
pFd-AFP	pRNAi derivative with a Fd::afp cassette	(Xu *et al.*, 2015)
pLyc048	pRNAi derivative with 3198up-mlsR-Fd::afp-3198down	This study
pCas9n-3198D	pRNAi derivative with a cas9 nickase and a gRNA targeting the 3198D site	(Xu *et al.*, 2015)
pCas9n-3198D with donor	pCas9n-3198D derivative with left arm-3P4::pta asRNA-right arm	This study

3.3.2 Bacterial strains and culture conditions

Escherichia coli Top10 (Invitrogen) was used for molecular cloning. Transformants were grown at 37 ℃ in Luria-Bertani medium supplemented with kanamycin (50 μg/ml) or chloramphenicol (15 μg/ml) when required. *Clostridium cellulolyticum* H10 and developed strains were cultured anaerobically at 34 ℃ in VM media supplemented with 2.0 g/L yeast extract and various carbon sources (Higashide *et al.*, 2011). Transformants of H10 and LM mutant were selected by erythromycin (15 μg/ml) or thiamphenicol (15 μg/ml). Colonies of *C. cellulolyticum* strains were developed on solid VM plates containing 1% (w/v) agar, 5 g/L cellobiose and antibiotics. Plasmid transformants were generated by transforming the corresponding plasmids. Chromosomal integrants, WT-G and WT-G-afp, were generated by transforming WT with pLyc045 and pLyc048, respectively. Chromosomal integrants, LM-G, LM-G-PTA and LM3P, were generated by transforming the LM mutant with pLyc045, pLyc046, and pCas9n-3198D-donor, respectively. All constructed strains are listed in Table 3.1.

3.3.3 C. cellulolyticum *transformation*

C. cellulolyticum electro-competent cells and methylated plasmids were prepared as previously described (Li *et al.*, 2014). Briefly, *C. cellulolyticum* strains were grown at 34 ℃ in liquid VM medium with 5 g/L cellobiose and 2 g/L yeast extract until reaching an OD600 =0.3-0.5. The cell culture was then chilled on ice and then centrifuged at 4 ℃ and 3,000 g for 8 min, and the cell pellets were washed at least three times with an equal volume of ice-cold anoxic electroporation buffer (270 mM sucrose, 1 mM $MgCl_2$ and 5 mM sodium phosphate buffer, pH 7.4). Lastly, competent cells made from every 10 ml of cell culture were resuspended in 100 μl chilled electroporation buffer for

further use. Plasmid DNA was methylated with MspI methyltransferase (NEB), followed by DNA purification and quantification. For each transformation, a 100 µl cell suspension was mixed with 2.0 µg of methylated plasmids and then electroporated in a 2-mm cuvette (1.25 kV, 5 msec, 1 square pulse) with a Gene Pulser Xcell (Bio-Rad) in the anaerobic chamber. After electroporation, cells were recovered for 12-24 h in antibiotic-free VM medium with 5 g/L cellobiose and 2 g/L yeast extract, and then selected by appropriate antibiotics on agar VM plates.

3.3.4 Enzyme activity assay

To measure enzyme activities, cell-free extracts were made from cellobiose-grown *C. cellulolyticum* strains at the mid-log phase using CelLytic B cell lysis reagent (Sigma). Crude extracts were centrifuged at 14,000 g at 4℃ for 10 min to remove insoluble cell debris. Then, the protein concentration was determined with a BCA assay kit (Thermo Scientific), using bovine serum albumin as a standard. Crude protein samples were stored on ice until assayed. One unit of activity is defined as the amount of enzyme that catalyzes the conversion of one micromole substrate per minute under the experimental conditions. The specific activity was defined as the units of enzyme activity per mg of total protein.

Acetate kinase activity was measured in the direction of acyl phosphate formation (Rose, 1955). The reaction was initiated by adding 0.4 µg of protein sample to 320 µl reaction mixture [200 mM Tris-HCl (pH 7.4), 10 mM ATP, 10 mM $MgCl_2$, 6% (w/v) hydroxylamine hydrochloride (neutralized with KOH before addition), and 267 mM potassium acetate]. The reaction was incubated at 25℃ for 10 min and stopped by adding 320 µl of 10% (w/v) ice-cold trichloroacetic acid. The experimental control

was made with boiled protein samples in the above reaction mixture. Color was developed by adding 320 µl 2.5% (w/v) $FeCl_3$ in 2.0 N HCl. The absorbance at 540 nm was measured with a Biowave II spectrophotometer (WPA). An extinction coefficient of 0.169/mM/cm was used to calculate the activity of acetate kinase.

Phosphotransacetylase activity was measured by monitoring the liberation of coenzyme A at 405 nm (Andersch *et al.*, 1983). The reaction was initiated by adding 2 µg of cell-free extracts to 1 mL of reaction mixture [0.1 M potassium phosphate buffer (pH 7.4), 0.2 mM acetyl-CoA, 0.08 mM 5, 5'-dithio-bis (2-nitrobenzoate)] and then incubated at 25℃ for 10 min. The experimental control was made with boiled protein samples in the above reaction mixture. The absorbance at 405 nm was measured with a Biowave II spectrophotometer (WPA). An extinction coefficient of 13.6/mM/cm was used to calculate phosphotransacetylase activity.

Aldehyde dehydrogenase activity was measured by monitoring NADH oxidation which decreases absorbance at 340 nm (Brown *et al.*, 2011). Protein samples (10 µl) were added to 1 mL reaction mixture [100 mM Tris-HCl (pH 7.6), 1 mM DTT buffer, 5 µM $FeSO_4$, 0.5 mM NADH, 55 mM acetaldehyde] and incubated at 34℃ for 20 min before absorbance measurement. The experimental control was made with boiled protein samples in the above reaction mixture. An extinction coefficient of 6.22/mM/cm was used to calculate aldehyde dehydrogenase activity.

3.3.5 Measurement of cell growth, cellulose consumption and fermentation products

C. cellulolyticum strains were revived in the VM medium with 5 g/L cellobiose, and antibiotic was added if necessary. The cellobiose-grown cultures at an OD_{600} of 0.5-0.7 were used for 1% inoculation into 50 ml fresh VM media with 5 g/L cellobiose, 10 g/L

or 50 g/L Avicel PH101 crystalline cellulose (Sigma) or 10 g/L xylan (Sigma). Each strain had three biological replicates. Cell growth on cellobiose was profiled by monitoring OD600 with a spectrophotometer. When grown on cellulose and xylan, 1 mL of cell culture was sampled periodically and then stored at -80 ℃ for characterizing fermentation kinetics. To quantify the end-point products, cell cultures grown on cellobiose, cellulose and xylan were collected after 6, 23 and 20 days, respectively.

Cell growth on cellulose and xylan was estimated by total protein measurement. The cells were lysed with 0.2 N NaOH/1% w/v SDS solution for 60 min at 25 ℃, and then neutralized with 0.8 N HCl. After centrifugation at 12,000 g for 10 min, the supernatant was used for protein quantification with a BCA assay kit. Then, the protein amount was plotted versus time to profile cell growth.

Cellulose in the fermentation broth was estimated by using a phenol-sulfuric acid method, with glucose as the standard (Dubois *et al.*, 1956). After cell lysis, the residual cellulose was washed twice with distilled water and then hydrolyzed into soluble sugars with 65% H_2SO_4. An aliquot of 150 µl diluted hydrolysate was mixed with 150 µl 5% phenol and 700 µl 98% H_2SO_4 and then incubated for 30 min at 25 ℃. Absorbance at 490 nm was determined with a Biowave II spectrophotometer. Glucose was used as a standard to calculate hexose equivalents.

To measure fermentation products (including lactate, acetate, ethanol, cellobiose and glucose), the fermentation broth was filtered through 0.2 µm filters, acidified with 0.025% H_2SO_4 and then subjected to high-performance liquid chromatography (HPLC) analyses with an Agilent 1200 system (Agilent Technologies) equipped with a variable-wavelength (190 to 600 nm) detector (with UV absorption measured at 245 nm) and an

ion-exclusion column (Aminex HPX-87H; 300 mm × 7.8 mm; Bio-Rad Laboratories, CA). HPLC operating parameters included a column temperature at 65 ℃, 0.025% sulfuric acid as the mobile phase at a flow rate of 0.6 ml/min and 50 µl sample injected (Hemme *et al.*, 2011). Referring to the corresponding standard curves, the concentration of each product was calculated.

The cellulose consumption was estimated by monitoring the change of hexose equivalents. The specific rate of product formation or cellulose consumption was the derivative of the time course plots (Desvaux *et al.*, 2000).

3.3.6 Quantitative real-time PCR

To compare the gene copy number and the transcript amount of *afp* gene in P-AFP transformant and G-AFP integrant, qRT-PCR was conducted as follows. Cell samples were collected from cellobiose-grown cultures at mid-log phase (OD_{600}=~0.45). To compare gene copy number, DNA was extracted by heating at 98 ℃ for 6 min. Heat-treated samples were centrifuged to remove insoluble cell debris. Then, the supernatants were subjected to qRT-PCR analysis using iTaq SYBR Green Supermix with ROX (Bio-Rad) on a Bio-Rad iQ5 thermal cycler. The *recA* gene in the genome was used as an internal calibrator to determine the copy number of *afp* gene. Primers used in qRT-PCR are listed (Table S3.1). Results were analyzed with the Pfaffl method (Pfaffl, 2001).

To compare the transcript amount of *afp* gene by qRT-PCR, cells were lysed by TRIzol Reagent (Invitrogen) followed by total RNA extraction and purification with NucleoSpin RNAII kit (Macherey-Nagel). SuperScript III Reverse Transcriptase (Invitrogen) was applied to convert RNA to cDNA by following the manufacturer's

protocol. cDNA products were diluted as appropriate and used as templates for qRT-PCR. Similarly, results were analyzed with the Pfaffl method using *recA* as the reference gene (Pfaffl, 2001).

3.3.7 Microscopy and flow cytometry

Fluorescence intensity of the anaerobic fluorescence protein was evaluated by fluorescent microscopy and flow cytometry. *C. cellulolyticum* strains at the mid-log phase were harvested, washed twice with the anaerobic PBS buffer and then suspended in the same buffer before loading onto microscope slides. Slides were imaged using Olympus BX51 fluorescence microscope equipped with optical filter sets with excitation at 490 nm and emission at 525 nm for the green fluorescence. The images were collected by an Olympus DP71 digital camera.

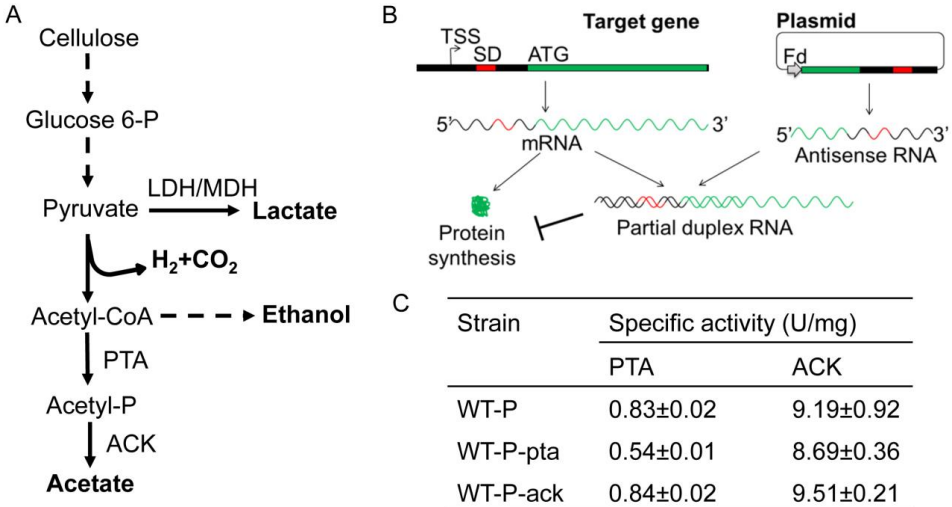
Flow cytometry analysis was performed on a BD Accuri C6 flow cytometer (BD Biosciences) (Li *et al.*, 2014). All samples were diluted with the anaerobic PBS buffer to similar concentrations, then run through the flow cytometer under aerobic condition following the manufacturer's instructions. The run limit was set up as 10,000 events at a slow flow rate, the threshold as 40,000 on FSC-H. The fluorescence was detected with a FL1 detector with a 530/30 filter. The data were collected and analyzed with the CFlow software.

3.4 Results and discussion

3.4.1 Plasmid-based antisense RNA expression

To test the use of asRNA molecules to repress gene expression in *C. cellulolyticum*, we targeted *pta* encoding phosphotransacetylase (PTA) and *ack* encoding acetate kinase (ACK), both of which are key to produce acetate from acetyl-CoA (Figure 3.1A) and

essential for cell survival (Li *et al.*, 2012). For each target gene, its 5' transcriptional region with a length of approximately 120 bp was inserted in a reverse orientation under the control of a ferredoxin promoter to produce asRNAs which will interfere with the stability and translation of target transcripts (Figure 3.1B) (Thomason *et al.*, 2010). The empty asRNA plasmid (pRNAi), customized pRNAi-pta and pRNAi-ack plasmids targeting *pta* and *ack*, respectively, were constructed and transformed into wildtype (WT), generating WT-P control, WT-P-pta and WT-P-ack transformants (where P



C

Strain	Specific activity (U/mg)	
	PTA	ACK
WT-P	0.83±0.02	9.19±0.92
WT-P-pta	0.54±0.01	8.69±0.36
WT-P-ack	0.84±0.02	9.51±0.21

Figure 3.1 (A) Pivotal metabolic pathways in *C. cellulolyticum*. Acetyl-CoA as a key intermediate metabolite, apart from being used to produce ethanol, can be converted to acetyl-phosphate by phosphotransacetylase (PTA, encoded by *pta* gene) and then to acetate by acetate kinase (ACK, encoded by *ack* gene). L-lactate dehydrogenase (LDH) and L-malate dehydrogenase (MDH) are functional in one-step lactate production from pyruvate. Dashed arrows refer to multiple enzymatic reactions. (B) Design of antisense RNAs (asRNAs) to repress *pta* and *ack* genes. For each target gene, the transcriptional region spanning from the predicted transcriptional start site (TSS) to the downstream site approximately 120-bp from the start codon (ATG), containing the Shine-Dalgarno sequence (SD), was amplified and reversely inserted downstream of the ferredoxin (Fd) promoter, generating the Fd::asRNA module. AsRNAs would interfere with the transcription, stability and translation of the target gene. (C) Enzyme assays of PTA and ACK in crude cell-free extracts. Mean and standard deviations of specific enzyme activities were calculated from three biological replicates.

means plasmid-based expression). Then, we examined the repression effect of designed asRNAs by measuring enzyme activities of PTA and ACK in these strains that were grown on 5 g/L cellobiose (Figure 3.1C). Our results showed that (i) PTA activity in WT-P-pta (0.54±0.01 U/mg) was decreased to 65% of WT-P control (0.83±0.02 U/mg) and WT-P-ack (0.84±0.02 U/mg); (ii) ACK activity was barely changed in WT-P-ack (8.69±0.36 U/mg) compared to WT-P (9.19±0.92 U/mg) and WT-P-pta (9.51±0.21 U/mg). The *pta* asRNAs performed much better than *ack* asRNAs in repressing the function of the target gene. The strain expressing *pta* asRNAs was further characterized.

There are a few possible reasons that could explain the observed difference in repression exerted by *pta* and *ack* asRNAs. AsRNA repression follows a threshold linear response (Georg *et al.*, 2011), which suggests that RNA repression only occurs when the abundance of asRNAs is higher than a certain threshold and then with a continuing increase in asRNA abundance, repression will gradually increase. One possibility is that WT-P-pta and WT-P-ack strains have different thresholds mainly depending on the abundance of *pta* or *ack* transcripts. Even though both asRNA expressing modules used the same promoter, it is possible that the abundance of asRNAs varies due to different ribonuclease vulnerabilities. In addition, the RNA structure is important to the physical binding between asRNAs and target transcripts that is necessary for RNA repression. RNA structure prediction (Gruber *et al.*, 2008) found that the *ack* target region is more likely to form a secondary structure (Figure S3.1), which may influence asRNA binding. Although we did not evaluate the extent that these factors could affect RNA repression, our study indicated the variability of

asRNA repression and suggested the potential importance of asRNA design and promoter activity in maximizing RNA repression.

3.4.2 Metabolic changes in knock-down strains

We measured the titers of three major metabolites (lactate, acetate and ethanol) at the end of batch fermentations to determine if acetate production was decreased. With 10 g/L cellulose, the WT-P-pta strain produced lactate, acetate, and ethanol in a molar ratio of 0.93:1.37:1, compared to 1.75:1.49:1 in WT-P control (Table S3.2). The acetate titer in WT-P-pta was decreased about 15% relative to the titer of WT-P (Figure 3.2). Interestingly, the lactate titer was decreased more than 50% in WT-P-pta, but ethanol production was not significantly changed (Table S3.2). When both strains were grown

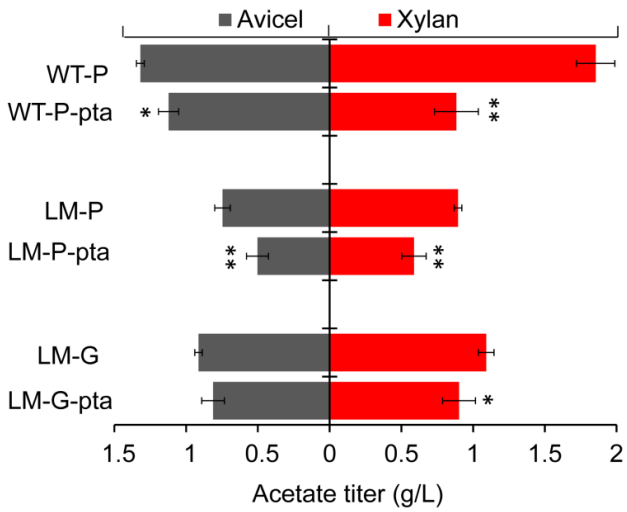


Figure 3.2 Comparison of acetate titers produced on 10 g/L Avicel cellulose (Left) and 10 g/L xylan (Right). Strain names are labeled on the left. Error bar represents the standard deviation of three replicate cultures. The asterisk (*) indicates statistically significant differences between the engineered strain and its corresponding control (student's t test, *P<0.05, **P<0.01).

on 10 g/L xylan, acetate became the major product, which is consistent with previous studies (Li *et al.*, 2012); strikingly, WT-P-pta substantially reduced acetate titer to less

than 48% of WT-P (Figure 3.2), corresponding to a molar ratio of acetate to ethanol of 4.11:1 in WT-P-pta versus 6:1 in WT-P (Table S3.2). Notably, the *pta* asRNAs expressed in WT performed very well in reducing acetate production even though carbon sources greatly affect metabolic profiles. The unexpected decrease in lactate titer on cellulose, as a side effect of manipulating acetate-producing genes, suggests a more sophisticated metabolic regulatory network in this strain, which is also supported by the decreased acetate production in the LM mutant that could rarely produce lactate (Li *et al.*, 2012). However, in *C. thermocellum* the Δ*pta* knockout mutant dramatically increased lactate titer (Argyros *et al.*, 2011), which is in contrast to the accompanying decrease in lactate titer in the *pta* knockdown mutant of *C. cellulolyticum*. It seems like *Clostridium* strains employ different strategies to coordinate metabolic networks. In addition, despite the operability of *pta* disruption in some strains, the resulting effectiveness in acetate formation varies a lot. For example, *pta* deletion reduced acetate by just 14% in *Clostridium tyrobutyricum* (Zhu *et al.*, 2005), but completely eliminated it in *C. thermocellum* (Argyros *et al.*, 2011) and *Thermoanaerobacterium saccharolyticum* (Shaw *et al.*, 2008).

Next, we transformed pRNAi-pta into the LM mutant to generate an LM-P-pta strain that should be deficient in both lactate and acetate production. A control strain, LM-P, was created in parallel by transforming pRNAi that cannot express any specific asRNAs. Metabolic profiling revealed that on 10 g/L cellulose, the LM-P control produced lactate, acetate and ethanol with a molar ratio of 0.04:0.55:1 (Table S3.2); LM-P-pta made negligible lactate, a 33% decrease in acetate titer (Figure 3.2) and an 86% increase in ethanol titer, resulting in a molar ratio of 0.001:0.20:1 (lactate: acetate:

ethanol). On 10 g/L xylan, the titers of lactate and acetate in LM-P-pta were decreased about 82% and 34% (Figure 3.2), respectively, and ethanol titer was slightly increased, corresponding to a molar ratio of 0.06:0.83:1 in LM-P and 0.01:0.51:1 in LM-P-pta (Table S3.2). Hence, with the customized asRNAs expressed in transformants, we successfully manipulated both lactate and acetate producing pathways simultaneously.

Comparing the molar ratio of the three major end products (lactate, acetate and ethanol) in the control and asRNA expressing strains, a general trend was apparent. Both WT-P-pta and LM-P-pta produced an equal molar amount of ethanol by generating less lactate and acetate, regardless of carbon source (Table S3.2). In another word, these repression strains recovered more carbon in the form of ethanol. For instance, when LM-P-pta was grown on cellulose, 83% of the carbons used to produce the three major metabolites were accounted for in the ethanol, 21% higher than the corresponding control (Table S3.2). This demonstrates a successful manipulation of essential metabolic genes to divert carbon flux towards ethanol production.

3.4.3 Chromosomal integration and functional analyses

In light of the effectiveness of *pta* asRNAs in reducing acetate production, we attempted to integrate the asRNA-expressing module into the genome of the LM mutant in such a way that the resulting integrants can work stably and desirably without using antibiotics. To do so, step-wise double-crossover recombination was applied (Heap *et al.*, 2012) (Figure 3.3A). The integration site was immediately downstream of the sole bifunctional acetaldehyde-CoA/alcohol dehydrogenase-encoding gene (*adhE*) in *C. cellulolyticum* (Ccel_3198). The specific integration did not change neighboring sequences but generated an artificial bicistronic operon containing the open reading

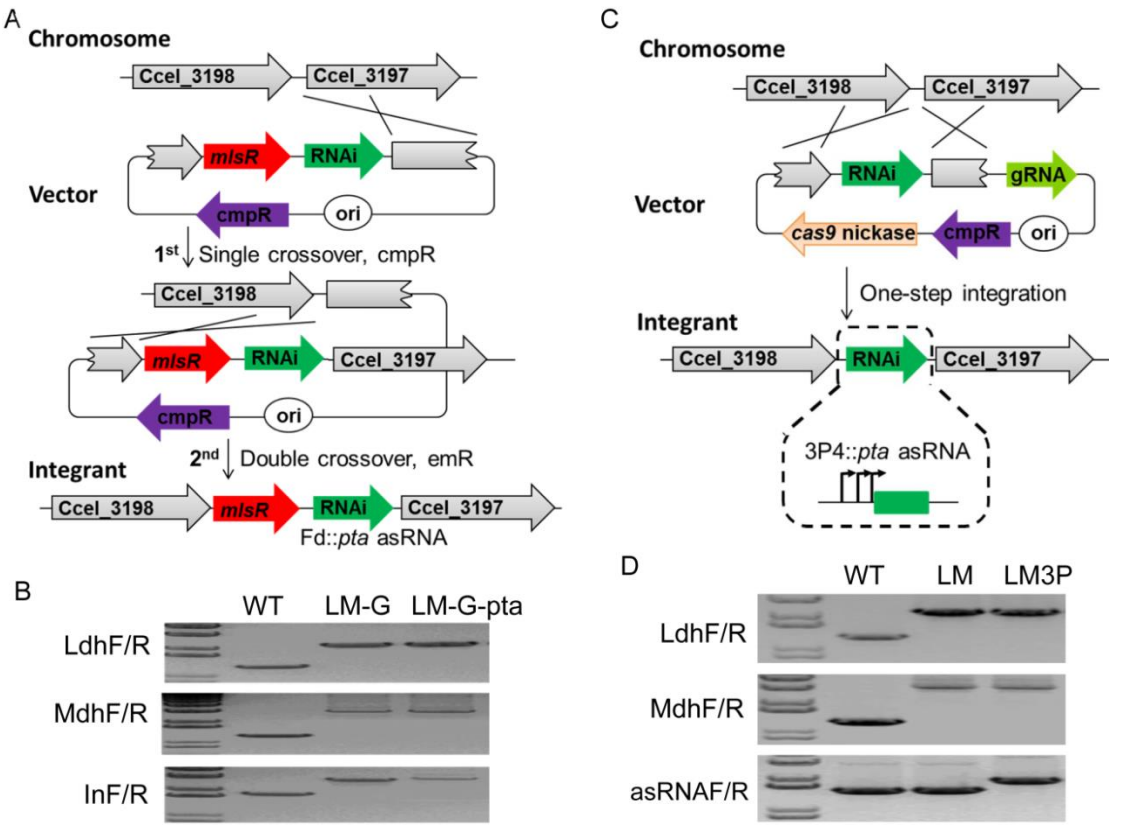


Figure 3.3 Chromosomal integration of functional modules via double-crossover recombination (A and B) and the Cas9 nickase genome editing tool (C and D). The integration site was located in the intergenic region between Ccel_3198 and Ccel_3197. (A) Generation of stable double-crossover clones, LM-G and LM-G-pta, using pLyc045 and pLyc046, respectively. The first step was to screen thiamphenicol-resistant single-crossover clones generated by plasmid integration. The second step was to select erythromycin-resistant double-crossover clones as a result of plasmid excision. Finally, modified genomic loci in candidate clones were verified by PCR with specific primers, LdhF/R for Δ*ldh* identification, MdhF/R for Δ*mdh* identification and InF/R for module integration (B). (C) Generation of stable chromosomal integrants, LM3P and LM3PS, by the Cas9 nickase genome editing tool. By transforming pCas9n-3198D-donor into the LM mutant, integrants were generated within a single step. (D) Modified genome loci in all integrants were then verified by PCR with specific primers, asRNAF/R for RNAi module integration. LM is a double mutant (Δ*ldh* Δ*mdh*); LM-G-pta and LM3P are triple mutants (Δ*ldh* Δ*mdh* Δ*pta*).

frames of *adhE* and *mlsR* under the control of the native *adhE* promoter, consequently enabling counter selection of double-crossover events with erythromycin once *mlsR* gene was expressed. During the screening of double-crossover events, two out of ten

randomly picked colonies were found to be pure LM-G controls (where G indicates genome/chromosome-based expression); however, no pure LM-G-pta integrants were directly isolated because unedited genomes were detected such that another round of plate streaking was required. Targeted integration in LM-G-pta and LM-G was verified by PCR amplification (Figure 3.3B) and amplicon sequencing.

The functionality of the integrated P4::*pta* asRNA module was evaluated by measuring PTA activity and fermentation products. In comparison, the crude extract of cellobiose-grown LM-G-pta integrant presented a lower PTA activity (0.92±0.04 U/mg) that was 89% of LM-G control (1.13±0.06 U/mg), indicating the integrated module was still functional but did not perform as well as the plasmid-based repression in WT-P-pta (Figure 3.1C). Metabolically, the acetate titer in LM-G-pta did not significantly reduce on 10 g/L cellulose but dropped 17% on 10 g/L xylan (Figure 3.2). The overall molar ratios (lactate: acetate: ethanol) were changed from 0.05:0.59:1 in LM-G to 0.01:0.31:1 in LM-G-pta when cultured on cellulose, and correspondingly from 0.09:1.13:1 to 0.04:0.84:1 on xylan (Table S3.2). In general, the integrant was not comparable with the aforementioned transformant in repressing enzymatic and metabolic activities. Previous studies have found that small RNA repression has quantitative characteristics distinct from those of protein-mediated repression (Levine *et al.*, 2008; Georg *et al.*, 2011). One such characteristic is the threshold-linear response as mentioned above. In this case, with a fixed transcription rate of chromosomal *pta* gene, switching from plasmid-based to chromosome-based asRNA expression presumably reduces asRNA dosage, which would weaken the repression of acetate formation.

3.4.4 Evaluation of the gene-dosage effect between transformants and chromosomal integrants

To determine if chromosomal integration mitigates gene activity and how strong the effect is, an *afp* gene encoding anaerobic fluorescent protein was introduced into either the plasmid (P-AFP) or the genome (G-AFP) and then their respective activities were visualized and compared. As expected, P-AFP presented much stronger signal intensity than G-AFP under fluorescent microscopy (Figure 3.4A). Then, quantification of the fluorescence signal by flow cytometry revealed that the signal intensity of P-AFP was 1.73-fold higher than that of G-AFP (Figure 3.4B); when compared to corresponding negative controls (P-CK and G-CK), P-AFP and G-AFP generated 2.79-fold and 1.65-fold greater fluorescent intensity, respectively. The lower signal intensity of G-AFP directly reflects a lower AFP activity and presumably indicates a lower amount of *afp* transcripts produced in G-AFP. Quantitative real-time PCR (qRT-PCR) analysis supported this assumption, showing that relative to G-AFP, P-AFP harbored a 12-fold higher abundance in *afp* gene copy number (Figure 3.4C) and a 36-fold higher abundance in *afp* transcript (Figure 3.4D). These results together indicate that chromosomal integration substantially altered the dosage of gene expression and then diminished gene activity. High-copy number pJIR750 derivatives, including pLyc17 used here to generate P-AFP transformants, have also been reported in *Clostridium perfringens*, which carried as many as 18 copies (Cheung *et al.*, 2009).

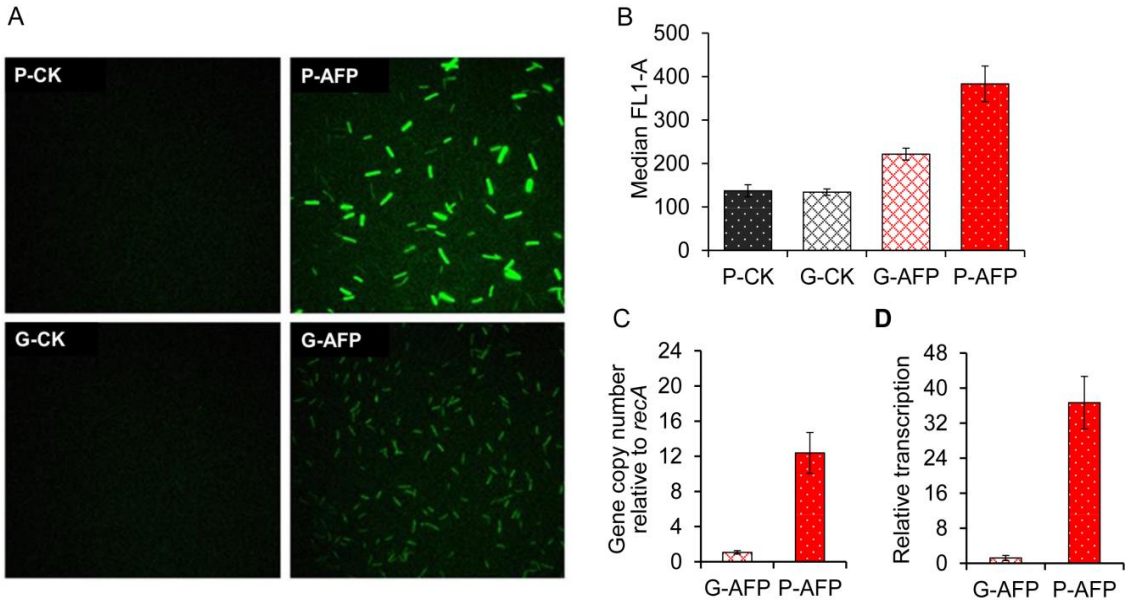


Figure 3.4 Expression of anaerobic fluorescent protein in the P-AFP transformant and the G-AFP integrant. (A) Fluorescence microscopy of cellobiose-grown cells at the mid-log phase. P-CK and G-CK were corresponding controls of P-AFP and G-AFP, respectively. (B) Quantification of fluorescent signal intensity with flow cytometry. (C) Relative *afp* gene copy number in both G-AFP and P-AFP by reference to the single chromosomal *recA* gene. (D) qRT-PCR comparison of *afp* transcript levels between G-AFP and P-AFP, with normalization to *recA* calibrator. Error bar represents the standard deviation of three replicates.

3.4.5 Improved repression of acetic acid production by a tandem repetitive promoter

To overcome the weakened asRNA repression observed with chromosomal integration, we attempted to improve promoter activity and increase asRNA production by generating a tandem promoter consisting of three P4 repeats, named 3P4. P4 is a 36-bp synthetic promoter with an activity comparable to the strong ferredoxin (Fd) promoter in *C. cellulolyticum* (Xu *et al.*, 2015). Switching to a stronger promoter and constructing an artificial promoter cluster have been used to enhance gene expression in both prokaryotes and eukaryotes (Li *et al.*, 2012; Jia *et al.*, 2014). Next, a 3P4::pta asRNA module was constructed and integrated into the LM genome at the same locus by a Cas9 nickase-based editing tool (Figure 3.3C), generating a LM3P integrant (Δ*ldh*

Δ*mdh* Δ*pta*). Mechanistically, the Cas9 nickase protein is directed by a customized guide RNA molecule to recognize the 23-bp target locus through base pairing, and then creates a DNA nick at the locus to trigger a template-dependent homologous recombination (Xu *et al.*, 2015). After transforming the single all-in-one vector, we randomly picked three antibiotic-resistant transformants, all of which were verified to be correct chromosomal integrants by PCR amplification (Figure 3.3D) and amplicon sequencing. The integrated asRNA module will express asRNAs to repress *pta* gene, independent of plasmid-borne Cas9 nickase and antibiotic utilization. Methodologically, although both double-crossover recombination and Cas9 nickase-triggered homologous recombination have the ability to integrate asRNA modules, the latter presents multiple advantages, such as markerless editing, one-step generation and high editing efficiency. Moreover, the Cas9 nickase-based tool has the advantage of using homologous arms as short as 0.2 kb to accomplish high-efficiency genome editing (Xu *et al.*, 2015).

Physiological characterization was performed in antibiotic-free medium. When grown on 5 g/L cellobiose, both LM and LM3P achieved similar biomass yields and presented similar growth rates (μ=0.13 h^{-1}), almost double the growth rate of WT (μ=0.08 h^{-1}) while LM3P's acetate titer decreased by 28% relative to LM (Figure S3.2). Since cellulose concentration significantly affects microbial physiology and metabolism (Desvaux *et al.*, 2000), these strains were further characterized on 10 g/L and 50 g/L cellulose. Profiling of metabolites in the end-point fermentation broth demonstrated a few significant differences. First, the acetate titer in LM3P was decreased by 27% on 10 g/L cellulose (Figure 3.5A), suggesting a much stronger gene repression triggered by the integrated 3P4::pta asRNA module than by the previous Fd::pta asRNA module in

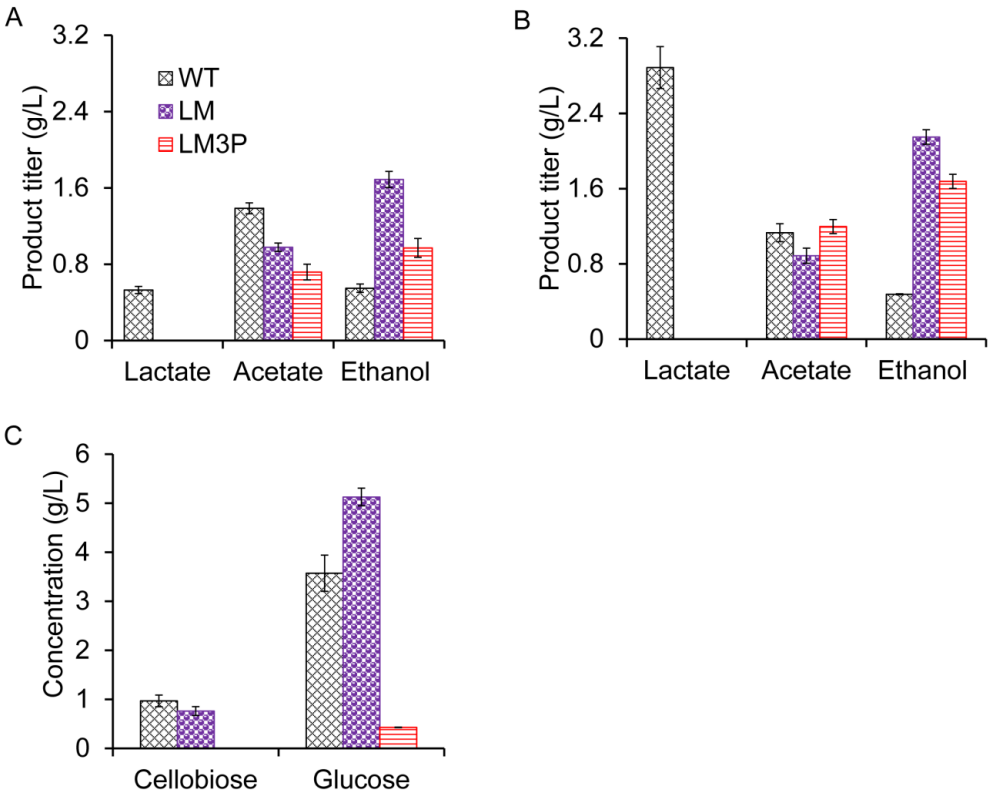


Figure 3.5 Measurements of major end products and released sugars in the end-point fermentation broth. The strains were cultivated on 10 g/L cellulose (A) and 50 g/L cellulose (B). (C) The concentrations of cellobiose and glucose in the fermentation broth were measured when at 50 g/L cellulose. Error bar represents the standard deviation of three replicate cultures.

LM-G-pta. The enhanced repression in LM3P even got close to the plasmid-based repression in LM-P-pta. However, on 50 g/L cellulose LM3P increased acetate by 35% in comparison to LM (Figure 3.5B). This difference in repression caused by varied cellulose loading suggested a complex and flexible metabolic regulation in LM3P, differing from previous reports that loading more cellulose (>6.7 g/L) would reduce the production of acetate and ethanol in WT (Desvaux *et al.*, 2000). Second, LM3P produced less ethanol than LM at both cellulose concentrations, with a greater decrease with 10 g/L cellulose (42%) than with 50 g/L cellulose (22%) (Figure 3.5A and 3.5B). This reduction was not due to a negative effect of chromosomal integration on the

neighboring *adhE* gene because the alcohol dehydrogenase activity responsible for acetaldehyde reduction was not reduced but instead increased in the crude extracts of LM3P (Table S3.3). It is possible that cellular redox balancing strategies changed the reducing power for ethanol production and the carbon flow for acetate production (Desvaux *et al.*, 2000). Lastly, LM accumulated a high level of glucose (5.13 g/L) and cellobiose (0.76 g/L) and correspondingly left less residual cellulose (Figure S3.3) in the fermentation broth while LM3P accumulated only a small amount of glucose (0.43 g/L) (Figure 3.5C). To determine how LM3P affects metabolic regulatory network and cellulose degradation, analyses of metabolomics and transcriptomics will provide valuable clues.

Additionally, we determined fermentation kinetics of LM and LM3P on 50 g/L cellulose. LM3P grew much slower (μ=0.006 h^{-1}) than LM (μ=0.013 h^{-1}), but it finally reached similar cell biomass (Figure 3.6A). Associated with cell growth, the cellulose utilization rate was reduced by 2.6 times in LM3P. While acetate and ethanol gradually accumulated with growth, the specific formation rates were approximately halved in LM3P (Figure 3.6B and 3.6C). Although acetate production ceased after LM entered into the stationary phase, its ethanol production continued for a much longer time. A similar trend was observed with LM3P fermentation. Moreover, LM started to accumulate cellobiose and glucose after cells entered into the stationary phase (Figure 3.6B). Glucose accumulation was obviously faster than cellobiose and seemingly lasted much longer. These fermentative characteristics provide clues to improve microbial consortia using LM as a cellulose degrader for biofuel production.

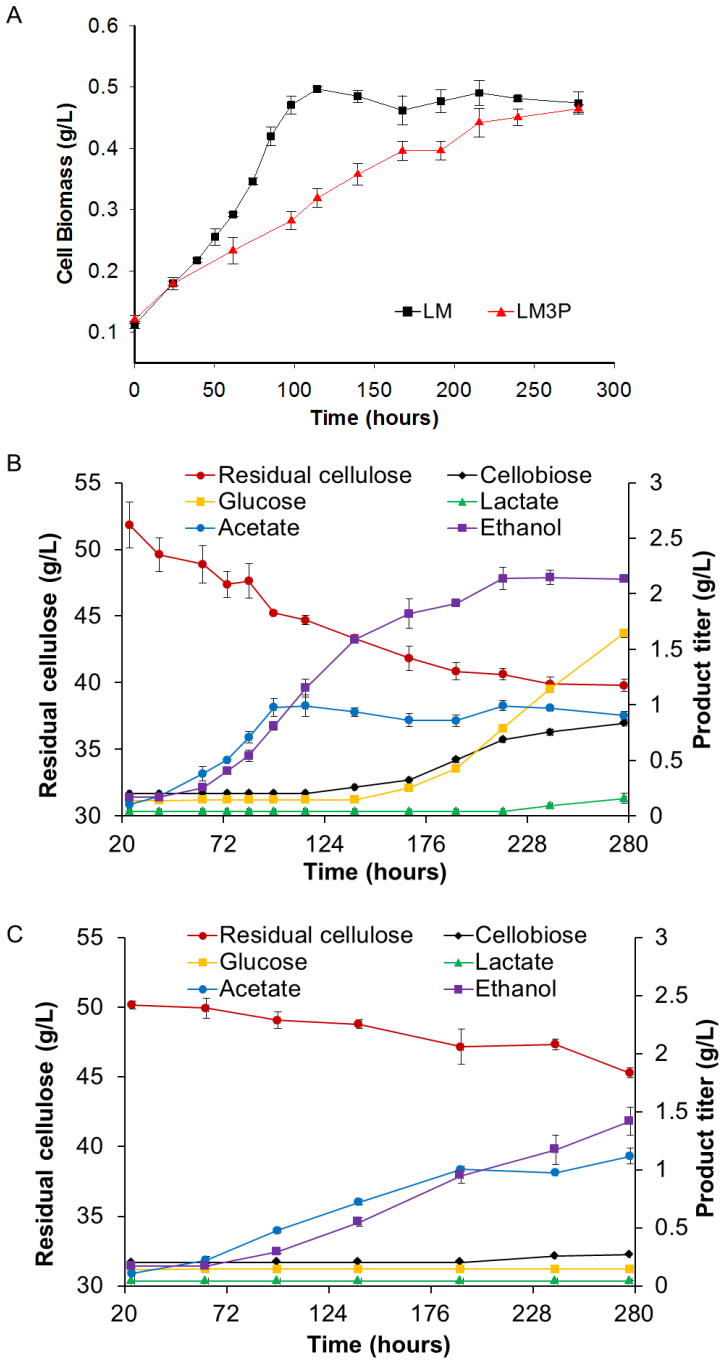


Figure 3.6 Fermentation kinetics of LM and LM3P on 50 g/L cellulose. Cell growth of both strains was profiled (A). Cellulose consumption and product formation were monitored during the fermentation of LM (B) and LM3P (C). Error bar represents the standard deviation of three replicates.

3.5 Conclusions

Antisense RNA-mediated repression worked well in both the *C. cellulolyticum* wildtype and LM mutant to repress *pta* expression thereby reducing acetate production in these strains. Combined utilization of gene repression and Cas9 nickase genome editing realized a one-step markerless integration of an upgraded antisense RNA-expressing module into the chromosome, genetically allowing stable manipulation of essential genes and providing a technical demonstration of the unmatched editing simplicity and efficiency of this system over double-crossover recombination. The tandem promoter strategy dramatically improved repression of acetate formation in the integrants. This combinatorial strategy significantly expanded our ability to manipulate more diverse genes for functional characterization and strain engineering.

Chapter 4 Dockerin-containing Protease Inhibitor Protects Key Cellulosomal Cellulases from Proteolysis in *Clostridium cellulolyticum*

4.1 Abstract

Cellulosomes are multienzyme machines for lignocellulosic biomass degradation in cellulolytic Clostridia. Better understanding of the mechanism of cellulosome regulation would allow us to improve lignocellulose hydrolysis. It is hypothesized that cellulosomal protease inhibitors would regulate cellulosome architecture and then lignocellulose hydrolysis. Here, a dockerin-containing protease inhibitor gene (*dpi*) in *Clostridium cellulolyticum* H10 was characterized by mutagenesis and physiological analyses. The *dpi* mutant had a decreased cell yield on glucose, cellulose and xylan, lower cellulose utilization efficiency, and a 70% and 52% decrease of the key cellulosomal components, Cel48F and Cel9E, respectively. Quantitative PCR showed that *cipC*, *cel48F* and *cel9E* all had similar transcript levels, although all were decreased by ~40% in the mutant compared to the wild type. This suggests that decreased cellulose degradation efficiency in the mutant may be caused by both lower expression of the *cip-cel* gene cluster and higher proteolysis of cellulosomal components. Disruption of *cel48F* and *cel9E* severely impaired cell growth on cellulose but *cel48F* disruption completely abolished cellulolytic activity. Purified recombinant Dpi showed inhibitory activity against cysteine protease. Taken together, Dpi protects key cellulosomal cellulases from proteolysis in H10. This study is the first to identify the physiological importance of cellulosome-localized protease inhibitors in Clostridia.

Keywords: cellulosome; cellulase; protease inhibitor; biofuels; *Clostridium cellulolyticum*

4.2 Introduction

Consolidated bioprocessing (CBP) of lignocellulosic biomass integrates the microbial activities of hydrolase production, saccharification, and fermentation into a single step, and is regarded as a promising approach for production of low-cost biofuels (Lynd *et al.*, 2002). Cellulolytic Clostridia (e.g., *Clostridium thermocellum* and *C. cellulolyticum*) as CBP-enabling candidates have been sequenced (Hemme *et al.*, 2010) and are being engineered with higher efficiency in cellulose hydrolysis and biofuel synthesis (Guedon *et al.*, 2002; Brown *et al.*, 2011; Nakayama *et al.*, 2013). Like some cellulose-degrading fungi (Dashtban *et al.*, 2009), these bacteria secrete diverse lignocellulose-degrading enzymes to synergestically decompose lignocellulosic biomass. Some of these enzymes are assembled onto cell surface-attached scaffoldin proteins by dockerin-cohesin interactions, generating multi-enzyme complexes called cellulosomes (Bayer *et al.*, 2004; Doi *et al.*, 2004; Fontes *et al.*, 2010). Biochemical studies on individual glycoside hydrolases have been widely conducted (Cantarel *et al.*, 2009) with the goal of boosting their industrial applications (Kuhad *et al.*, 2011). To date, however, only limited reports on the *in vivo* roles of glycoside hydrolases in cellulolysis are available (Perret *et al.*, 2004; Tolonen *et al.*, 2009; Olson *et al.*, 2010). Proper cellulosome assembly is critical to efficiently degrade cellulose when compared with free hydrolases (Schwarz, 2001; Maamar *et al.*, 2004). To accomplish CBP in Clostridia, bridging several knowledge gaps (e.g., physiological functions of cellulosomal components, regulatory mechanisms of cellulosome maintenance and gene expression) is necessary.

Cellulosomal heterogeneity is reflected in the varied abundance of components in mature cellulosomes when grown on different carbon sources (Han *et al.*, 2005; Gold *et al.*, 2007; Blouzard *et al.*, 2010). The synergistic catalysis of glycoside hydrolases is important for efficient cellulolysis (Lynd *et al.*, 2002). For example, of 62 predicted dockerin-containing proteins in the genome of *C. cellulolyticum*, 50 were identified in isolated cellulosomes, 36 when grown on cellulose, 30 on xylan, and 48 on hatched wheat straw (Blouzard *et al.*, 2010). The 26 kb *cip-cel* gene cluster in *C. cellulolyticum* containing 12 genes (*cipC*, *cel48F*, *cel8C*, *cel9G*, *cel9E*, *orfX*, *cel9H*, *cel9J*, *man5K*, *cel9M*, *rgl11Y*, and *cel5N*) produces two large transcripts, a 14-kb mRNA carrying the first five coding sequences and a less abundant 12-kb mRNA with the coding sequences of the genes located in the 3′ part of the cluster (Maamar *et al.*, 2006). CipC, Cel48F, and Cel9E are three major cellulosomal components in *C. cellulolyticum* (Maamar *et al.*, 2004; Perret *et al.*, 2004). Previous studies showed *cipC* disruption and *cel48F* repression severely impaired cellulolysis (Maamar *et al.*, 2004; Perret *et al.*, 2004). However, how microorganisms adapt and maintain their cellulosomes under different environmental conditions remains a mystery.

In addition to dockerin-containing glycoside hydrolases, other enzymes (e.g., esterases, polysaccharide lyases, chitinase and peptidases) have been predicted and/or found to be on cellulosomes (Gold *et al.*, 2007; Blouzard *et al.*, 2010). Kang *et al.* cloned and studied three serine protease inhibitors, Serpin1-3, from *C. thermocellum* (Kang *et al.*, 2006). Serpin1 was able to interact with *cipA* cohesion and inhibit subtilisin activity. Several cysteine peptidase inhibitors that are likely cellulosomal components in *C. cellulovorans* also exhibited inhibitory activities against

representative plant proteases, papain and ficin (Meguro *et al.*, 2011). Proteomics studies on isolated *C. cellulolyticum* cellulosomes also identified a Chagasin_I42 component that might be a cysteine protease inhibitor (Blouzard *et al.*, 2010). Bacterial proteases are involved in several biological processes including protein turnover, sporulation and conidial discharge, germination, enzyme modification, nutrition, and regulation of gene expression (Rao *et al.*, 1998). Considering the localization of cellulosomal protease inhibitors, it was speculated that they might be responsible for self-protection to avoid proteolysis of exogenous proteases (Meguro *et al.*, 2011), or for cellulosome remodeling (Schwarz *et al.*, 2006). So far, the physiological importance of these inhibitors has not been investigated.

C. cellulolyticum as a non-ruminal mesophilic cellulolytic model is relatively susceptible to genetic manipulation (Petitdemange *et al.*, 1984; Desvaux, 2005). A dockerin-containing protease inhibitor gene (*dpi*) (*Ccel_1809*) from *C. cellulolyticum* H10 was chosen to determine the *in vivo* functions of this kind of cellulosome-localized protease inhibitor. The protein encoded by the *dpi* gene has been identified in active cellulosomes (Blouzard *et al.*, 2010). In this study, we hypothesized that the cellulosomal protease inhibitor Dpi would be enzymatically functional and affect insoluble carbon utilization by regulating cellulosomal components. To test these hypotheses, a *dpi* mutant was characterized at the phenotypic, physiological and protein levels. We discovered that Dpi was able to effectively block cysteine protease inhibitor activity, protect key *C. cellulolyticum* cellulosomal cellulases, and allow cells to maintain high-efficiency cellulolysis. Three additional mutants, a *trans*-complementation strain (*dpi/over*) and *cel48F* and *cel9E* mutants were constructed to

further identify the physiological importance of *dpi*, *cel48F* and *cel9E* genes in degrading cellulose. This study provides new insights into our understanding of cellulosomal protease inhibitor-mediated protection of cellulosomal components from proteolysis in *C. cellulolyticum*.

4.3 Results

4.3.1 dpi mutant construction and phenotypic analysis

To examine whether the cellulosome-localized protease inhibitor Dpi plays an important role in insoluble carbon utilization, a *dpi* mutant was constructed using a mobile group II intron based gene inactivation system (Supporting information: Figure S4.1) (Heap *et al.*, 2010; Li *et al.*, 2012). Growth of the mutant was examined on both soluble (glucose and cellobiose) and insoluble substrates (cellulose and xylan). With 10 g/L cellobiose, there was no significant difference observed between the *dpi* mutant and WT in terms of growth rate and maximal biomass (Figure 4.1A). However, with 10 g/L glucose, the mutant showed a 22% decrease in maximal cell density compared to WT although no difference was observed in the growth rate during exponential phase (Figure 4.1B). With 10 g/L cellulose, the mutant presented a slower growth rate and its maximal biomass was 52% of WT (Figure 4.1C). Similar results were observed on xylan, which showed a 24.4% decrease in maximal *dpi* mutant biomass (Figure 4.1D). Therefore, the inactivation of the *dpi* gene affected cell growth on glucose, cellulose and xylan but not cellobiose.

The cellulose degradation efficiency of the *dpi* mutant and WT was also examined. For all time points tested, the mutant left higher amounts of cellulose residue in the fermentation broth (Figure 4.1E). After entering stationary phase, 46.7$\pm$10.2%

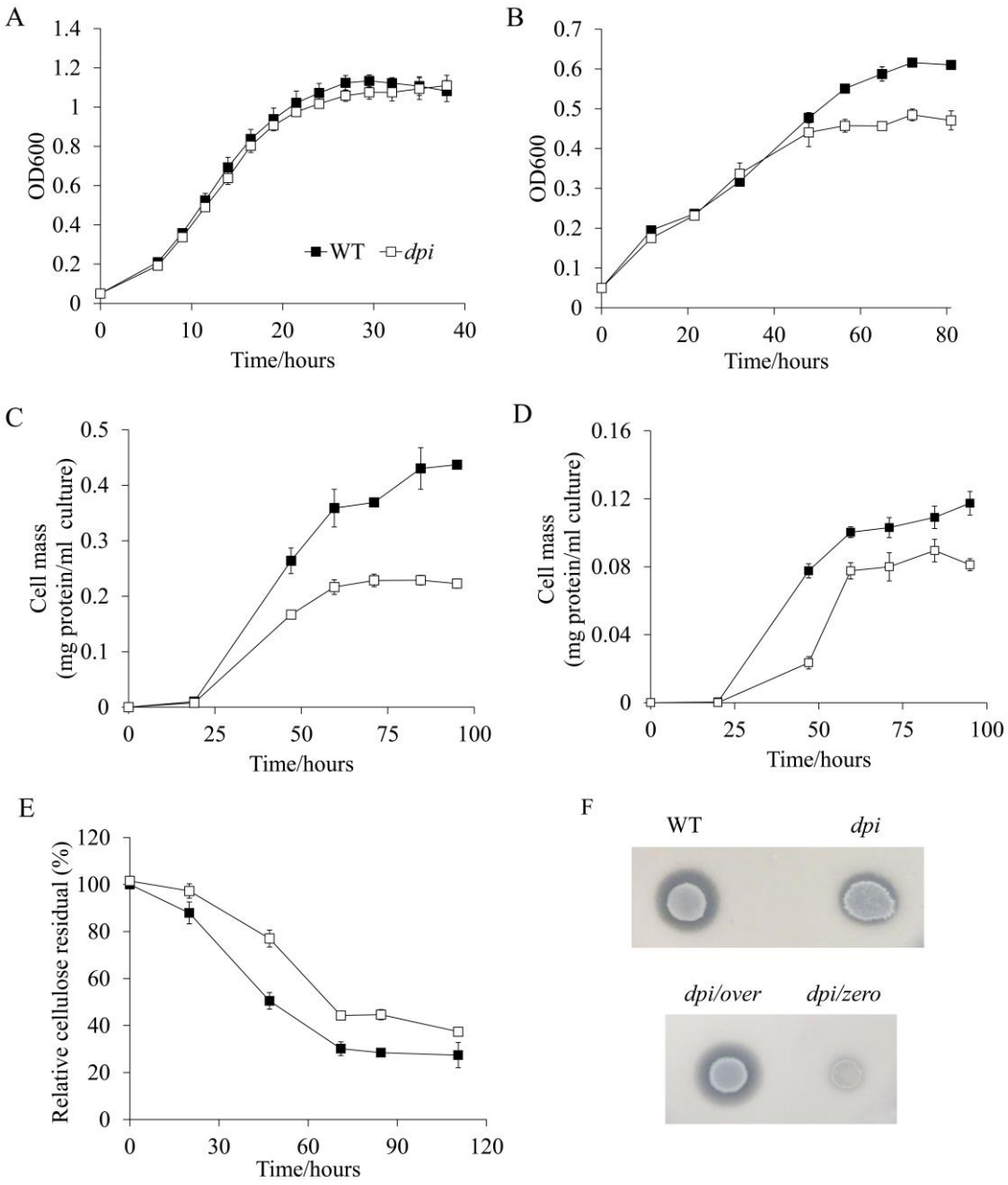


Figure 4.1 Growth profiling of WT, *dpi* mutant, *dpi/over* and *dpi/zero* strains. Cell densities of WT and dpi mutant on 10 g/L cellobiose (A) and 10 g/L glucose (B) were estimated by monitoring OD600. Cell mass obtained on 10 g/L Avicel cellulose (C) and 10 g/L xylan (D) were determined by total protein quantification. Cellulose residual percentage (E) was calculated by dividing the cellulose residual amount by the initial cellulose input. The means and standard deviations were calculated from three independent measurements. Avicel degradation tests (F) of WT, *dpi* mutant, *dpi/over* and *dpi/zero* strains were performed on cellulose-containing top-agar plates.

more cellulose remained in the mutant culture compared to the WT. To visualize differences in cellulose consumption between the WT and mutants, a hydrolysis test on cellulose-containing top-agar plates was performed (Maamar *et al.*, 2004). In this test, cellulose degradation results in a transparent halo surrounding colonies, with a larger halo indicating higher amounts of cellulose degradation. The WT developed a large halo while the mutant developed a smaller halo (Figure 4.1F). *Trans*-complementation of the *dpi* mutant (*dpi/over* strain) restored the cellulolytic phenotype to produce a halo similar to WT (Figure 4.1F). These results confirmed that inactivation of *dpi* negatively affected cellulose utilization in *C. cellulolyticum*.

4.3.2 Molecular analysis of the mutant cellulolytic system

To investigate how the mutant reduced cellulose utilization, we isolated cellulosome fractions (Fc) and free extracellular fractions (Ff) from cellulose-grown cultures in the mid-log phase and then compared protein component patterns by SDS-PAGE. There were three prominent components in the WT Fc (Figure 4.2A), which is consistent with previous reports that CipC, Cel48F and Cel9E were the three most abundant cellulosomal components (Maamar *et al.*, 2004; Perret *et al.*, 2004). The most significant difference observed in the mutant was the reduced abundance of two major bands in Fc, (labeled B1 and B2). There were two additional minor bands also showing decreased density (labeled B3 and B4) in the mutant. B1- B4 bands were verified to be Cel48F, Cel9E, Cel9J and Cel9M, respectively by mass spectrometry analysis (Table S4.1). Interestingly, with equal Fc loading, the density of CipC was not altered due to *dpi* disruption. Analysis of Ff also showed that some bands, but not all, were obviously

altered in abundance between the WT and mutant. Thus, disruption of *dpi* gene significantly altered key cellulosomal components on the CipC scaffoldin.

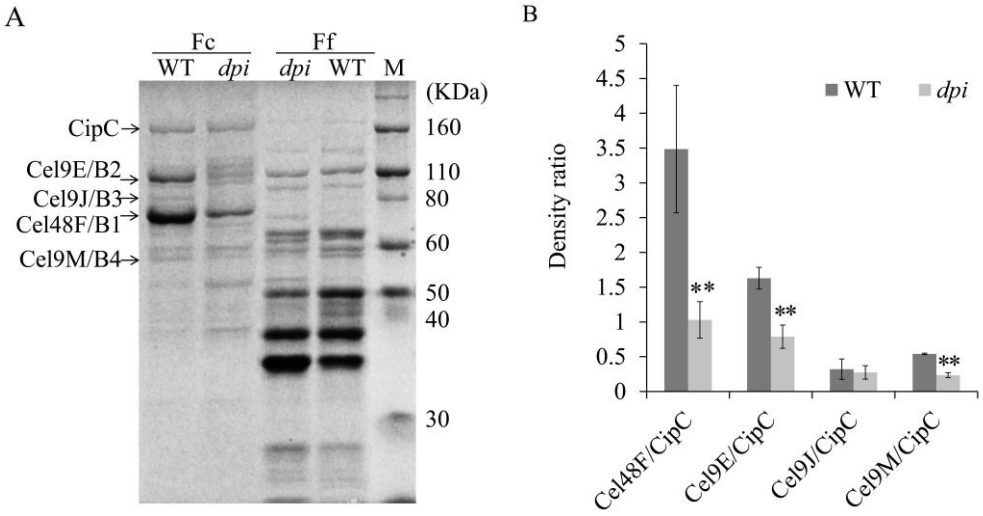


Figure 4.2 Composition of the cellulolytic system of the dpi mutant and WT. A. SDS-PAGE analysis of cellulosome fraction (Fc) and free extracellular protein fraction (Ff) isolated from 10 g/L Avicel cellulose-grown WT and *dpi* mutant at the mid-logarithmic phase. Bands labeled with enzyme names on the left were identified by mass spectrometry. B. Densitometry analysis of several enzymes in Fc fractions. Ratios of Cel48F/CipC, Cel9E/CipC, Cel9J/CipC and Cel9M/CipC were calculated based on staining intensity for each protein. The means and standard deviations were calculated from gels of three biological replicates. The asterisks denote significant difference between WT and *dpi* mutant (**$p < 0.01$, by Student's t test).

CipC as a structural protein has eight cohesion domains for assembly of dockerin-containing enzymes (Desvaux, 2005). The relative abundances of the above four enzymes on CipC were quantified by densitometry analysis (Figure 4.2B). The ratio of Cel48F/CipC was 4.48 in WT but dropped to 1.03 in the mutant. Similarly, the *dpi* mutation caused the Cel9E/CipC ratio to drop from 1.63 to 0.79, and the Cel9M/CipC ratio from 0.54 to 0.24. Correspondingly, the ratios of Cel48F/CipC, Cel9E/CipC and Cel9M/CipC decreased by 70%, 52% and 56.5% in the mutant ($p <$

0.01), respectively. However, the Cel9J/CipC ratio did not show a statistically significant change between WT and the mutant ($p > 0.05$). Therefore, the reduced abundance of several cellulosomal components on scaffoldin may further explain the lowered cellulolysis.

4.3.3 Quantification of cipC, cel48F *and* cel9E *transcripts, and cellulsome production*

The altered abundances of two prominent cellulosomal components (Cel48F and Cel9E) between the WT and mutant could be caused by two possibilities: selective proteolysis and differential transcript levels. *cipC, cel48F* and *cel9E* are all located in the *cip-cel* gene cluster and co-transcribed (Maamar *et al.*, 2006), so changes in the level of transcription of these genes would occur simultaneously. qPCR analysis revealed that the mutant had very similar amounts of *cipC, cel48F* and *cel9E* transcripts during exponential growth, indicating that the difference in component abundance (Figure 4.2A) is independent of transcription level (Figure 4.3A). However, the mutant transcription level was around 60% of that observed in the WT. A lower expression level of the *cip-cel* gene cluster would reduce CipC availability and lead to less cellulosomal assembly on the cell surface. To test this, cellulosome productivity, which equals the ratio of isolated cellulosome to total cellular biomass, was determined in cellulose-grown cells at the mid log phase. The WT and *dpi* mutant both presented similar productivity (Figure 4.3B), 0.14 ± 0.01 mg cellulosome complex per mg cellular biomass. Thus, *dpi* inactivation did not affect cellulosome productivity, which further supports the hypothesis that the reduced cellulolysis observed in the *dpi* mutant is caused by the decreased abundance of several major cellulosomal components.

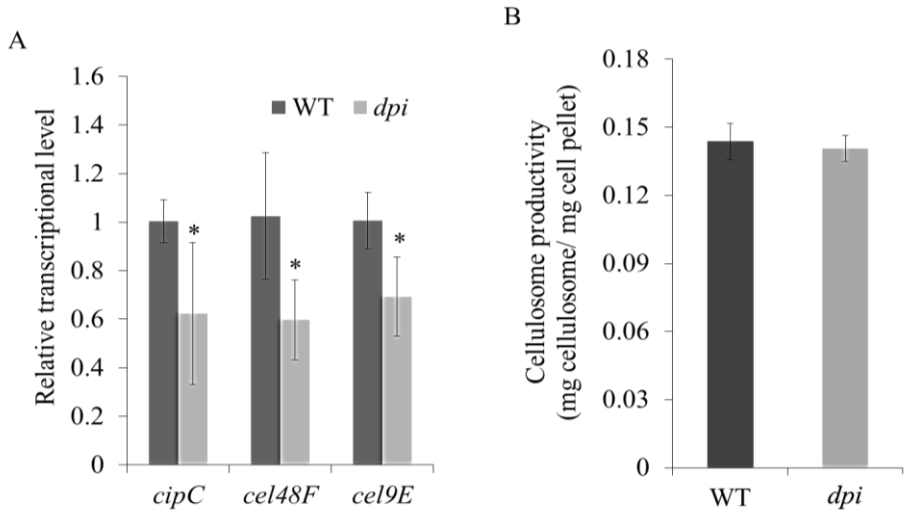


Figure 4.3 Comparison of transcript levels and cellulosome productivity. Relative transcript levels (A) of the genes *cipC*, *cel48F* and *cel9E* in WT and *dpi* mutant grown on 10 g/L Avicel cellulose at the mid-logarithmic phase were compared by normalizing with the expression of the calibrator gene *recA*. Cellulosome productivity (B) of cellulose-grown WT and *dpi* mutant was calculated by dividing the isolated cellulosome amount by the total protein amount in the cell pellet. The means and standard deviations were calculated from the values of three biological replicates. The asterisk means significant difference between WT and *dpi* mutant (*$p < 0.05$, by Student's t test).

4.3.4 Characterization of cel48F *and* cel9E *mutants*

To verify the importance of Cel48F and Cel9E in cellulolysis, *cel48F* and *cel9E* mutants were created (Figure S4.2). There was no obvious growth detected with the *cel48F* mutant on 10 g l^{-1} Avicel cellulose (Figure 4.4A), indicating that Cel48F might be a pivotal cellulase for cellulolytic activity in this bacterium. This result was consistent with previously reported results using antisense RNA to knock down *cel48F* expression (Perret *et al.*, 2004). Also, direct *cel48F* inactivation had a more obvious effect on cellulolysis than antisense RNA. The *cel9E* mutant showed very weak growth on Avicel cellulose with a 64.5±0.9% decrease in cell mass compared with WT (Figure 4.4A). Additionally, although both Cel48F and Cel9E are cellobiohydrolases responsible for degrading cellulose to soluble sugars, disruption of *cel48F* and *cel9E*

unexpectedly affected cell growth on 5 g l^{-1} cellobiose, producing 29.4±0.7% and 24.9±1.2% less biomass in the stationary phase compared with WT, respectively (Figure 4.4B), indicating that Cel48F and Cel9E might also play a role in cell growth on cellobiose.

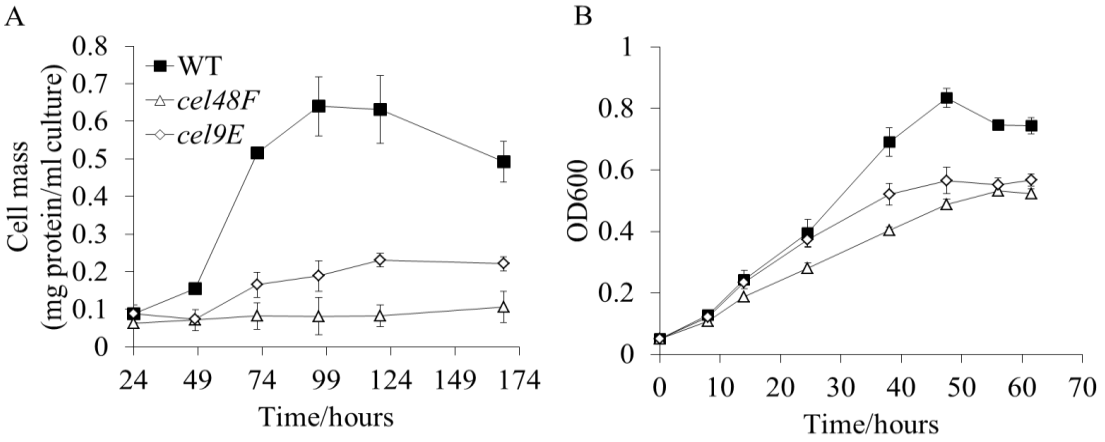


Figure 4.4 Growth curves for WT, cel48F and cel9E mutants grown on 10 g/L Avicel cellulose (A) and 5 g/L cellobiose (B). The means and standard deviations were calculated from three independent measurements at each time point.

To evaluate the potential polar effect caused by the insertion of a mobile group II intron, the transcript amounts of *cel8C* and *orfX*, the first downstream genes of *cel48F* and *cel9E*, respectively, were compared between WT and mutants. Since the *cel48F* mutant cannot grow on cellulose and the *cip-cel* gene cluster is expressed on cellobiose (Mussolino *et al.*, 2012), the polar effect was evaluated in cellobiose-grown cells. qPCR analyses revealed that transcripts of *cel8C* in the *cel48F* mutant and *orfX* in the *cel9E* mutant were reduced to 42% and 39%, respectively (Figure 4.5). This means that the polar effect occurred in both mutants, which might also partially contribute to the observed defect in cellulolysis. However, the polar effect was insufficient to cause the total abolishment of *cel48F* mutant growth on cellulose. Therefore, these defects

observed in the *cel48F* and *cel9E* mutants are a combinational effect of gene inactivation and polar effect on downstream genes.

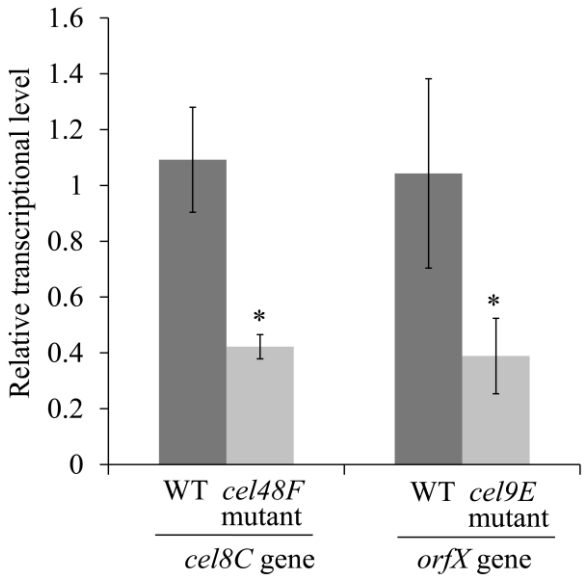


Figure 4.5 Detection of the polar effect by quantitative real-time PCR. Relative transcript levels of the first downstream genes, *cel8C* in *cel48F* mutant and *orfX* in *cel9E* mutant were compared with these in WT by normalizing with *recA*. All strains were grown on 5 g/L cellobiose at the mid-logarithmic phase. The means and standard deviations were calculated from three biological replicates. The asterisk means significant difference between WT and mutants (*$p < 0.05$, by Student's t test).

4.3.5 Expression and enzymatic activity assay of recombinant Dpi protein

The *dpi* gene putatively encodes a dockerin-containing protease inhibitor (Dpi). Motif scanning of its putative peptide sequence (316 aa) predicted a signal peptide at the N terminal (1-36 aa), a dockerin_1 domain in the middle (88-108 aa), and two Chagasin_I42 domains (135-223 aa and 226-315 aa) (Letunic *et al.*, 2012) (http://smart.embl.de/). Using the structure prediction tool Phyre (Kelley *et al.*, 2009), we constructed a visible model of the Dpi protein (44-316 aa) consisting of a type I dockerin domain and two Chagasin_I42 domains (Figure 4.6A). The predicted type I

dockerin domain is very similar to reported crystalline structures (Lytle *et al.*, 2001; Pinheiro *et al.*, 2008). The middle Chagasin_I42 domain has three conserved loops (DE, BC and FG) able to form a flexible wedge that may block the active site of cysteine protease according to previous studies (Figueiredo da Silva *et al.*, 2007; Casados-Vazquez *et al.*, 2011). However, the C-terminal Chagasin_I42 domain presents a much less compact structure even though conserved amino acids building up these three key loops exist, suggesting that these two Chagasin_I42 domains might have different enzymatic features.

The encoding sequence without the N-terminal signal peptide (44-316 aa) was cloned into pET28a (+) and then expressed in *E. coli*. The recombinant Dpi harboring a His tag at the N terminal was produced with high yield (in lane 2, Figure 4.6B) and purified with Ni (+) affinity chromatography under native conditions (in lane 4, Figure 4.6B). The inhibitory activity of the recombinant Dpi was examined against commercial trypsin, chrymotrypsin, papain and pepsin. In this test, only papain was effectively inhibited by Dpi (Table 1). This result is in accordance with functional prediction since papain belongs to the cysteine protease family (Rawlings *et al.*, 2012). Moreover, in the reaction with 0.85 nM papain, the residual papain activity was gradually decreased by increasing Dpi dose (Figure 4.6C). At a concentration of 1.8 nM of Dpi, half of the maximal papain activity was repressed (IC_{50}=1.8 nM). When the Dpi concentration reached 3.31 nM (the molecular ratio of Dpi to papain is 3.89), less than 2% protease activity remained. Therefore, Dpi is an effective inhibitor of cysteine protease.

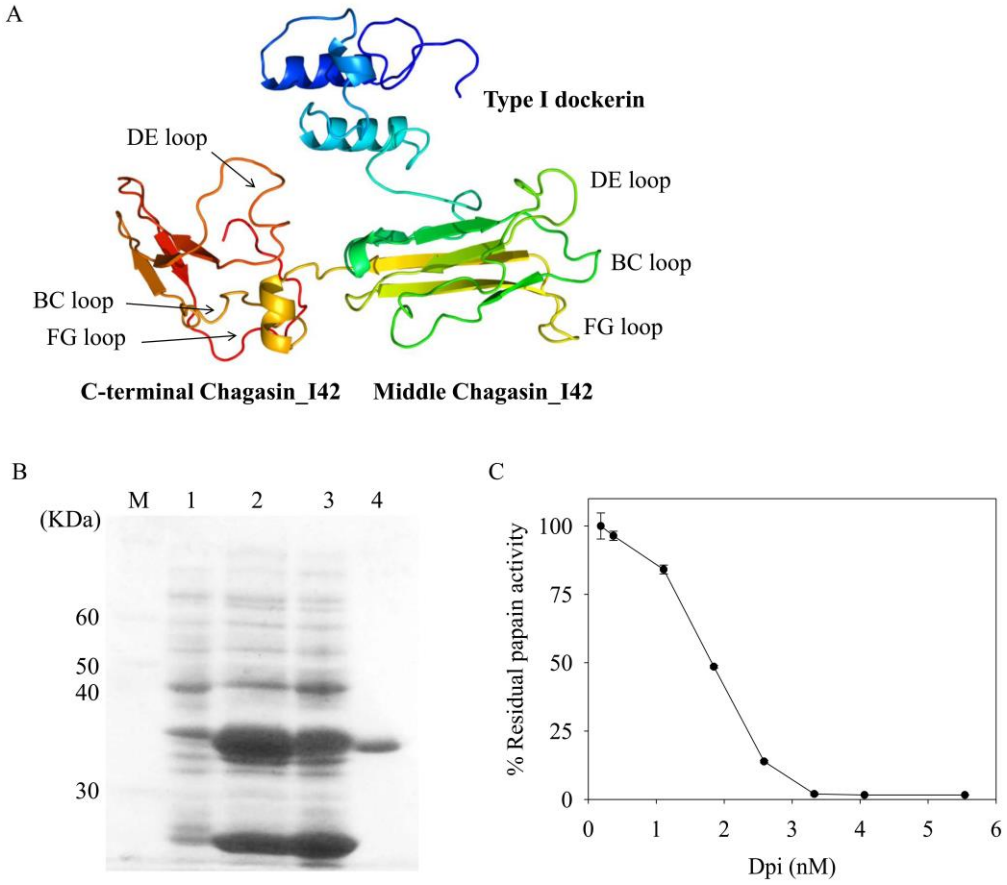


Figure 4.6 Characterization of Dpi protein. A. Modeling structure of Dpi protein (44-316 aa) established by Phyre based on the conserved type I dockerin and chagasin_I42 domains. B. SDS-PAGE analysis of protein samples from non-induced (lane 1) and induced (lane 2) *E. coli* harboring pET28a(+)-dpi vector, supernatant of induced cell lysate (lane 3) and purified Dpi (lane 4) after affinity chromatography. C. Inhibitory efficiency of the purified Dpi against papain. The residual proteolytic activities of 0.85 nM papain were determined with the increase of Dpi dose from 0.18 nM to 5.54 nM. The half maximal inhibitory concentration (IC50) was 1.8 nM at this given condition. The means and standard deviations were calculated from three independent measurements.

Table 4.1 Inhibitory activity of the recombinant Dpi against commercial proteases.

Peptidase	Property	% Inhibitory activity
Papain	A cysteine endopeptidase	52.81±1.05
Trypsin	A pancreatic serine protease	2.58±2.15
Chymotrypsin	A serine endopeptidase	ND
Pepsin	A aspartate protease	ND

ND, no inhibition was detected.

4.4 Discussion

Blouzard *et al.* identified Dpi as a cellulosomal component upon cell growth on different substrates (cellulose, xylan, and wheat straw) of *C. cellulolyticum* (Blouzard *et al.*, 2010). The present study reports the physiological functions of this cellulosome-localized protease inhibitor. Disruption of *dpi* affected cell growth on glucose, cellulose and xylan, but not on cellobiose. The *dpi* mutant grown with glucose entered into the stationary phase slightly earlier than WT, which could be caused by a temporal expression of *dpi* or its protease targets. The growth phase-dependent expression of a cell wall-associated cysteine protease has been found in *Staphylococcus epidermidis* (Oleksy *et al.*, 2004). If the antagonistic activity of Dpi against protease targets was disrupted or abolished, the resulting hyperactive proteolysis would do damage to functional proteins/enzymes essential for cell growth. Even though glucose and cellobiose are both soluble carbon substrates, the cellobiose-grown mutant did not exhibit obvious differences from WT. It seems like cellobiose catabolism is not associated with cellulosomal Dpi functionalization. In addition to cell growth changes on cellulose and xylan, the mutant also presented lower efficiency in cellulolysis due to impairment of key cellulosomal components. Disturbance of cellulosomal composition negatively affects enzymatic activities to hydrolyze insoluble carbons (Maamar *et al.*, 2004; Perret *et al.*, 2004) , thus reducing the amount of usable sugar available to support cell growth. Dpi is thus physiologically associated with cell growth and biomass utilization in a substrate-dependent manner.

The existence of cellulosomal protease inhibitors raises questions as to how the cellulosome is maintained or modified under diverse environmental conditions

(Schwarz *et al.*, 2006). Similar inhibitors from *C. thermocellum* and *C. cellulovorans* exhibited inhibitory activities against bacterial and plant proteases (Kang *et al.*, 2006; Meguro *et al.*, 2011). In the present study, *C. cellulolyticum* Dpi was shown to be an effective inhibitor of cysteine protease. It is the sole protease inhibitor encoded by *C. cellulolyticum* according to the MEROPS database (Rawlings *et al.*, 2012). Functionalization of cellulosomal Dpi depends on cysteine proteases that can be secreted out of the cell. Whole genome mining uncovered seven of 25 potential secretory peptidases belonging to the cysteine protease family and can be further divided into three subfamilies, C1A (Ccel_2442), C82 (Ccel_2590) and C40 (Ccel_0747, Ccel_1652, Ccel_1956, Ccel_2128 and Ccel_2940) (Rawlings *et al.*, 2012). Interestingly, Ccel_2442 from C1A, carrying a dockerin domain and two Chagasin domains, is a papain-like cysteine protease, which is considered as the most probable target of Dpi but has never been identified in active cellulosomes. Both C82 and C40 are involved in bacterial cell-wall modification. Further efforts will focus on identifying the *in vivo* inhibitory targets of Dpi.

CipC scaffoldin has been used as an internal calibrator to quantify the relative abundance of cellulosomal components (Perret *et al.*, 2004). Our analysis determined that the abundances of Cel48F, Cel9E and Cel9M relative to CipC were decreased significantly in the mutant. These changes were largely caused by higher proteolysis induced by Dpi loss, not by different transcript amounts because all of the encoded genes were co-transcribed in the *cip-cel* gene cluster (Maamar *et al.*, 2006) and there was no statistically significant difference observed in the transcript amounts of *cipC*, *cel48F* and *cel9E* in the mutant. Additionally, the lack of the *dpi* gene also reduced the

amounts of transcripts from the *cip-cel* gene cluster in an unknown manner. This could be caused by reduced transcriptional activity and/or differential RNA stability. The lowered transcript levels did not significantly reduce cellulosome productivity. The poor correlation between RNA transcript and protein abundance has been reported in both prokaryotic and eukaryotic cells because of various biological factors (e.g. RNA abundance, RNA secondary structure, ribosome occupancy, codon bias, amino acid usage and protein half-lives) and methodological constraints (e.g. detection sensitivity and experimental error and noise) (Minczuk *et al.*, 2006). Interestingly, CipC abundance was similar in both the mutant and WT and several bands in the Ff fraction also showed similar abundance in both the mutant and WT. These results suggest that the proteolysis may be nonrandom and target specific proteins.

Even though glycoside hydrolases are important to cellulose saccharification, the contribution of each family to cellulolysis is still under active investigation. Families 48 and 9 are major cellulosomal components (Maamar *et al.*, 2004; Perret *et al.*, 2004; Blouzard *et al.*, 2010; Olson *et al.*, 2010). The disruption of the *cel48F* gene in *C. cellulolyticum* completely eliminated cell growth on cellulose. This defect is more severe than the report based on RNAi-mediated knockdown of *cel48F* expression (Perret *et al.*, 2004). The deletion of Cel48S from *C. thermocellum* resulted in a 40% decrease in cellular yield and 35% lower activity on Avicel cellulose (Olson *et al.*, 2010). However, the essential role of Cel48 in cellulolytic processes needs further evaluation because of the polar effect caused by intron insertion. A more severe polar effect was observed in a *cipC*Mut1 mutant that was created by IS insertion into the *cipC* gene in *C. cellulolyticum* (Maamar *et al.*, 2004). The polar effect in the *cipC*Mut1

mutant blocked the generation of 7.5 kb-long transcripts that were long enough to carry both *cipC* and the downstream *cel48F*. However, the transcript level of cel8C, which is immediately downstream in the cel48F mutant, remained at 42%. The disruption of the sole family 9 glycoside hydrolase in *C. phytofermentans* abolished cellulose degradation activity (Tolonen *et al.*, 2009). However, the *C. cellulolyticum cel9E* mutant only showed a 64.5% decrease in cellular biomass on cellulose, which was also a combinational effect of gene inactivation and polar effect. Taken together, these studies showed that the importance of Cel48 and Cel9 varied in cellulose-degrading *Clostridium* species. Interestingly, the loss of Cel48F and Cel9E reduced cell yield on cellobiose. A similar result was also observed due to Cel48S deletion in *C. thermocellum* (Olson *et al.*, 2010). Thus, Cel48F and Cel9E as key cellulosomal cellobiohydrolases could exert broader influences on cellular metabolism.

Combined with previous studies (Kang *et al.*, 2006; Schwarz *et al.*, 2006; Meguro *et al.*, 2011) and our findings, a conceptual model for Dpi-mediated regulation of cellulosome activity is proposed (Figure 4.7). In *C. cellulolyticum*, the *cip-cel* gene cluster and many other genes encode and secret diverse structural proteins and hydrolases. Cellulosomes are assembled on the cell surface with a diversity of components (e.g., glycoside hydrolases 48 family and 9 family, protease inhibitors) by dockerin-cohesin interaction (Carvalho *et al.*, 2003; Bayer *et al.*, 2004; Doi *et al.*, 2004; Blouzard *et al.*, 2010; Fontes *et al.*, 2010). Glycoside hydrolases on cellulosomes synergistically degrade diverse insoluble carbon substrates into soluble sugars to support cell growth. Under normal conditions, hydrolase activity is stablized by protease inhibitors such as Dpi in *C. cellulolyticum*, serpins in *C. thermocellulum* (Kang

et al., 2006), and cyspins in *C. cellulovrans* (Meguro *et al.*, 2011), which inhibit proteolysis, allowing the cells to continue to degrade availabe substrates at high efficiency. However, if the protease-inhibitor balance is disrupted by the introduction of exogenous proteases from plant biomass, some glycoside hydrolases (e.g., Cel48F and Cel9E in *C. cellulolyticum*) will be preferentially destroyed by the hyperactive proteolytic activity. Then the evacuated cohesin domains will be occupied by other available dockerin-containing components, leading to a dynamic change of cellulsomal composition.

The disturbance of the protease-inhibitor balance would down-regulate major transcripts of the *cip-cel* gene cluster, but not cellulosome productivity. The hyperactive proteolytic activity would also appear when protease and inhibitor genes are differentially expressed. It is possible that facing new carbon sources, cells may adjust protease-inhibitor expression leading to a proteolysis-dependent removal of the initially incroporated major cellulsomal components, thus allowing new substrate-induced enzymes to be assembled into the cellulosome. It should be noted that although supported by experimental data, further investigation is needed to substantiate this model.

In conclusion, this study uncovered the physiological role of a dockerin-containing protease inhibitor in protecting key cellulosomal cellulases from proteolysis, and identified the *in vivo* importance of two major cellulosomal components, Cel48F and Cel9E in crystalline cellulose degradation. This study suggests a mechanism by which cellulase stability may be enhanced via controlling protease/inhibitor activity or cellulase protein engineering to improve lignocellulose hydrolysis efficiency.

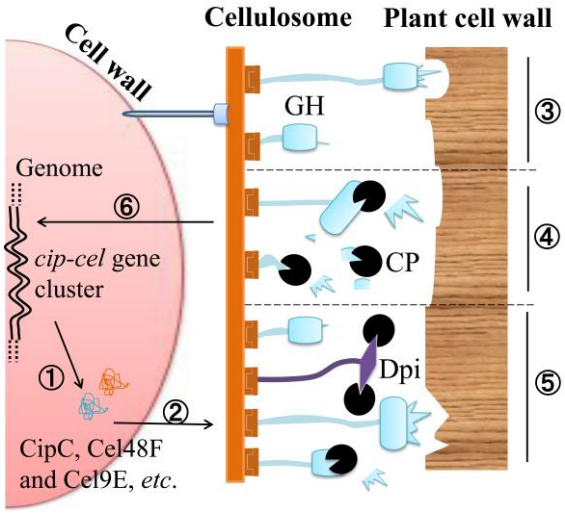


Figure 4.7 A conceptual model of Dpi-mediated regulation of cellulosomal activity in C. cellulolyticum. The *cip-cel* gene cluster expresses major cellulosomal components (e.g. CipC, Cel48F and Cel9E) (①) which assemble to form cell surface-bound cellulosomes (②) responsible for lignocellulose degradation (③). Bacteria-/plant biomass-derived cysteine proteases attack cellulosomal glycoside hydrolases (e.g. Cel48F and Cel9E) (④), thus reducing cellulolytic activity. Cellulosome-localized Dpi is able to block these proteases and protect cellulosomal components from proteolytic damage (⑤). Conversely, loss of the antagonistic balance due to differential expression or external protease attack will cause proteolysis of key cellulosomal components and simultaneously allow other dockerin-containing components to be incorporated, and also down-regulate the expression of the *cip-cel* gene cluster (⑥) with unknown mechanism. CP, cysteine proteases from plant biomass or bacteria; CipC, CipC scaffoldin; Dpi, dockerin-containing protease inhibitor of cysteine proteases; GH, glycoside hydrolases.

4.5 Materials and Methods

4.5.1 Bacterial strains and growth conditions

The bacterial strains and plasmids used in this study are listed in Table 4.2. *Escherichia coli* Top10 (Invitrogen) and Rosetta™ 2(DE3) pLysS strain (Novagen) were used for cloning and protein expression, respectively. *E. coli* transformants were grown at 37 ℃ in Luria-Bertani medium supplemented with kanamycin (50 μg ml^{-1}) and/or chloramphenicol (15 μg ml^{-1}) when required. *C. cellulolyticum* H10 and the

developed strains (including *dpi*, *cel48F*, *cel9E*, *dpi/over* and *dpi/zero*) were cultured anaerobically at 34 ℃ in modified VM medium supplemented with yeast extract (2.0 g l^{-1}) and various carbon sources (Higashide *et al.*, 2011). The medium was supplemented with erythromycin (15 μg ml^{-1}) or thiamphenicol (15 μg ml^{-1}) as appropriate. Colonies of each strain were isolated on solid VM medium containing 1% (weight/volume) agar and amended with cellobiose (5 g l^{-1}) and erythromycin (15 μg ml^{-1}) or thiamphenicol

Table 4.2 Bacterial strains and plasmids constructed to study the *dpi* gene.

Strain or plasmid	Relevant characteristics	Reference
Strains		
Escherichia coli Top10	F^- *mcr*A Δ(*mrr-hsd*RMS- *mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1	Invitrogen
E. coli Rosetta 2(DE3) pLysS	F^- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE2 (Cam^R)	Novagen
C. cellulolyticum H10	Wildtype	Petitdemange, *et al.*,1984
dpi	A group II intron was inserted into *dpi* ORF at 171 nt	This study
dpi/zero	*dpi* mutant background with pClostron3-*dpi/zero* plasmid	This study
dpi/over	*dpi* mutant background with pClostron3-*dpi/over* complementary plasmid.	This study
cel48F	A group II intron was inserted into *cel48F* ORF at 764 nt	This study
cel9E	A group II intron was inserted into *cel9E* ORF at 653 nt	This study
Plasmids		
pLyc1217Er	Kan^r in *E.coli*, Em^r in *Clostridium*, Fd promoter, pWH199 derivative	Li, *et al.*,2012
pClostron3	CMP^r in *E.coli*, TMP^r in *Clostridium*, Fd promoter, pJIR750a derivative	This study
pClostron3-*dpi/zero*	pClostron3 derivative with deletion of group II intron and LtrA	This study
pClostron3-*dpi/over*	pClostron3 derivative with *dpi* ORF driven by Fd promoter	This study
pET28a(+)-*dpi*	*dpi* coding sequence was ligated into NdeI-NotI-linearized pET28a(+)	This study

Abbreviations: *dpi*, *docterin-containing protease inhibitor* gene; Em^r, erythromycin resistant; Kan^r, kanamycin resistant; Fd, ferridoxin; LtrA, intron-encoded protein.

(15 μg ml^{-1}) as appropriate. For making cellulose-containing top-agar plates, sterile Avicel cellulose mixed with un-solidified VM agar was overlaid on the solidified VM agar. When required, thiamphenicol (15 μg ml^{-1}) was added. Cells (10 μl at OD_{600} =0.4) were dropped on the plates and then incubated at 34 ℃.

4.5.2 Plasmid construction and transformation of C. cellulolyticum H10

C. cellulolyticum mutants were generated by group II intron insertion. Before transformation, the intron region of *E. coli- C. cellulolyticum* shuttle vector pLyc1217Er was modified (Higashide *et al.*, 2011; Li *et al.*, 2012). Based on the online intron design tool (http://clostron.com/), we chose anti-sense integration sites at 171 bp, 764 bp and 653 bp downstream of the start codons of *dpi*, *cel48F* and *cel9E* genes, respectively and then synthesized four PCR primers (Table S4.2) for each intron modification, including IBS, EBS1d, EBS2 and EBSu. One-step crossover PCR using these four primers and pLyc1217Er as the template gave intron amplicon which was used to replace the original intron region after digestion with XmaI and BsrGI. The vectors were confirmed by sequencing and then were used for *C. cellulolyticum* transformation (Li *et al.*, 2012) , generating *dpi*, *cel48F* and *cel9E* mutants.

For the *dpi* mutant complementation, the plasmid pClostron3-*dpi/over* harboring the intact *dpi* ORF driven by a *C. pasteurianum* ferridoxin (Fd) promoter was constructed (Graves *et al.*, 1986). Using *C. cellulolyticum* genomic DNA as template, primers Dpi-overF and Dpi-overR were used to amplify the ORF (Table S4.2). PCR product was ligated into pClonstron3. The resulting plasmid confirmed by sequencing was then named pClostron3-*dpi/over*. The empty plasmid without any ORF downstream of the Fd promoter was named pClostron3-*dpi/zero* and was used as a negative control.

The *dpi* mutant transformed with pClostron3-*dpi/over* and pClostron3-*dpi/zero* generated *dpi/over* and *dpi/zero* strains, respectively. The plasmids were transferred to *C. cellulolyticum* by electroporation as previously described (Li *et al.*, 2012).

For expressing the recombinant Dpi in *E. coli*, the coding sequence of the *dpi* gene was cloned into the pET28a(+) vector (Novagen). The coding region was amplified by PCR using primers NtDpiF and NtDpiR (Table S4.2). An 879 bp amplicon digested with NdeI and NotI were cloned into NdeI-NotI-linearized pET28a(+), resulting in pET28a(+)-*dpi*. The final plasmid carries the *dpi* coding sequence fused in frame at its N terminus with a sequence encoding hexahistidine residues (His tag). The plasmid was transformed into Rosetta™ 2(DE3)pLysS competent cells to produce recombinant proteins according to the manufacturer's instruction.

4.5.3 Growth and cellulose degradation measurement

C. cellulolyticum growth on glucose (10 g l^{-1}) or cellobiose (10 g l^{-1}) was measured by monitoring OD_{600}. But on Avicel cellulose (10 g l^{-1}) or xylan (10 g l^{-1}), cell growth was determined by measuring bacterial protein content using the Pierce®BCA Protein Assay Kit (Thermo Scientific). Residual cellulose in cultures was estimated by using the phenol-sulfuric acid method, with glucose as the standard (Dubois *et al.*, 1956).

4.5.4 RNA isolation and quantitative real-time PCR

Total RNA was extracted from cellulose (10 g l^{-1}) -grown *C. cellulolyticum* cells at the mid-logarithmic phase by TRIzol® Reagent (Invitrogen). The RNA yield and integrity was determined with spectrophotometry and gel electrophoresis, respectively. And then reverse transcription was conducted by using SuperScript® III Reverse Transcriptase

(Invitrogen). cDNA products were diluted as appropriate and used as the templates. Quantitative real-time PCR was performed using iTaq SYBR Green Supermix with ROX (Bio-Rad) on a Bio-Rad iQ5 thermal cycler. Gene-specific primers used for transcript quantification are listed in Table S2. The thermal cycling conditions were as follows: 95 ℃ for 3 min, 40 cycles of 95 ℃ for 15 s, 55 ℃ for 15 s and 72 ℃ for 45 s. The *recA* gene was used as an internal calibrator (Stevenson *et al.*, 2005). Relative expression level was calculated with the Pfaffl Method (Pfaffl, 2001).

4.5.5 Fractionation of extracellular proteins

The *C. cellulolyticum* strains were grown on VM medium with cellulose (10 g l^{-1}). During mid-logarithmic phase the culture was filtered through a 3 μm-pore size GF/D glass fiber (Whatman). The penetration fluid was centrifuged to collect the supernatant containing the free extracellular protein fraction (Ff) and then concentrated with acetone precipitation. The cellulose retained on the filter was used to isolate bound proteins which mainly contain the cellulosome fraction (Fc) as previously described (Maamar *et al.*, 2004). Protein concentration was determined by the Pierce®BCA Protein Assay Kit (Thermo Scientific) according to the manufacturer's instruction.

4.5.6 Expression and purification of recombinant Dpi protein

To express recombinant Dpi protein, Rosetta™ 2(DE3) pLysS strain carrying pET28a(+)-*dpi* vector with an OD_{600} of 0.7 was induced with 0.2 mM IPTG at 25 ℃ for 15 h. The induced cells were harvested by centrifugation and then lysed by using CelLytic™ B 2× (Sigma) according to the manufacturer's protocol. The lysates were centrifuged and filtered with 0.2 μm filters (Sigma). The supernatant lysate was purified using a HisTrap HP 1 ml column (GE Healthcare) according to the manufacturer's

instruction. The eluate was fractionized during the washing step and the purity of each fraction was evaluated by SDS-PAGE. The fractions with pure recombinant protein were pooled, dialyzed and concentrated with an Amicon concentrator in 50 mM phosphate buffer (pH 7.2). Protein concentration was quantified using the Pierce®BCA Protein Assay Kit (Thermo Scientific). The recombinant protein was supplemented with 50% glycerol and then stored at -20 ℃ for further analysis.

4.5.7 SDS-PAGE analysis and MS identification

Protein samples from *E. coli* and *C. cellulolyticum* were subjected to SDS-PAGE using 10% resolving gels and mini electrophoresis units (Bio-Rad). Gels were stained with Coomassie blue. For densitometry analysis, decolored gels were scanned and analyzed with MYImage (Thermo Scientific).

To identify proteins in the gel, mass spectrometry was performed as follows. Protein bands excised from gel were subjected to in-gel trypsin digestion with reduction and alkylation as previously described (Wilm *et al.*, 1996). Then tryptic peptides were applied to HPLC and MS/MS analysis with the DionexUltiMate 3000 and ABI MDS SciexQstar Elite respectively. MS/MS data collected was submitted to in-house MASCOT (Matrix Science) server for protein identification against the NCBInr (02-2012) protein database.

4.5.8 Inhibitory activity test of the recombinant Dpi

The inhibitory activity of the recombinant Dpi was tested on commercial proteases including trypsin, chymotrypsin, papain and pepsin (All from Sigma) by using EnzChek® Protease Assay Kits (Invitrogen). Trypsin, chymotrypsin and pepsin were dissolved in 0.001 N HCl, making stock solutions (0.1 mg ml^{-1}). Papain was dissolved

in 50 mM sodium acetate (pH 4.5) with the concentration of 0.5 mg ml^{-1}. Following the manufacturer's instructions, the proteolytic reactions were performed in various working buffers, trypsin (20 μg ml^{-1}) and chymotrypsin (3.75 μg ml^{-1}) in 10 mM Tris-HCl (pH 7.8), pepsin (25 μg ml^{-1}) in 20 mM sodium acetate (pH 4) and papain (10 μg ml^{-1}) in 10 mM MES (pH 6.2), all of which were supplemented with the native or boiled recombinant Dpi (25 μg ml^{-1}) and substrate casein (5 μg ml^{-1}). All reactions were incubated at 37℃ for 1 hour before detecting the fluorescence using excitation and emission filters of 595 nm and 630 nm, respectively. Percentage of inhibition was calculated by dividing the difference in fluorescence values of reactions with the boiled Dpi from those with intact Dpi by the corresponding control reactions with the boiled Dpi, and then multiplying by 100.

The inhibitory capacity of Dpi towards papain activity was determined. Each reaction consists of casein substrate (5 μg ml^{-1}), papain (0.85 nM) and various concentrations of the native or boiled Dpi (0-5.54 nM) in 10 mM MES (pH 6.2). The boiled Dpi was used in control groups at each concentration. Before adding casein substrate, other components were mixed in advance and incubated at 37℃ for 15 min. Reactions were incubated at 37℃ for 1 hour and then the fluorescence was measured using excitation and emission filters of 595 nm and 630 nm, respectively. All assays were made in triplicate. IC50 was defined as the concentration of Dpi required for achieving 50% inhibition of papain.

Chapter 5 Carbohydrate Utilization in *Clostridium cellulolyticum* Differentially Relies on Catabolite Regulation System

5.1 Abstract

Clostridium cellulolyticum is a consolidated bioprocessing (CBP) bacterium, able to perform one-step bioconversion of lignocellulosic biomass into biofuels. However, carbohydrate utilization, irrespective of insoluble lignocellulose or its hydrolysates, needs to be improved to reduce the cost of biofuel production. Considering the importance of carbon catabolite regulation (CCR) in substrate utilization and many other biological processes in microorganisms but it is as yet undescribed in CBP-enabling bacteria, we investigated the CCR in *C. cellulolyticum* at the physiological, genetic, and transcriptomic levels. Our bioinformatic analysis indicated that this bacterium has a reduced CCR due to the absence of the sugar transporting phosphotransferase system in the genome, while the regulatory system of CCR is presumably retained and built up with the orthologs of a bifunctional HPr kinase/phosphorylase (HprK), a Crh protein, and LacI family members. This bacterium exhibited a very mild reverse CCR in light of the simultaneous assimilation of both hexoses and pentoses, and the promoting effect of glucose on the consumption of other sugars. Characterization of CCR component mutants revealed that both *hprK* and *crh* genes were tightly associated with the assimilation of monomer sugars, rather than cellobiose. Inactivation of either the *crh* gene or a LacI member regulator gene *ccpA* completely abolished cell growth on cellulose, which is the first genetic evidence showing the indispensability of Crh and CcpA in cellulose degradation. With microarray-based transcriptomic analysis of mutants that were cultivated on a soluble

sugar mix, the *crh* mutant exhibited a significant regulatory role in altering the expression of approximately 10% genes in the genome, some of which putatively encode transcriptional regulators, signal transduction components, and ATP-binding cassette transporters; transcriptional comparison between the single and double mutants of two additional LacI members, *lfpC2* and *lfpC3*, indicated that both functional specificity and redundancy occurred between these two genes; in contrast, the *ccpA* mutant just caused a minimal impact on physiological and transcriptional features which are totally distinct from its growth defect on insoluble cellulose. This study sheds very first light on the genetic importance of CCR-mediated regulation in cellulose degradation, provides systematic understanding of carbohydrate utilization in *C. cellulolyticum*, and also exploits potential candidate genes for catabolic engineering to improve lignocellulose bioconversion.

Keywords: carbon catabolite regulation, LacI transcriptional regulator, signal transduction, cellulose degradation, *Clostridium cellulolyticum*

5.2 Introduction

With an increasing concern over declining fossil fuels and worsening environmental issues (O'Neill *et al.*, 2002; Pacala *et al.*, 2004; Lewis *et al.*, 2006), bioconversion of lignocellulosic biomass to biofuels and other bioproducts are gaining significant prominence (Naik *et al.*, 2010). Lignocellulose is synthesized by photosynthetic CO_2 fixation in plants; accordingly alternative fuels made from lignocellulosic biomass are sustainable and carbon-neutral (Lynd *et al.*, 2002; Doi, 2008; Liao *et al.*, 2016). Hydrolysis of lignocellulose into soluble sugars (e.g., pentoses, hexoses, and oligosaccharides) and fermentation of resulting sugars into end products are two major steps during lignocellulose bioconversion (Lynd *et al.*, 2002). Making use of loaded substrates and fermentable hydrolysates will contribute to making the whole process more cost-effective (Balan, 2014). Although bacteria evolutionarily obtain metabolic versatility and flexibility with respect to substrate utilization (Petitdemange *et al.*, 1984; Goerke *et al.*, 2008), the flexibility usually causes diauxic cell growth and stepwise utilization of fermentable sugars if some are preferred (frequently glucose) (Moses *et al.*, 1966; Ng *et al.*, 1982; Singh *et al.*, 2008), which will result in a lower substrate utilization efficiency and a longer fermentation time and naturally bring economically unfavorable factors to the production process (Lynd *et al.*, 2002). Herein, it is necessary to understand the mechanisms of sugar uptake, the physiological responses and even systems-level regulatory mechanisms in candidate bacteria, especially when supplied with a mixture of fermentable sugars.

As one of the most important regulatory phenomena in bacteria, carbon catabolite regulation (CCR) enables bacteria to selectively use preferred carbon sources

by down-regulating the activities for using non-preferred or secondary substrates (Goerke *et al.,* 2008). Although different mechanisms (e.g., transcription control and translation control) of CCR are employed in bacteria to reach the same regulatory outcomes (Goerke *et al.*, 2008; Deutscher *et al.*, 2014), the phosphoenolpyruvate (PEP)-carbohydrate phosphotransferase system (PTS) is commonly required in signal transduction that can lead to CCR. Take the CCR of Firmicutes as an example. The PTS transports and phosphorylates sugars by using PEP as phosphoryl donor for the phosphorylation cascade formed by enzyme EI, histidine protein (HPr), and sugar-specific EII complex (Deutscher *et al.*, 2006). On the other hand, phosphorylation status of some PTS components, such as HPr and EII components, exerts regulatory functions by interacting with transcriptional regulators or non-PTS transporters (Deutscher *et al.*, 2006; Deutscher, 2008; Deutscher *et al.*, 2014). HPr plays a pivotal role in coordinating sugar uptake and transcriptional regulation of catabolic genes by changing its phosphorylation status (Deutscher *et al.,* 1995). There are two highly conserved amino acids (His15 and Ser46) in HPr that can be phosphorylated but with distinct mechanisms. P-Ser46-Hpr is catalyzed by a bifunctional HPr kinase/phosphorylase (HprK) (Deutscher *et al.*, 1983; Martin-Verstraete *et al.*, 1999; Mijakovic *et al.*, 2002), which is considered as a molecular sensor of intracellular glycolytic intermediates (Jault *et al.*, 2000; Mijakovic *et al.*, 2002); P~His15-HPr is generated during sequential phosphoryl transfer for sugar uptake (Postma *et al.*, 1993; Deutscher *et al.*, 1995). Therefore, cytoplasmic HPr exists in four forms (HPr, P~His-HPr, P-Ser-HPr and doubly phosphorylated HPr), by which carbon catabolism-associated processes are finely regulated (Deutscher *et al.*, 1995; Reizer *et al.*, 1996; Schumacher *et al.*, 2004;

Horstmann *et al.*, 2007). Reports showed that as many as 5-10% of all bacterial genes are in the control of CCR (Blencke *et al.*, 2003; Liu *et al.*, 2005). HPr-associated transcriptional regulators, such as catabolite control protein A (CcpA), tend to have pleiotropic functions, such as carbon utilization, nitrogen utilization, sporulation, pilus biogenesis, biofilm formation, toxin production and so forth (Ren *et al.*, 2012). Many of these biological processes are directly or indirectly associated with the overall fermentation performance (Ren *et al.*, 2012; Mitchell, 2016). There will be many possibilities to improve sugar conversion once we got better understanding of bacterial CCR.

CCR studies have been done in some sugar fermenting bacteria, such as *Escherichia coli* (Deutscher *et al.*, 1983; Schumacher *et al.*, 2004), *Bacillus subtilis* (Goerke *et al.*, 2008; Deutscher *et al.*, 2014), and *Clostridium acetobutylicum* (Grimmler *et al.*, 2010; Xiao *et al.*, 2011; Ren *et al.*, 2012). However, we know much less or even little about CCR in consolidated bioprocessing (CBP) bacteria, such as *Clostridium cellulolyticum* and *Clostridium thermocellum*. CBP-enabling bacteria are capable of directly using lignocellulose as carbon and energy source to produce end products such that they can accomplish the whole conversion with a single step, obviously superior to sugar fermenting bacteria which routinely rely on prerequisite lignocellulose hydrolysis by commercial enzymes or lignocellulose degrading microorganisms (Lynd *et al.*, 2002). *C. cellulolyticum,* as one of CBP representatives, is an anaerobic, mesophilic and cellulolytic model bacterium with industrial potential. It can grow on insoluble and soluble carbon sources (e.g., cellulose, xylan, cellobiose, glucose, xylose, and arabinose) (Petitdemange *et al.*, 1984; Li *et al.*, 2012; Xu *et al.*,

2013), among which cellobiose supports fast cell growth and is widely used in experiments. In comparison, *C. thermocellum* ferments cellulose and simple sugars, but not xylan (Ng *et al.*, 1981). However, no particular studies on CCR mechanisms have been conducted in CBP candidates, including *C. cellulolyticum*. According to previous reports, *C. cellulolyticum* CCR seems to have some interesting peculiarities. First, a low concentration of glucose induced instead of repressed the expression of the *cip-cel* operon which is indispensable to cellulose degradation (Xu *et al.*, 2013). Second, a catabolite-responsive element (*cre*) in the promoter of the *cip-cel* operon, which is the specified binding site of CcpA in many bacteria, played a negative role in the expression of a reporter gene in *C. cellulolyticum* (Abdou *et al.*, 2008). Third, no HPr orthologs were predicted in the genome, suggesting the native PTS could be problematic. Finally, despite the finding of CcpA homolog in the genome, two neighboring LacI member regulators, *lfpC2* and *lfpC3*, presented a strong negative correlation with the transcription level of the *cip-cel* operon (Xu *et al.*, 2013). With these clues, we speculate that *C. cellulyticum* CCR may have very distinct mechanisms responsible for the uptake of soluble sugars and the regulation of cellulose degrading genes.

As of the potential significance in fundamental molecular discoveries and application/engineering-oriented practices towards efficient lignocellulose bioconversion, we aimed to systematically understand the biological functions of CCR in *C. cellulolyticum*. To begin with, CCR genes were mined from the genome sequence and carbon catabolite repression was evaluated experimentally. Then, we created and characterized a Δ*hprK* knockdown mutant, five single and one double knockout mutants

of potential CCR components (i.e., Δ*crh*, Δ*ccpA*, Δ*lfpC2*, Δ*lfpC3*, Δ*lfpC2&3*) which were generated by the one-step Cas9 nickase-based genome editing tool. Finally, microarray-based transcriptomic analysis was performed to dissect gene functions. The outcomes of this study will help to engineer superior strains with optimum fermentation performance on a variety of complex feedstocks.

5.3 Results and discussion

5.3.1 In silico analysis of CCR components in C. cellulolyticum

Considering the important role of CCR in sugar utilization and many other biological processes, we did in silico genome mining of CCR components in *C. cellulolyticum*. Some of the following discoveries have been discussed previously (Abdou *et al.*, 2008). First, an HprK ortholog (Ccel_2293) was found, showing high identity with well-characterized HprK proteins and containing a conserved nucleotide-binding motif and a downstream signature sequence (Figure S5.1) (Galinier *et al.*, 1998; Hanson *et al.*, 2002). It is presumed to be functional in altering the phosphorylation status of HPr or Crh proteins.

Second, a catabolite repression HPr (Ccel_0806), which is an HPr paralogue named Crh in *B. subtilis* (Galinier *et al.*, 1997), was predicted in the *C. cellulolyticum* genome. Crh and HPr have different functions. In *B. subtilis* Crh plays a regulatory role like HPr does, by altering the phosphorylation status of its conserved Ser46 by HprK; however, it is not functional in sugar uptake due to the absence of the conserved His15 to form a high-energy phosphate bond (van den Bogaard *et al.*, 2000; Xiao *et al.*, 2011; Deutscher *et al.*, 2014). When looking into several cellulosome-producing bacteria (Figure S5.2), we found *Clostridium papyrosolves* also lacks any HPr orthologs but has

a sole Crh ortholog in the genome. Yet, other cellulose degraders like *C. thermocellum*, *Clostridium cellulovorans* and *Acetivibrio cellulolyticus* have HPr proteins instead of Crh.

Third, we did not find any orthologs of enzyme I (EI) and diverse sugar-specific enzyme II (EII), both of which are key components of PTS-mediated sugar uptake in many bacteria. Together with the lack of HPr, it is probable that *C. cellulolyticum* already lost the PTS to transport sugars. Instead, ATP-binding cassette (ABC) transporters may play an important role in sugar uptake. According to the TransportDB database (http://www.membranetransport.org), 62 genes in the *C. cellulolyticum* genome belong to the sugar-related ABC superfamily. Besides, a few genes are associated with the sugar-related major facilitator superfamily. These non-PTS sugar transport systems are supposed to endow this bacterium with the ability to grow on a diversity of carbon sources.

In terms of regulatory mechanisms, Crh phosphorylation in *B. subtilis* also mediates the binding of CcpA, one of LacI family transcriptional regulators, to tune the expression of downstream catabolic genes. *C. cellulolyticum* has five putative LacI genes, among which Ccel_1005 has the identity of 34% and similarity of 55% to that of *B. subtilis* CcpA (O'Neill *et al.*, 2002; Xu *et al.*, 2013). Seemingly, *C. cellulolyticum* contains all components responsible for the signal transduction of CCR. With all aforementioned features, we can question how sugars are transported in *C. cellulolyticum*, how sugar utilization is coordinated or regulated by CCR, and even what biological processes are under the control of CCR.

5.3.2 Evaluation of carbon catabolite repression in C. cellulolyticum

To evaluate CCR in *C. cellulolyticum*, we measured cell growth and substrate consumption on a single sugar and sugar mix. With a sole sugar as the carbon source, cells grew fastest on cellobiose, which was almost two-fold higher than on glucose or xylose (Table 5.1). This result is consistent with previous reports (Petitdemange *et al.*, 1984). In comparison with cell growth on cellobiose, dual sugars containing cellobiose and another simple sugar, glucose or xylose, supported a slower cell growth rate. More importantly, we did not observe a diauxic growth when any two of pentose, hexose and cellobiose were present simultaneously (i.e., cellobiose and glucose, cellobiose and xylose, glucose and xylose) (Figure S5.3). In terms of sugar consumption, the presence of a simple sugar (glucose or xylose) with cellobiose, or glucose with xylose, did not show phased sugar assimilation. Interestingly, the addition of glucose did not repress but dramatically promoted the consumption rate of cellobiose, which was increased from 1.88 mmol/g/h to 4.37 mmol/g/h (Table 5.1); however, the glucose consumption rate were oppositely decreased in the dual sugar mix (3 mmol/g/h) relative to in the sole glucose (4.56 mmol/g/h). With the addition of xylose, similar changes were observed that the cellobiose consumption rate was increased but xylose itself had a decreased consumption rate (Table 5.1). Similarly, the xylose consumption rate was also significantly increased by the presence of glucose, concomitant with a decrease in the consumption rate of glucose when compared to the counterpart of sole glucose as a carbon source. These results indicate that *C. cellulolyticum* can use various sugars simultaneously, irrespective of pentoses, hexoses, and simple cellodextrin, and it can assimilate cellobiose faster than glucose. In *C. thermocellum*, cellobiose was utilized in

preference to glucose (Ng *et al.*, 1982). These observations are distinct from reports on the strong glucose-induced CCR and diauxic shifts in *C. acetobutylicum* (Yu *et al.*, 2007), *Lactococcus lactis* (Solopova *et al.*, 2014), and *Escherichia coli* (Loomis *et al.*, 1967), all of which have a significant preference for glucose over other sugars. This type of CCR without preference for glucose is called reverse CCR, which has only been found in very few bacteria, such as *Bifidobacterium longum* (Parche et al., 2006), *Streptococcus thermophilus* (van den Bogaard et al., 2000), and *Pseudomonas aeruginosa* (Collier et al., 1996). In addition, a study found that a small amount of glucose even enhanced cellulose degradation in *C. cellulolyticum* (Xiao *et al.*, 2011). These aforementioned features, no diauxic growth and simultaneous assimilation of multiple sugars, bring merits to this stain as a wonderful mesophilic candidate to make the most out of nutrients during lignocellulose bioconversion.

Table 5.1 Characteristics of cell growth and substrate consumption with a sole or dual sugar(s).

Carbon source	Substrate[a]	μ_{max} (h^{-1})	Yx/s (g cells/g substrate)[b]	$q_{substrate}$ (mmol/g/h)
Sole	Cellobiose	0.145	0.23	1.88
	Glucose	0.065	0.08	4.56
	Xylose	0.071	0.10	4.82
Dual	Cellobiose	0.109	0.07	4.37
	Glucose		0.20	3.00
	Cellobiose	0.096	0.10	2.84
	Xylose		0.22	2.91
	Glucose	0.071	0.11	3.74
	Xylose		0.08	6.28

a, all values were determined using data at the mid-log phase with three biological replicates.

b, an optical density of 1 at 600 nm corresponded to 0.5 g (dry weight) per liter (Gehin *et al.*, 1996).

5.3.3 Physiological characterization of the hprK *knockdown mutant*

To dissect the role of CCR in sugar utilization, we conducted reverse genetic studies of three possible CCR genes, *hprK*, *crh* and *ccpA*, and two additional LacI member regulators, *lfpC2* and *lfpC3*. Mutants were generated and characterized as below.

HprK as the sensor of glycolytic intermediates passes down the signal by phosphorylating Ser46 of HPr or Crh (Jault *et al.*, 2000; Mijakovic *et al.*, 2002). The changed phosphorylation status affects the binding affinity of HPr or Crh to transcriptional regulators and sequentially the binding of regulators on the promoters of catabolic genes. Rationally, knockout of this sensor would affect the regulatory role of CCR in sugar catabolism. Initially, the ClosTron method (Li *et al.*, 2012) and our newly developed Cas9 nickase-based genome editing (Xu *et al.*, 2015), were sequentially applied to disrupt the gene; however, we failed to get any Δ*hprK* knockout mutants, indicating the essentiality of this gene under the test condition. To circumvent the difficulty, we switched to antisense RNA-mediated interference (RNAi) to knockdown instead of knockout the *hprK* gene. To do so, the partial gene sequence with the length of 150 bp starting from the predicted transcriptional start site was inserted downstream of a strong ferredoxin promoter in an inverted orientation such that the opposite strand will be transcribed to produce RNA molecules that are completely complementary with *hprK* transcripts. The resulting RNA duplexes will trigger RNA degradation or block translation to lower down gene expression (Thomason *et al.*, 2010). After plasmid construction, electroporation and colony screening, a verified pRNAi-hprK strain was obtained and then subject to growth profiling and measurements of both sugar consumption and fermentation products on glucose, cellobiose, sugar mix or cellulose

as a carbon source. Interestingly, relative to the empty pRNAi control, the pRNAi-hprK strain dramatically reduced cell growth on glucose (Figure 5.1A) and slightly changed the yield of cell biomass on a sugar mix and cellulose (Figure 5.1B and C); however, its growth on cellobiose was not significantly affected (Figure 5.1D). The weak growth on glucose may indicate the defectiveness of glucose assimilation in pRNAi-hprK. To examine it, sugar consumption was monitored during cell growth on a sugar mix containing cellobiose, D-glucose, D-xylose and L-arabinose. The results showed that (Figure 5.2): 1) the RNAi-hprK strain surely lost the ability to consume glucose since the glucose concentration was not changed during its visible growth (Figure 5.1B); 2) simultaneously, the mutant dramatically decreased the consumption rate of xylose and arabinose when compared with the pRNAi strain; 3) cellobiose utilization was not affected in the knockdown mutant; 4) the control pRNAi strain presented the simultaneous assimilation of diverse sugars, including pentoses and hexoses, which is consistent with the results as described above; and arabinose was consumed at the lowest rate. Obviously, these results indicate that the *hprK* gene is very critical to glucose assimilation, also important to the utilization of both xylose and arabinose, but almost dispensable to cellobiose metabolism.

Apart from the decreased cell biomass on cellulose, the pRNAi-hprK knockdown strain hydrolyzed more cellulose and released 3.16-fold higher amount of soluble reducing sugar in the fermentation broth (Figure S5.4A and S5.4B). It seems like *hprK* repression enhanced cellulose hydrolysis under the test condition. Metabolically, the knockdown strain reduced acetate titer but reversely increased ethanol production irrespective of carbon sources, such as cellulose and cellobiose

(Figure S5.4C and S5.4D). These results indicate that the *hprK* gene or its associated signal transduction is somehow linked with cellular metabolisms.

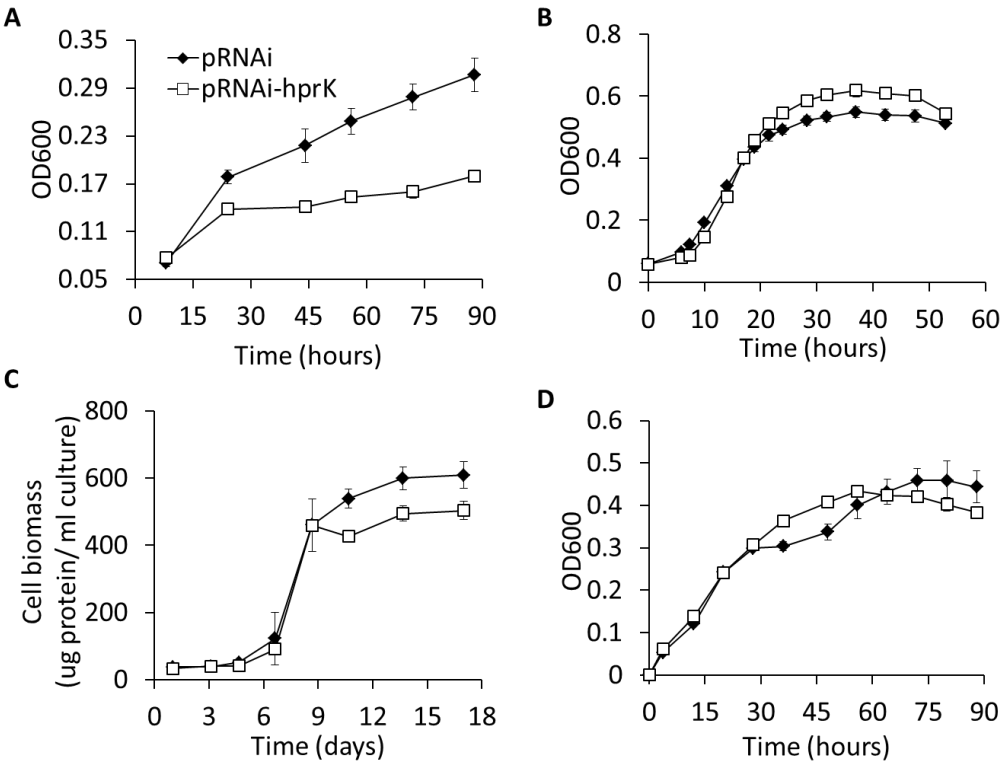


Figure 5.1 Growth profiling of pRNAi and pRNAi-hprK strains. All tests were performed in the defined VM medium with corresponding substrates: 10 g/L D-glucose (A); a sugar mix consisting of 3 g/L cellobiose, 2.5 g/L D-glucose, 2 g/L D-xylose, and 1 g/L L-arabinose (B); 15 g/L Avicel cellulose (C); 15 g/L cellobiose (D). The mean and standard deviation are shown for three biological replicates at each time point.

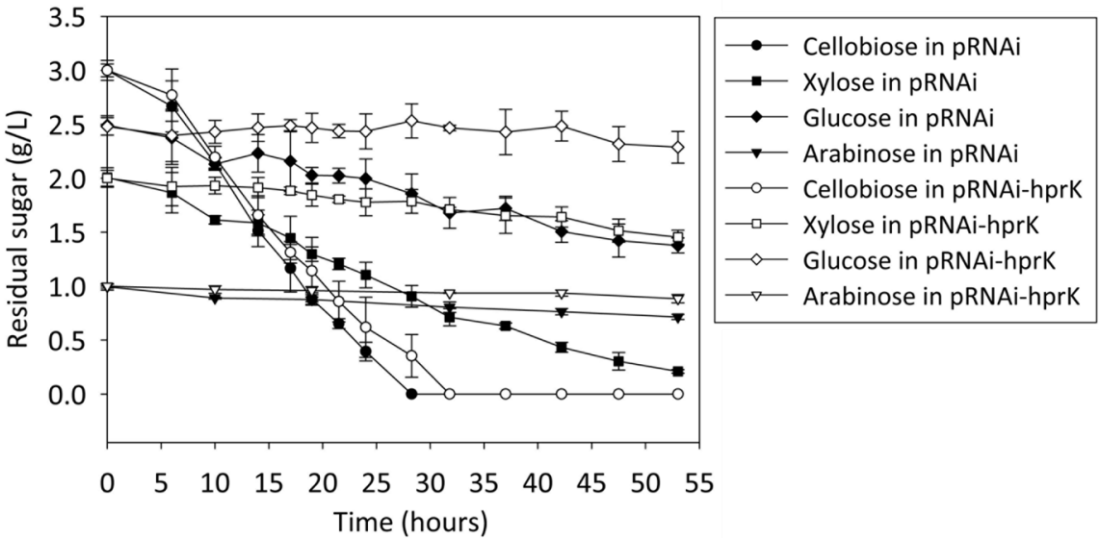


Figure 5.2 Measurements of residual sugars during cell growth. Both pRNAi and pRNAi-hprK strains were cultivated in the defined VM medium with a sugar mix containing 3 g/L cellobiose, 2.5 g/L D-glucose, 2 g/L D-xylose, and 1 g/L L-arabinose. Residual sugars were measured with high performance liquid chromatography. The mean and standard deviation are shown for three biological replicates at each time point.

5.3.4 Mutagenesis and characterization of the crh *gene and three LacI regulator genes*

LacI family transcriptional regulators consist of helix-turn-helix DNA-binding domains and ligand-binding domains (Ravcheev *et al.*, 2014). During evolution, LacI members have diverged molecular determinants of the DNA and ligand specificity, leading to varied functions (Ravcheev *et al.*, 2014). *C. cellulolyticum* has five putative LacI family members, among which CcpA ortholog was predicted. The putative Crh protein, once phosphorylated at Ser46 by HprK, may play a similar regulatory role as reported in *B. subtilis* (Schumacher *et al.*, 2006), serving as a cofactor of CcpA regulator to tune gene expression. It is interesting that two other LacI members *lfpC2* and *lfpC3* presented a strong negative correlation with the transcription of the *cip-cel* operon (Xu *et al.*, 2013). However, little is known about the roles of these regulators in cellulose degradation and sugar assimilation in this bacterium. To interrogate their functional specificities and

even molecular mechanisms, we generated knockout mutants of the *crh* gene and three LacI members, including *ccpA, lfpC2,* and *lfpC3*. To inactivate these genes and minimize polar effects on neighboring genes, we applied the Cas9 nickase genome editing tool to precisely insert a customized 7 bp DNA fragment harboring a restriction enzyme site into the targeted gene, presumably resulting in frameshift mutations. By doing so, corresponding single mutants (Δ*crh*, Δ*ccpA*, Δ*lfpC2*, Δ*lfpC3*) and one double mutant (Δ*lfpC2&3*) were generated and verified by PCR identification, amplicon digestion of inserted restriction enzymes and amplicon sequencing (Figure S5.5). It is worth noting that the double mutant was generated with a single step by transforming an all-in-one vector harboring two customized gRNAs and two homologous regions to target both *lfpC2* and *lfpC3* genes simultaneously. Then, we characterized these mutants by monitoring cell growth on different carbon sources and measuring substrate utilization.

The Δ*crh* mutant grew much slower than the control and other mutants on agar plates irrespective of glucose, xylose or arabinose used as the sole carbon source; however, its growth on cellobiose was similar to all other strains (Figure 5.3), which was further verified in the cellobiose liquid medium in terms of cell growth rate and maximal cell yield (Figure S5.6). Such distinct sugar-dependent effects of *crh* inactivation on cell growth suggested that Crh is associated with the metabolism of simple sugars (i.e., glucose, xylose, and arabinose) but cellobiose. To exclude the effect of one sugar on the utilization of others, we cultivated these strains with a sugar mix, consisting of cellobiose, glucose, xylose, and arabinose, and then compared the consumption of each sugar (Figure 5.4). Basically, all mutants used up cellobiose

quickly. The Δ*crh* mutant almost halved the consumption of glucose, xylose, and arabinose; strikingly, among all mutants tested here, it was the only one with a decreased utilization of glucose. Together, the Δ*crh* mutant presented the very similar changes in sugar assimilation as were observed in the *hprK* knockdown mutant. It is not surprised because Crh is a mediator responsible for transmitting the signal perceived by HprK (Jault *et al.*, 2000; Mijakovic *et al.*, 2002; Schumacher *et al.*, 2006). These results also indicated that cellobiose assimilation is seemingly independent of HPrK/Crh-dependent CCR.

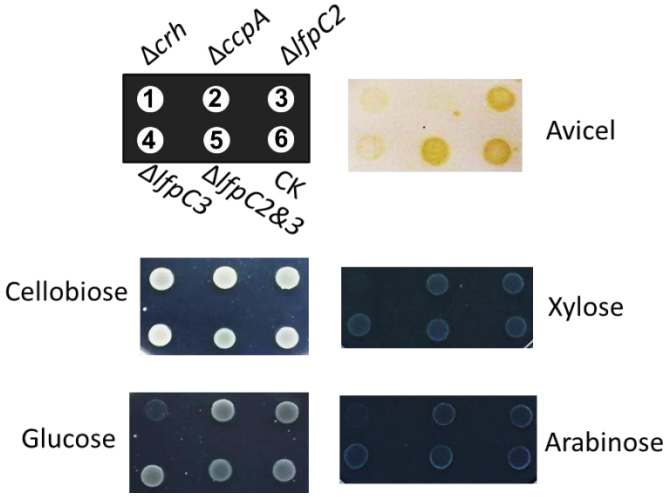


Figure 5.3 Effect of mutagenesis on colony development. The same amount of cells was inoculated on the defined solid VM medium with different carbon sources. The relative position of each strain on the plates is indicated in the scheme panel. Each test was run with multiple replicates. Sugar plates and Avicel cellulose-topping plates were imaged after 7 days and 14 days, respectively.

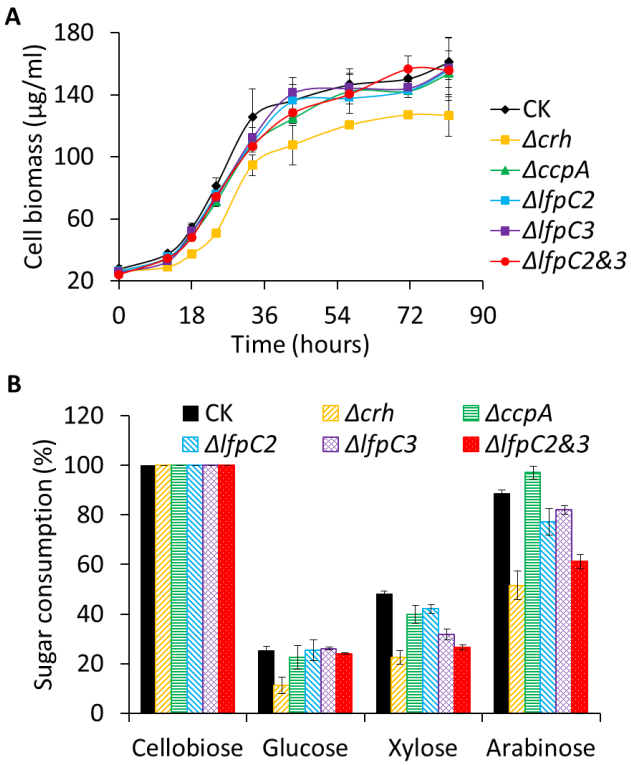


Figure 5.4 Cell growth and sugar utilization in the defined medium with a sugar mix as carbon source. The sugar mixture consisted of 3 g/L cellobiose, 2.5 g/L D-glucose, 2 g/L D-xylose, 1 g/L L-arabinose. Cell growth was indicated by cell biomass at each time point (A). In the endpoint fermentation broth, the consumption percentage of each sugar was calculated and presented (B).

Among the mutants of LacI member regulators, the Δ*ccpA* mutant slightly reduced xylose utilization but increased arabinose utilization a little bit (Figure 5.4). Both Δ*lfpC2* and Δ*lfpC3* mutants reduced the consumption of xylose and arabinose, and the impact was much more obvious in the dual mutant, Δ*lfpC2&3*. As above noted, none of these three LacI regulators were significantly related to glucose metabolism but the *crh* gene did. It is possible that Crh-dependent CCR employed other transcriptional regulators or regulator-independent regulatory ways to promote glucose utilization. The three LacI regulators appeared to be associated with the assimilation of both xylose and arabinose probably via their involvement in the transcriptional control of sugar transport

genes and catabolic genes. In the *C. cellulolyticum* genome, *lfpC2* and *lfpC3* genes are physically located in a bicistronic operon, sharing high identity and similarity with each other. The accumulative effect in the double mutant indicates the possibility of direct or indirect co-regulation of these two regulators in xylose and arabinose utilization.

Cellulose as an insoluble carbon source has to be enzymatically hydrolyzed into soluble sugars prior to sugar uptake and metabolism in bacteria. The cellulolytic activity of bacterial cells can be intuitively compared on cellulose agar plates by observing transparent halos formed during cell growth. Interestingly, distinct from the obvious growth of Δ*lfpC2* and Δ*lfpC2&3* mutants like the control strain did, other mutants (i.e., Δ*crh*, Δ*ccpA*, and Δ*lfpC3*) presented negligible cell growth (Figure 5.3). Similar changes were observed in the liquid medium with Avicel cellulose that both Δ*crh* and Δ*ccpA* mutants barely degraded cellulose and the Δ*lfpC3* mutant reduced cellulose degradation relative to the remaining strains (Figure S5.7). It is clear that both *crh* and *ccpA* genes are essential to cellulolysis and cell growth on cellulose. Between *lfpC2* and *lfpC3* genes, *lfpC3* is more likely to be positive in cellulose degradation; however, *lfpC2* conditionally affects cellulose degradation only when *lfpC3* is dysfunctional. This assumption is supported by our observation that the weakened cellulolysis in Δ*lfpC3* was restored by introducing an additional mutation in *lfpC2*. Maybe *lfpC2* negatively intervenes an alternative way to influence cellulolysis in parallel. These defective phenotypes were restored in our complementation tests.

5.3.5 Link differentially expressed genes to biological processes

To explore an overall picture of biological functions of these genes, we conducted transcriptomic analysis for all knockout mutants with gene expression microarrays.

Considering the complexity of lignocellulose and inducible effects of some sugars occurring at the transcriptional level in bacteria, we cultivated all strains in the defined medium with a mixture of cellobiose, glucose, xylose, and arabinose to mimic sugar diversity, and collected cells at three time points during the exponential phase, one transitional point at the early stationary phase and three time points during the stationary phase. Time-course sampling provides greater flexibility to detect phase-dependent gene expression/regulations and analyze holistic correlations of genes of interest by constructing co-expression networks in the future. After total RNA extraction and DNA labeling, microarray hybridization was conducted on Agilent 8-array slides, which contains 13, 098 probes with 50 nt in length able to interrogate the transcript levels of 94% protein encoding genes in *C. cellulolyticum*. Differentially expressed genes (DEGs) in each mutant relative to the control, including both up-regulated and down-regulated genes, are identified as genes with a log2 fold-change above 1 (or below -1) and an adjusted p value below 0.05. To validate microarray results, quantitative PCR (qPCR) was applied to examine the fold changes of 10 selected DEGs with a wide range of up-regulated or down-regulated expression. Regression analysis of their log2 fold changes obtained with microarray and qPCR provided an R2 value of 0.94 (Figure S5.8), indicating that our microarray analysis gave an accurate report of transcript levels in this study.

Considering phase-dependent gene expression, we determined DEGs in each mutant during the exponential and stationary phase separately. In general, during the exponential phase Δ*crh* possessed the largest number of DEGs (133 genes), followed by the dual mutant Δ*lfpC2&3* (70 genes), Δ*lfpC3* (49 genes), Δ*lfpC2* (38 genes), and least

in Δ*ccpA* (7 genes). The number of up-regulated and down-regulated genes was summarized in Table 5.2. In total, 231 genes were differentially expressed between mutants and the control, among which 148 were annotated genes. Hierarchical clustering analysis of these 231 DEGs showed that many DEGs were only highly influenced in the Δ*crh* mutant as indicated in the group II and III; the remaining DEGs in the group I and IV varied in transcriptional levels amongst these mutants (Figure 5.5A). The Venn diagram indicated the portion of shared and unique DEGs (Figure 5.5B). To explore the functional relevance of DEGs in each mutant, gene set enrichment was conducted to identify enriched (or depleted) gene ontology (GO) terms in the lists of DEGs. All GO terms, corresponding DEGs, gene annotations, and fold changes in transcriptional levels were all listed in Table S5.3.

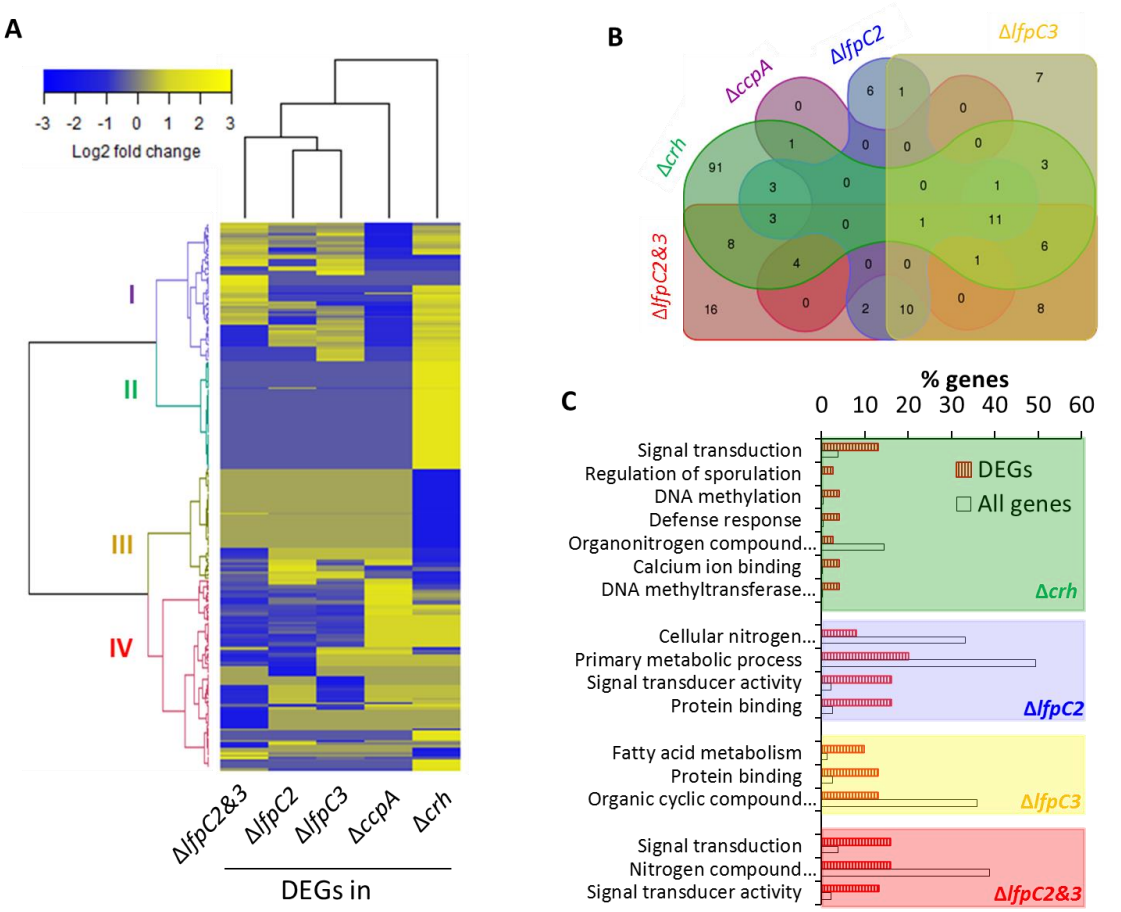


Figure 5.5 Microarray-based transcriptomic analyses in mutants at the exponential phase. (A) Hierarchical clustering analysis of 231 differentially expressed genes (DEGs) that present in at least one of the five mutants (Δ*crh*, Δ*ccpA*, Δ*lfpC2*, Δ*lfpC3*, and Δ*lfpC2&3*). Up-regulated and down-regulated genes are indicated by yellow and blue, respectively; and brightness indicates the magnitude of log2 fold changes. Expression patterns at the gene level were clustered into four colorized branches (I-IV). (B) Venn diagram analysis of DEGs in each mutant. It separates shared and common DEGs among mutants. (C) Gene ontology (GO) enrichment-based functional profiling of the DEGs in mutants. All GO terms listed here were specifically enriched or depleted in the DEGs of indicated mutants relative to all genes in *C. cellulolyticum* ($p<0.01$, two sided Fisher's exact test). The GO number of each term was listed in Table S5.3.

133 DEGs in the Δ*crh* mutant had significant enrichments for regulation of sporulation, signal transduction, defense response, Calcium ion binding, DNA methylation and DNA-methyltransferase activity, and organonitrogen compound biosynthetic process (Figure 5.5C). More specifically, eight DEGs putatively encode components of two-component systems (TCS), including histidine kinases and response regulators, the majority of which presented (6/8) up-regulated expression (Table S5.5). Since TCS components are usually but not exclusively involved in regulating neighboring genes, in-depth analysis found that these influenced TCSs may be associated with pilus biogenesis (Ccel_0502-0503), hemicellulose degradation and/or resulting hydrolysates transport (Ccel_1223 and Ccel_1227), sporulation (Ccel_1894-1895), xylose ATP-binding cassette (ABC) transporters (Ccel_1982-1987), citrate/malate metabolism (Ccel_2269-2270). There are a few of DEGs associated with ABC transporters presumably responsible for amino acid transport (Ccel_1631 and Ccel_2587), xylose uptake (Ccel_1223, Ccel_1987, and Ccel_2686) (Table S5.5). Transcriptional changes in these sugar-related TCS mechanisms and sugar transporters add explanations to the decreased assimilation of monomer sugars as observed above (Figure 5.3 and 5.4B). In addition, there are six DEGs putatively encoding transcriptional regulators of a diversity of families such as AbrB, MarR, AraC, XRE, and LacI. It means that Crh is important for signal transduction from the HprK sensor to diverse signal receivers. Venn diagrams (Figure 5.5B) showed the Δ*crh* mutant had 91 unique DEGs that were not shared with any other mutants at the exponential phase. All these results together indicate that the *crh* gene has diverse functions in this bacterium, and also suggest that Crh-mediated signal transduction may rely on other transcriptional

regulators or unknown mechanisms, slightly on the three interrogated regulators under the test condition.

Table 5.2 The number of differentially expressed genes in sugar mixture-grown mutants at the exponential and stationary phases.

Mutant	Exponential			Stationary		
	Up-regulated	Down-regulated	Total	Up-regulated	Down-regulated	Total
Δ*crh*	81	52	133	195	69	264
Δ*ccpA*	1	6	7	0	0	0
Δ*lfpC2*	17	21	38	1	3	4
Δ*lfpC3*	20	29	49	1	0	1
Δ*lfpC2&3*	30	40	70	39	38	77

Among these regulator mutants, the Δ*ccpA* mutant presented a very few DEGs with no enriched GO terms, which indicates a weak or narrow functionality of the CcpA regulator under the test condition. In the Δ*lfpC2* mutant (Figure. 5.5C), its DEGs had four GO terms enriched, including signal transducer activity, protein binding, cellular nitrogen compound metabolic process, and primary metabolic process. In comparison, the Δ*lfpC3* mutant had very different enrichments (Figure 5.5C), such as fatty acid metabolic process organic cyclic compound binding, even though it additionally shared the protein binding term with the Δ*lfpC2* mutant. The double mutant Δ*lfpC2&3* had three enriched GO terms (i.e., signal transduction, nitrogen compound metabolism, and signal transducer activity) all of which presented in the Δ*lfpC2* mutant (Figure 5.5C), but for each GO term, there were more DEGs present in the double mutant. As to the composition of DEGs, the Δ*lfpC2*, Δ*lfpC3* and Δ*lfpC2&3* mutants shared a lot of DEGs (Figure S5.9A), and simultaneously the Δ*lfpC2&3* mutant had the largest number of unique DEGs (28 out of its 70 DEGs). It seems like that the functional redundancy and

specificity occur between *lfpC2* and *lfpC3* genes. This assumption is also supported by the observations that 1) the Δ*lfpC2&3* double mutant had one more TCS gene (Ccel_0944) in the DEG set (Table S5.5), in addition to two others (Ccel_0994 and Ccel_2125) that were shared with both Δ*lfpC2* and Δ*lfpC3*; 2) although a diversity of ABC transporters was affected in both single and double mutants, only the Δ*lfpC2* mutant and the double mutant exhibited changes in sugar-related ABC transporters (Ccel_0145 and Ccel_2686 in Δ*lfpC2*; Ccel_1987 in the double mutant). These observations may explain the decreased utilization of xylose and arabinose in the double mutant. Furthermore, all of these four regulator mutants (i.e., Δ*ccpA,* Δ*lfpC2*, Δ*lfpC3,* and Δ*lfpC2&3*) shared the same DEG with a down-regulated expression, which is Ccel_3075, putatively encoding a Xenobiotic Response Element (XRE) family transcriptional regulator. XRE family regulators have diverse functions, which are related to DNA methylation, cell development, antitoxin, and prophage repression (RegPrecise database).

At the stationary phase, all mutants exhibited dramatic changes in terms of DEGs and enriched GOs (Figure 5.6). First, the single mutants (Δ*ccpA*, Δ*lfpC2,* and Δ*lfpC3*) showed very few or even none DEGs (Table 5.2). Second, the Δ*crh* mutant increased DEGs to 264 genes accounting for 7% of genes in this bacterium, the majority of which were up-regulated (195/264) and strikingly phase-dependent (194/264) since 73% DEGs were exclusively influenced at the stationary phase (Figure S5.9B). Also, a lot of genes were specifically up-regulated in this mutant as indicated in the branch IV (Figure 5.6A). All of its DEGs were enriched with response to stimulus, cobalamin biosynthetic process, signal transducer activity, heme biosynthetic process, cell

communication, and cellular component organization (Figure 5.6C). The vast impacts at the stationary phase in the Δ*crh* mutant indicate the important role of Crh in regulation. The DEGs and associated biological processes could help cells to cope with or respond to unfavorable conditions, such as insufficient nutrients and stress surroundings that usually occur at the stationary phase of cell growth. Third, the Δ*lfpC2&3* mutant had 77 DEGs with enrichments in protoporphyrinogen IX biosynthetic process, chlorophyll metabolic process, binding, and precorrin-2 dehydrogenase activity, all of which are so different from those at the exponential phase (Figure S5.9C). The precorrin-2 dehydrogenase belonging to the family of oxidoreductases participates in porphyrin and chlorophyll metabolism. In addition, Venn diagram analysis showed that at this stage (Figure 5.6B), 86% of DEGs in the Δ*crh* mutant (227/264) were exclusively influenced; approximately half of DEGs in Δ*lfpC2&3* were uniquely co-regulated by *lfpC2* and *lfpC3*. Interestingly, GO enrichments showed that both Δ*crh* and Δ*lfpC2&3* mutants influenced the metabolism of isocyclic compounds (e.g., porphyrin and chlorophyll). These results suggest that LfpC2 and LfpC3 are redundantly involved in the signal transduction from the upstream Crh.

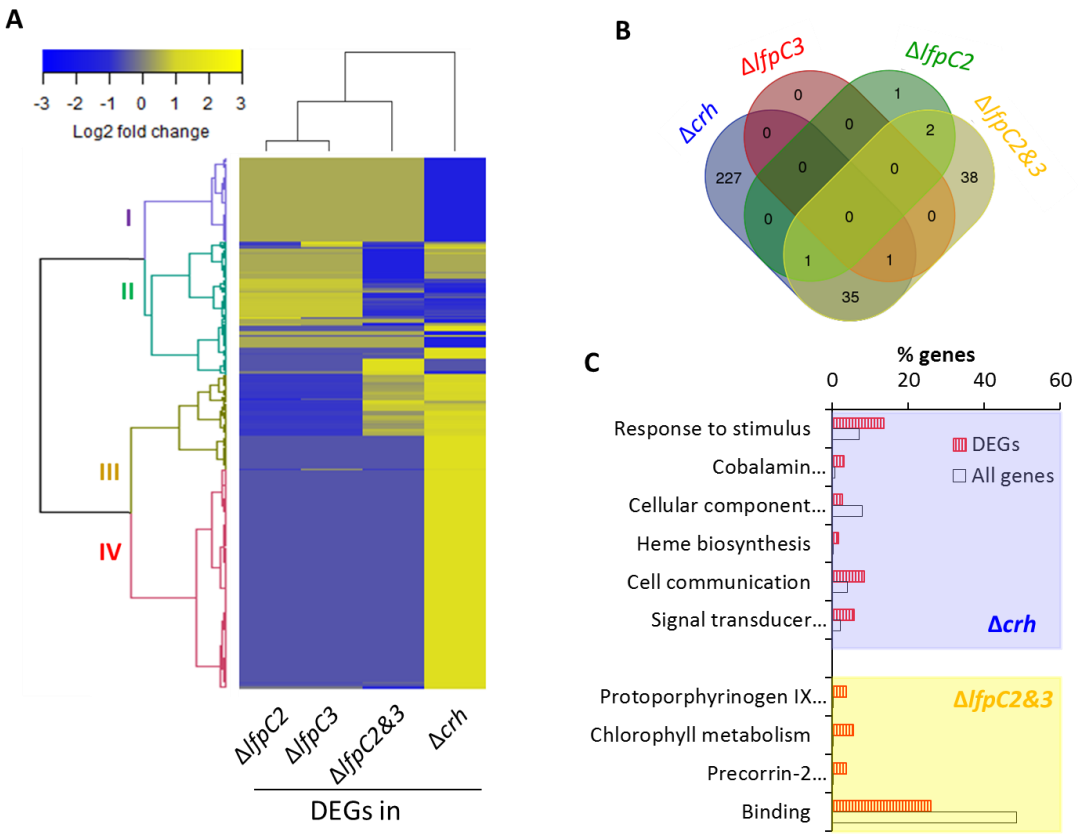


Figure 5.6 Microarray-based transcriptomic analyses in mutants at the stationary phase. (A) Hierarchical clustering analysis of 375 differentially expressed genes (DEGs) that present at least one of the four mutants (Δ*crh*, Δ*lfpC2*, Δ*lfpC3*, and Δ*lfpC2&3*). Up-regulated and down-regulated genes are indicated by yellow and blue, respectively; and brightness indicates the magnitude of log2 fold changes. Expression patterns at the gene level were clustered into four colorized branches (I-IV). (B) Venn diagram analysis of DEGs in each mutant. It separates shared and common DEGs among mutants. (C) Gene ontology (GO) enrichment-based functional profiling of the DEGs in mutants. All GO terms listed here were specifically enriched or depleted in the DEGs of indicated mutants relative to all genes in *C. cellulolyticum* ($p<0.01$, two sided Fisher's exact test). The GO number of each term was listed in Table S5.4.

In general, our microarray data showed a broad and important role of Crh in regulating diverse genes, more prominently involved in signal transduction. Loss of the Crh interactor, CcpA, did not exhibit noticeable changes at least in terms of its impact on gene expression and cell growth on soluble sugars tested above. It is possible that CcpA has been degenerated or specialized for other particular uses. The other two LacI regulators, LfpC2 and LfpC3, exhibited functional redundancy and specificity as manifested by DEG analysis and aforementioned sugar assimilation, whereas LfpC2 is a more likely downstream signal receiver of Crh, in particular for signal transduction and nitrogen-related metabolism.

By combining our physiological and transcriptomic data with previous reports in *C. cellulolyticum*, we proposed a rudimentary model for carbon catabolite regulation in this bacterium (Figure 5.7). Different from the well-studied CCR in *B. subtilis* (Deutscher *et al.*, 2006; Deutscher, 2008), this bacterium does not have the sugar-transporting PTS system but has functional regulatory components of CCR, including HprK, Crh and CcpA. When usable sugars are limited, the cells are normally at the status of insufficient energy or intermediates that are mainly produced from glycolysis. At this time, HprK acts as a phosphorylase to remove the phosphate group from the P-Ser-Crh. The unphosphorylated Ser-Crh could be a high-affinity cofactor of CcpA and then the regulatory complex formed by CcpA and Ser-Crh positively regulates the transcription of genes encoding carbohydrate-active enzymes (CAZyme). It is probable that this complex directly binds to the *cre* operator of the *cip-cel* operon and then tunes the expression of key cellulosomal genes (Abdou *et al.*, 2008), which build up cellulosomes on the cell surface in an unknown way. Once lignocellulose degrading

enzymes are produced, the extracellular lignocellulose will be efficiently hydrolyzed into soluble sugars, including oligosaccharides, cellobiose, glucose, xylose and arabinose. When more sugars available, the cells will assimilate sugars to product more energy and glycolytic intermediates that will trigger the kinase activity of HprK to phosphorylate Ser-Crh using ATP. The resulting P-Ser-Crh will mediate the binding of transcriptional regulators (TFs), such as LfpC2, to directly or indirectly tune the expression of sugar-specific two-component systems (TCS) responsible for the uptake of xylose and arabinose, possibly glucose. The sensor kinase of TCS, once sensing xylose/arabinose with the help of extracellular sugar/solute-binding proteins (SBP), will activate its response regulator (RR) by phosphorylation and then enhance the expression of xylose/arabinose-specific ATP-binding cassette (ABC) transporters. According to studies on HPr proteins (Deutscher *et al.*, 2006; Deutscher, 2008), it is possible that the P-Ser-Crh may directly interact with ABC transporters to differentiate the uptake efficiency of different sugars. Since we did not observe any significant impacts of CCR mutants on cellobiose utilization, it seems like some undescribed regulatory mechanisms are involved in the assimilation. Current studies in *C. cellulolyticum* cannot exclude the involvement of other sugar-specific/-nonspecific secondary transporters (ST) in sugar uptake (Saier, 2000). Catabolically, pentoses, such as xylose and arabinose, once transported into the cells, need to be converted in the pentose phosphate pathway (PPP) before entering into glycolysis and downstream metabolism. With these proposals, much more effort is required to decipher the exact mechanisms at the molecular level.

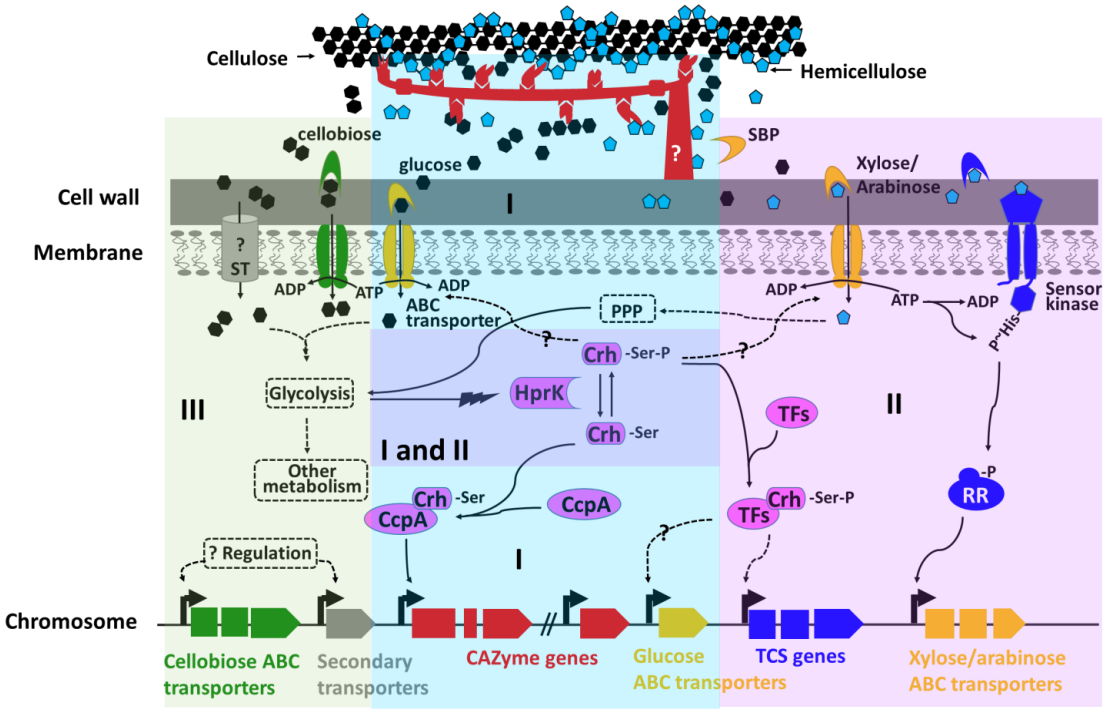


Figure 5.7 A model of carbon catabolite regulation in *C. cellulolyticum*. The regulatory system of CCR (shaded area I and II) senses the intracellular energy status, regulates lignocellulose degradation (shaded area I) by controlling key carbohydrate degrading enzymes (CAZyme), and conditionally activates two-component systems (TCS) to sense available sugars by sugar-/solute-binding proteins (SBP) and sensor kinases prior to activating downstream response regulators (RR) and sugar-specific ATP-binding cassettes (ABC) transporters (shaded area II). Cellobiose catabolism and possible secondary transporters (ST) for sugar uptake may be regulated in an unknown way (shaded area III). Pentoses, such as xylose and arabinose, are normally processed by the pentose phosphate pathway (PPP) before catabolized through glycolysis. Gene colors on the chromosome correspond to the colors of their protein products on/around the membrane. Dashed curves and question marks indicate multi-step processes and undetermined relationships, respectively.

5.4 Conclusion

In light of the importance of carbon utilization during lignocellulose bioconversion, *C. cellulolyticum* CCR was systematically studied at the physiological, genetic and transcriptomic levels. Our results indicate that this bacterium has a very mild reverse CCR as manifested by the simultaneous utilization of multiple sugars and the promoting effect of glucose on the consumption of other sugars. CCR components, HprK and Crh,

were genetically proven to be tightly associated with the assimilation of monomer sugars, rather than cellobiose. We also provided the first genetic evidence to show the indispensability of Crh and CcpA regulator in cellulose degradation. Moreover, our transcriptomic analysis revealed the significant regulatory role of Crh in gene expression at both exponential and stationary phases, the functional specificity and redundancy of LfpC2 and LfpC3 regulators, and in contrast the minimal impact of CcpA on physiological and transcriptional traits when soluble sugars are used as the carbon source. With such new insights into how sugars are utilized in *C. cellulolyticum*, many interesting questions arise worthy of deep investigation. For example, does the phosphorylated Crh mediate the binding affinity of CcpA to tune the transcription of downstream genes? Does CcpA directly regulate the transcription of the *cip-cel* operon by binding the promoter? How does Crh or its phosphorylation status affect TCS and sugar transporters? As well, why does cellobiose utilization not depend on the CCR regulatory system? Our study further understanding these unresolved mechanisms will provide many possibilities to engineer bacteria with high-performance carbon utilization during lignocellulose bioconversion.

5.5 Materials and methods

5.5.1 Plasmids and bacterial strains

To construct the pRNAi-hprK plasmid able to express antisense RNA molecules, a partial transcriptional region of the hprK gene, spanning from the predicted transcriptional start site to the downstream site approximately 170 bp away from the start codon, was amplified with primers, hprK_RNAiF and hprK_RNAiR, and then ligated with KpnI-and-PvuI linearized pRNAi backbone.

To precisely edit the chromosomal genes (i.e., crh, ccpA, lfpC2, and lfpC3), the Cas9 nickase-based genome editing was applied (Xu *et al.*, 2015). For each target gene, a high-specificity 23 bp target site with the format of 5'-(N)20NGG-3' (N=A/T/G/C) was selected first. Then, corresponding all-in-one vectors, consisting of an Fd::cas9n gene, a customized gRNA gene and a homologous donor template, were constructed according to our previous report (Xu *et al.*, 2015). The donor template for editing the *crh* gene was modified to contain a 7-bp DNA fragment with a HindIII enzyme site; other donors targeting *ccpA*, *lfpC2*, and *lfpC3* also had a 7-bp DNA insert but with a EcoRI enzyme site. All these inserts destroyed the selected target sites in donors and then allowed us to identify desired mutants by amplicon digestions. For inactivating both *lfpC2* and *lfpC3* genes simultaneously, the all-in-one vector was constructed with a cas9 nickase, two customized gRNA genes and two specific donor templates. For each complementation experiment, a pair of specific primers was designed to amplify the whole gene cassette from the chromosome. The resulting amplicon contained all possible elements, including the predicted promoter, the ribosome binding site, the open reading frame, and the downstream transcriptional terminator. The plasmid backbone was amplified from pLyc1217Er vector (Li *et al.*, 2012), with primers pErF and pErR. Then, PCR products after DNA purification were assembled together using the Gibson assembly kit. All primers were listed in Table S5.1.

All plasmids were constructed in *E. coli* Top 10 and then verified by DNA sequencing. The plasmids were transformed into *C. cellulolyticum* H10 by electroporation as previously described (Li *et al.*, 2014). Before transformed into the wildtype H10, plasmids pRNAi and pRNAi-hprK were methylated with MspI

methyltransferase (NEB) (Li *et al.*, 2012). The remaining plasmids were transformed into the H10 control strain without DNA methylation. All *E. coli* and *C. cellulolyticum* strains were screened and cultivated as previous described (Xu *et al.*, 2015). The only exception was that *C. cellulolyticum* strains were grown in the defined VM medium which uses a vitamin complex solution to replace yeast extract. Plasmid curing was conducted for each knockout mutant via serial transferring in the antibiotic-free medium (Li *et al.*, 2012). The types and concentrations of carbon sources used to grow *C. cellulolyticum* strains were indicated in the experiments below. Plasmids and bacterial strains were listed in Table S5.2.

5.5.2 Measurement of cell growth, sugar consumption, cellulose consumption, and fermentation products

C. cellulolyticum strains were revived in the VM medium with 2 g/L yeast extract and 5 g/L cellobiose. Antibiotic was added if necessary. Then, fresh cellobiose-grown cultures at an OD600 of 0.5-0.7 were inoculated in the defined VM medium supplemented with vitamin solution and carbon sources, which could be cellobiose, D-glucose, D-xylose, L-arabinose, sugar mixture, or Avicel cellulose. With aforementioned soluble sugars, cell growth was profiled by monitoring OD600 with a spectrophotometer, and time-course samples were taken for HPLC analysis to measure the amount of residual sugars at the corresponding time point. Cell growth on Avicel cellulose, residual cellulose, and fermentation products were determined using the methods described in the section 3.3.5. Colony development was performed on the defined solid VM medium with Avicel cellulose on the agar top (Xu *et al.*, 2014), or a sole sugar as a carbon source, such as 5 g/L cellobiose, 2 g/L D-glucose, 2 g/L D-

xylose, 2 g/L L-arabinose. The specific rate of sugar consumption ($q_{substrate}$) in mmol per gram of cells per hour was the derivative of the time course plots (Desvaux *et al.*, 2000). To compare the growth of *C. cellulolyticum* strains on agar plates with different carbon sources, 10 µl of cells at the same density during the exponential phase were dropped on the agar plates and then incubated at 34 ℃ under anaerobic condition.

5.5.3 Microarray hybridization and data analysis

All *C. cellulolyticum* strains (i.e., Δ*crh,* Δ*ccpA,* Δ*lfpC2,* Δ*lfpC3,* Δ*lfpC2&3,* and the control) were cultivated in the defined VM medium with a mixture of 3g/L cellobiose, 2.5g/L D-glucose, 2g/L D-xylose, 1g/L L-arabinose, in order to mimic the complexity of released sugars during lignocellulose degradation and then capture possible carbon catabolite regulations due to the presence of other sugars. Each strain had three biological replicates. Cell samples with the volume of 10 ml were collected at seven time points: during the exponential phase (21h, 25h, and 28h), at the early stationary phase (40h) and during the stationary phase (54h, 58h, and 61h). After centrifugation at 4 ℃ and 5000 × g for 10 min, cell pellets were retained, immediately frozen with liquid nitrogen, and then stored at -70 ℃. Total RNA was extracted using TRIzol (Invitrogen) and purified using NucleoSpin RNAII kit (Macherey-Nagel) according to the manufacturer's instructions. Then, RNA integrity was estimated by running agarose gels; RNA purity and concentration were measured with NanoDrop spectrophotometer.

For microarray hybridization, we designed 13, 098 probes with 50 nt in length able to interrogate the transcript levels of 94% protein encoding genes in *C. cellulolyticum* and then sent the design to manufacture 8-array slides by Agilent. For each RNA sample, total RNA with the amount of 0.6 µg was reversely transcribed to

Cyanine 3-labeled cDNA using Reverse transcriptase III (Invitrogen) in the reaction with cyanine 3-labeled dUTP. Genomic DNA (gDNA), which was extracted from the control strain using GenElute bacterial genomic DNA kit (Sigma), was labeled by incorporating cyanine 5-labeled dUTP using Klenow DNA polymerase. Each reaction for gDNA labeling contained 1.5 μg qualified gDNA; the resulting product was used for eight hybridizations. Prior to hybridization process, all labeled cDNA and gDNA were purified with QIAquick PCR purification reagents (Qiagen) and SpinSmart columns (Denville Scientific Inc), and then lyophilized for use. Labeled cDNA and gDNA were mixed in the hybridization master buffer containing 8% formamide, followed with denaturing at 95 ℃ for 3 min, incubation at 37 ℃ for 30 min and finally loading onto the array. Hybridization was carried out at 67 ℃ and 20 rpm for 22 h. Slides were washed and then scanned using NimbleGen MS200 (Roche) with the following settings: two-channel scanning, 2 μm scanning resolution, 100% laser strength, 30% gain percentage. With Agilent feature extraction 11.5, all digital images were manually checked to confirm the gridding quality, from which raw data was extracted.

Microarray data analysis was performed using limma package in R (Ritchie *et al.*, 2015). First, probes with both qualified green and red signals were screened (single-to-noise ratio>2, signal-to-background ratio >1.3, coefficient of variation<0.8, minimal gMeansigal>150, and minimal rMeansignal>50) (He *et al.*, 2008). Second, the mean signals of each probe were applied to background correction by subtraction, whithin-array normalization by loess, and then between-array normalization by quantile. Third, with all normalized data, gene probes with significantly different expression levels were identified using limma's linear model and then evaluated by empirical Bayes methods.

The transcriptional level of genes was calculated by averaging the values of qualified probes only if half or more probes of this gene were qualified. In this study, differentially expressed genes (DEGs) refer to genes with a log2 fold-change above 1 (or below -1) and an adjusted p value below 0.05. Hierarchical clustering of all DEGs occurred in any mutant was performed using the gplots package with the ward.D method (Warnes, 2009). Venn diagram graphs were generated with the online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Blast2go was applied to do GO enrichment with the Fisher's exact test ($p<0.01$, two sided) (Conesa *et al.*, 2005).

5.5.4 Quantitative PCR

To validate microarray results, ten differentially expressed genes presented at the exponential phase of either Δ*crh* or Δ*lfpC2&3* were selected for the analysis of quantitative PCR. RNA samples of the control strain and mutants Δ*crh* and Δ*lfpC2&3*, were the same as these ones used for microarray hybridization. For each biological replicate, the same amount of total RNA at each time point during the exponential phase was pooled together. Then, SuperScript III Reverse Transcriptase (Invitrogen) was applied to convert 1 μg of pooled RNA to cDNA by following the manufacturer's protocol. cDNA products were diluted 10 times before used as PCR templates. qPCR was performed using iTaq SYBR Green Supermix with ROX (Bio-Rad) on a Bio-Rad iQ5 thermal cycler. The thermal cycling conditions were as follows: 95 ℃ for 3 min, 40 cycles of 95 ℃ for 15 s, 55 ℃ for 15 s and 72 ℃ for 30 s. The *recA* gene was used as an internal calibrator (Stevenson *et al.*, 2005). Relative transcript levels of these ten genes in mutants was calculated with the Pfaffl Method (Pfaffl, 2001). Finally, the fold

changes validated by qPCR were log2 transformed prior to regression analysis with microarray data. Gene-specific primers are listed in Table S5.1.

Chapter 6 Summary and Outlook

Lignocellulosic biofuels have the potential to mitigate the pressure on energy supply and environmental sustainability. However, the pace of microbial engineering towards efficient and cost-effective biofuel production has been hampered mainly due to our insufficient understanding of biological systems in potential microbes and the lack of simple and efficient genome editing tools for functional genomics studies and genetic engineering-oriented practices. By using the cellulolytic model organism *C. cellulolyticum*, this study contributes to alleviating these challenges in two ways: first, by developing robust genome editing tools that allow microbiologists and microbial engineers to efficiently manipulate both essential and non-essential genes in microbes; and second, by doing comprehensive studies on key metabolic genes, cellulose-degrading cellulosomes, and catabolite regulation systems in the CBP-enabling *C. cellulolyticum*. Major achievements are summarized below.

First, this work timely modified the bacterial Cas9 system to edit the *C. cellulolyticum* genome. Previous reported strategies (e.g., Cas9-HR and Cas9-NHEJ) (Cong *et al.*, 2013; Dicarlo *et al.*, 2013; Gratz *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013; Yang *et al.*, 2013) failed to edit the genome in this bacterium because of the limited ability of the native DNA repair systems to fix Cas9-induced DNA breaks. The method developed here successfully circumvented the toxicity of DNA breaks by applying the Cas9 nickase protein to generate a single nick to trigger homologous recombination. It presented the advantage of marker-independent gene delivery and versatile editing in a single step at a very high editing efficiency and specificity. Because it needs a very short 23 bp target sequence (van der Ploeg, 2009), this method

has a very high target site density in bacterial genomes such that nearly all genes can be targeted using this method. Furthermore, Chapter 5 demonstrated its ability to perform multiplex editing, for instance, simultaneous modification of two target genes using a single all-in-one vector. Obviously, these outstanding features make this Cas9 nickase-based genome editing tool (or called a single nick-triggered homologous recombination strategy) unmatchable by current widely-used editing tools, such as double cross-over recombination (Heap *et al.*, 2012) and TargeTron (Enyeart *et al.*, 2013; Esvelt *et al.*, 2013). This method tremendously speeds up functional genomics studies in *C. cellulolyticum* as described in Chapter 5. This exemplary strategy can be expanded to other microbes (including those sensitive to DNA breaks) to facilitate microbial genome editing for fundamental and applied research. To meet the demand of metabolic engineers, the genetic cargo capability of this method needs to be improved in order to integrate multiple genes or very long DNA fragments into the chromosome. To help the users to select a perfect target site and skip difficult design procedures, we are endeavoring to construct an integrated Cas9 target web database for sequenced microbes, which can provide user-friendly functions for target visualization and alignment with functional elements (e.g., promoters, conserved protein domains, and terminators), comparison of target sets, as well as one-stop primer design for plasmid construction and mutant identification.

Second, the essential acetate producing pathway was stably manipulated by delivering an enhanced antisense RNA expression module into the chromosome using the Cas9 nickase editing tool. The effectiveness of plasmid-based RNA repression was obvious; however, chromosome-based repression appeared to be so weak. The reason

we identified in this study was that switching from plasmid-based expression to chromosome-based expression dramatically decreased gene dosage, along with much less functional gene products formed in the end. Microbial engineers should keep that in mind if they want to express foreign genes/pathways in chromosomal integrants. The challenge was solved by one-step integration of a tandem promoter-driven RNA expression module using the Cas9 nickase editing tool. It sets an example of stable manipulation of essential metabolic genes in microbes and in plasmid/antibiotic-independent microbial fermentation. Technically, antisense RNAs successfully repressed the *pta* gene in Chapter 3 and the *hprK* gene in Chapter 5, whereas it did not work very well for the *ack* gene. The variation in RNA repression indicates the importance of antisense RNA design (e.g., RNA targeting region, promoter strength and RNA structure).

Third, three cellulosomal components, Dpi, Cel48F and Cel9E, were proven to be important for cellulolysis in *C. cellulolyticum*. Purified recombinant Dpi showed in vitro inhibitory activity against cysteine protease. Besides a statistically significant change in the expression of the *cip-cel* operon, inactivation of this Dpi encoding gene dramatically disturbed cellulosome stoichiometry, particularly a sharp decrease in the abundance of major cellulosomal components, Cel48F endocellulase and Cel9E exoglucanase (Maamar *et al.*, 2004; Perret *et al.*, 2004). Our study then verified the indispensable contribution of Cel48F and Cel9E cellulases in cellulose degradation. Taken together, this study connected the functional relationships between cellulosomal protease inhibitors and other cellulosomal enzymes. Although a few cellulosomal protease inhibitors have been characterized (Kang *et al.*, 2006; Meguro *et al.*, 2011), our

study provides the first evidence showing the in vivo importance of cellulosomal protease inhibitors in protecting cellulosomal components from proteolysis. The findings of key cellulases and protease inhibitor-mediated cellulosome maintenance may suggest other ways to improve cellulose hydrolysis.

Fourth, carbohydrate utilization in *C. cellulolyticum* differentially relied on the CCR system. This bacterium has a partial CCR in the genome without any predictable components of PTS (Abdou *et al.*, 2008). It just showed a very mild reverse CCR in light of simultaneous assimilation of multiple sugars and no preference to glucose (Goerke *et al.*, 2008). These rare merits can help bacteria to make the most out of carbons that are loaded during fermentation. With our systematic characterization of CCR mutants (e.g., Δ*hprK*, Δ*crh*, and Δ*ccpA*), it seems like cellobiose assimilation is independent of CCR under our test condition, but the utilization of monomers (both pentoses and hexoses) and insoluble cellulose are tightly associated with CCR. More importantly, while the potential regulatory role of CCR in cellulose degradation was proposed a long time ago (Abdou *et al.*, 2008; Xu *et al.*, 2013), this study provides the first genetic evidence to show the indispensability of Crh and CcpA in cellulose catabolism. Apart from that, this study also provide an overall view of biological functions of CCR components, such as the significant regulatory role of Crh, the functional specificity and redundancy of LfpC2 and LfpC3 regulators, and the minimal impact of CcpA on physiological and transcriptional traits when soluble sugars are used as the carbon source. As new insights into this unique CCR were generated in this study, many interesting questions arise. Much more effort is needed to solve the detailed mechanisms behind our observations.

In conclusion, our study provided novel insights into the physiological and genetic importance of a series of genes associated with sugar assimilation, cellulose degradation and even cellular metabolism in the consolidated bioprocessing candidate *C. cellulolyticum*. Aforementioned discoveries will direct microbial engineers to develop more feasible strategies to improve lignocellulose bioconversion. The developed Cas9 nickase-based genome editing tool and its derivative, Cas9 nickase-assisted RNA repression, will naturally facilitate the pace of functional genomics studies in microbes and microbial engineering for application to real-world problems.

The results from this study and other associated projects that I have been involved in are mainly reflected in the publications as listed below.

1. Tao Xu, et al. CTD: an integrative CRISPR-Cas targets database for genome modification and regulation in microorganisms. (In preparation)
2. Tao Xu, et al. Physiological, genetic and transcriptomic deciphering of the reduced carbon catabolite regulation system in the bioenergy-related *Clostridium cellulolyticum*. (In preparation)
3. Tao Xu, Yongchao Li, Joy D. Van Nostrand, Zhili He, Jizhong Zhou. Cas9 nickase-assisted RNA repression enables stable and efficient manipulation of essential metabolic genes in *Clostridium cellulolyticum*. (Submitted)
4. Tao Xu, Yongchao Li, Zhou Shi, Christopher L. Hemme, Yuan Li, Yonghua Zhu, Joy D. Van Nostrand, Zhili He, Jizhong Zhou. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. Applied and Environmental Microbiology. 2015, 81:4423-4431
5. Tao Xu, Yongchao Li, Joy D. Van Nostrand, Zhili He, Jizhong Zhou. Cas9-based tools for targeted genome editing and transcriptional control. Applied and Environmental Microbiology. 2014, 80(5):1544
6. Tao Xu, Yongchao Li, Zhili He, Jizhong Zhou. Dockerin-containing protease inhibitor protects key cellulosomal cellulases from proteolysis in *Clostridium cellulolyticum.* Molecular Microbiology. 2014, 91(4):694-705
7. Yongchao Li, Tao Xu, Timothy J. Tschaplinski, Nancy L. Engle, Yunfeng Yang, David E. Graham, Zhili He and Jizhong Zhou. Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation abolishment and carbon alleviation.

Biotechnology for Biofuels. 2014, 7:25

8. Laura Bartley, Tao Xu, Chengcheng Zhang, Hoang Nguyen, Jizhong Zhou. Switchgrass Biomass Content, Synthesis, and Biochemical Conversion to Biofuels. In: Compendium of Bioenergy Plants – Switchgrass. Luo, H., and Wu, Y. (eds), CRC Press, Tailor & Francis Group. 2014, 109-169.

Appendix A: Supplementary Tables

Table S2.1 Target sites selected for genome editing in *C. cellulolyticum*.

Table S2.2 Summary of primers used in this study.

Table S2.3 Bioinformatic analysis of targeting space in multiple bacteria.

Table S3.1 List of primers used in this study.

Table S3.2 Measurement and comparison of product titers and molar ratios in the fermentation broth of *C. cellulolyticum* strains grown on cellulose and xylan.

Table S3.3 Measurement of alcohol dehydrogenase activity in crude extracts.

Table S4.1 Mass spectrometry analysis of gel slices from SDS-PAGE.

Table S4.2 Oligonucleotide primers in this study.

Table S5.1 Primers used to study carbon catabolite regulation.

Table S5.2 Bacterial strains and plasmids constructed to study carbon catabolite regulation.

Table S5.3 List of genes and their associated gene ontology (GO) terms in mutants at the exponential phase.

Table S5.4 List of genes and their associated gene ontology (GO) terms in mutants at the stationary phase.

Table S5.5 List of selected DEGs related to two-component system (TCS), transporter and transcriptional regulators (TF).

Table S2.1 Target sites selected for genome editing in *C. cellulolyticum*.

Target ID	Name	23 bp Target site (N20+NGG)	Specificity	GC%[a]
Ccel_0614	pyrF	TATGAAATGTATGGAATTGATGG	High	25%
Ccel_2866	mspI	ATTAAAGAAGGGTACTCTATAGG	High	30%
Ccel_0374	β-gal	AGAAGGTTTCGTTTGGGGTACGG	High	45%
Ccel_3198[b]	3198D	AAGTAAGAAACATTTGGTTCCGG	High	30%
Ccel_0728	X21	AAAATAACTCTTACACCAAACGG	Low	25%
Ccel_0728	X22	AATGTAACTCTTACACCAAACGG	Low	30%

a, GC content in 20-bp target recognition sequence (red); b, intergenic region downstream of Ccel_3198.

Table S2.2 Summary of primers used in this study.

Primer	Sequence	Note
P4F	GGAATTCTAGACATAATATATTGACAAATTTATTTTTTAAAGTT	P4 promoter generation
P4R	CGGGGTACCTCCTAACAACTTAATTTTAACTTTAAAAAATAAATTT	P4 promoter generation
PromF	CAGCTATGACCATGATTACGCGTTGCAACAAATTGATGAG	P4 promoter PCR
PromR	GAGTTTTGCGTTGATCATTGATAAGTACCTCCTAACAACTTAATTATAC	P4 promoter PCR
Cas9nF	GTATTCAATAGGACTGGCAATAGGAACAAATAGCGTAGGATGGGCAGTAATTACA	*cas9* nickase generation
Cas9nR	CGCTATTTGTTCCTATTGCCAGTCCTATTGAATACTTTTTATCCATATGA	*cas9* nickase generation
P4gRF	TTATGGTACCCGGGGATCCTCGTTGCAACAAATTGATGAGCAATG	gRNA retargeting
gRNAR	CAACTGTTGGGAAGGGCGATCCGCGTCTAGAGCCGATCGA	gRNA retargeting
P4gRR	CTAAAACGCAGGTGAGTACAACCTGCCGTACAACTTAATTTTAACTTTAAAAAATAAATTTGT	gRNA control
gRCKF	GTTGTACGGCAGGTTGTACTCACCTGCGTTTTAGAGCTAGAAATAGCAAGT	gRNA control
P40614R	CTAAAACTCAATTCCATACATTTCATACAACTTAATTTTAACTTTAAAAAATAAATTTGT	*pyrF* gRNA
G0614F	GTTGTATGAAATGTATGGAATTGAGTTTTAGAGCTAGAAATAGCAAGT	*pyrF* gRNA
P42866R	CTAAAACATAGAGTACCCTTCTTTAATCAACTTAATTTTAACTTTAAAAAATAAATTTGT	*mspI* gRNA
G2866F	GTTGATTAAAGAAGGGTACTCTATGTTTTAGAGCTAGAAATAGCAAGT	*mspI* gRNA
P40374R	CTAAAACTACCCCAAACGAAACCTTCTCAACTTAATTTTAACTTTAAAAAATAAATTTGT	*β-gal* gRNA
G0374F	GTTGAGAAGGTTTCGTTTGGGGTAGTTTTAGAGCTAGAAATAGCAAGT	*β-gal* gRNA
P43198DR	CTAAAACGAACCAAATGTTTCTTACTTCAACTTAATTTTAACTTTAAAAAATAAATTTGT	3198D gRNA
G3198DF	GTTGAAGTAAGAAACATTTGGTTCGTTTTAGAGCTAGAAATAGCAAGT	3198D gRNA
0614LF	GAATTTTATTATGGTACCCGGGTCTTGGTTTGAAAGGCAATCCT	1-kb LH of *pyrF* donor
0614LR	CATATTTACAGGTTTCCTGGAAAGCAATCAATGTAAGCAAGCTGTGGCTTAACTGCGGGAACCTG	1-kb LH of *pyrF* donor
0614RF	CAGGTTCCCGCAGTTAAGCCACAGCTTGCTTACATTGATTGCTTTCCAGGAAACCTGTAAATATG	1-kb RH of *pyrF* donor
0614RR	CTCATCAATTTGTTGCAACGAGCTCAAGACCTGTTATCTCATTTCTTTTG	1-kb RH of *pyrF* donor
2866LF	GAATTTTATTATGGTACCCGGGAAGTCTGTAGCAACAGATTCTAGTTGTTCC	1-kb LH of *mspI* donor
2866LR	CCATTTTAAATTGCTTTTCTTGATTTGGGTAATTCTATATTAATCCCTAATTCATTTTTAAGATTATTTAGC	1-kb LH of *mspI* donor
2866RF	GCTAAATAATCTTAAAAATGAATTAGGGATTAATATAGAATTACCCAAATCAAGAAAAGCAATTTAAAATGG	1-kb RH of *mspI* donor
2866RR	CTCATCAATTTGTTGCAACGAGGACCACGCTTTTTGCTTGGATAAGTCC	1-kb RH of *mspI* donor
0374-F1	GAATTTTATTATGGTACCCGGGCAGGACCATGAAGGAACATATGAC	1-kb LH of *β-gal* donor
0374LR	TGATATGATGCTGTTGCCGTACGATATCCAAACGAAACCTTCTTTGAAT	LH of *β-gal* donor
0374RF	ATTCAAAGAAGGTTTCGTTTGGATATCGTACGGCAACAGCATCATATCA	RH of *β-gal* donor
0374RR1	GCTCATCAATTTGTTGCAACGAGCTAGTTACCCAGTACAGAGTTTCC	1-kb RH of *β-gal* donor
0374LF0.5	CTGAATTTTATTATGGTACCCGGGGCGGAGATACCTGAAGAACTTG	0.5-kb LH of *β-gal* donor
0374RR0.5	TTGCTCATCAATTTGTTGCAACGAGCCACCTGGTTTACCGGAAGCTT	0.5-kb RH of *β-gal* donor

0374LF0.2	TATTATGGTACCCGGGGTGTAGGTGATAAAGTTGAGCATAAG	0.2-kb LH of *β-gal* donor
0374RR0.2	CAATTTGTTGCAACGAGACCTATATGCCTTAATACCGATTTC	0.2-kb RH of *β-gal* donor
0374LF0.1	TATTATGGTACCCGGGGTGCTGGAATGAAAAGACTGATG	0.1-kb LH of *β-gal* donor
0374RR0.1	CAATTTGTTGCAACGAGCATCAATTTTGCCTTTCATCCTG	0.1-kb RH of *β-gal* donor
X21gRNAF	GTTGAAAATAACTCTTACACCAAAGTTTTAGAGCTAGAAATAGCAAGT	X21 gRNA
X21gRNAR	CTAAAACTTTGGTGTAAGAGTTATTTTCAACTTAATTTTAACTTTAAAAAATAAATTTGT	X21 gRNA
X22gRNAF	GTTGAATGTAACTCTTACACCAAAGTTTTAGAGCTAGAAATAGCAAGT	X22 gRNA
X22gRNAR	CTAAAACTTTGGTGTAAGAGTTACATTCAACTTAATTTTAACTTTAAAAAATAAATTTGT	X22 gRNA
X21LF	TATTATGGTACCCGGGTATCGTTAATTAAAAATCTAATAAAAAGTGATTATAAAAAATATC	1-kb LH of X21 donor
X21LDR	TCTGTAATTCACTAATTCCATTGAATGGTGTAAGAGTTATTTTAGTATCT	1-kb LH of X21 donor
X21RDF	AGATACTAAAATAACTCTTACACCATTCAATGGAATTAGTGAATTACAGA	1-kb RH of X21 donor
X21RR	CAATTTGTTGCAACGAGCCATAGTACCGTCACCGAAAG	1-kb RH of X21 donor
X22LF	TATTATGGTACCCGGGAAACGGTAGTGTAACTATAGTTCCCGGTATTCAACCTACAAAGGA	1-kb LH of X22 donor
X22LDR	ATGTCAAACCTGTGATACCTTTGAATGGTGTAAGAGTTACATTTACATTT	1-kb LH of X22 donor
X22RDF	AAATGTAAATGTAACTCTTACACCATTCAAAGGTATCACAGGTTTGACAT	1-kb RH of X22 donor
X22RR	CAATTTGTTGCAACGAGAACCACATATAATAGGACATAGC	1-kb RH of X22 donor
pLMD1	CTTGAAGACGAAAGGGCCTCGTGAT	pBR322 backbone PCR
pLMD4	AATTCTCATGTTTGACAGCTTATCATCGATAAG	pBR322 backbone PCR
pLMD2	GGCCCTTTCGTCTTCAAGTCTTGGTTTGAAAGGCAATC	2-kb *pyrF* fragment
pLMD3	CTGTCAAACATGAGAATTCCAGTTACTTGGAGTTTTAC	2-kb *pyrF* fragment
pLMD5	ATTCCATACATTTCATAGTAAGC	pLMD-arm backbone PCR
pLMD8	TGATGGATTGATTGCTTTCC	pLMD-arm backbone PCR
pLMD9	ACTATGAAATGTATGGAATAGCATGCCGGAGCAAATGAG	6kb λ DNA PCR
pLMD10	AAAGCAATCAATCCATCACAGCAGCTCCTTGCCGAGAT	6kb λ DNA PCR
pLMD11	ACTATGAAATGTATGGAATGGCTGCTCTGAAGGCGGTGT	3kb λ DNA PCR
pLMD12	AAAGCAATCAATCCATCAAGGCCAGATACTGCGAGGTG	3kb λ DNA PCR
P3198D01	CTTGAAGACGAAAGGGCCTC	3198D donor
P3198D02	AATTCTCATGTTTGACAGCTTATC	3198D donor
P3198D03	CCCTTTCGTCTTCAAGAATGAGGGATTTCAAACCTG	3198D donor
P3198D04	CTGTCAAACATGAGAATTTTACTTGCCGTAGTAAACTTTC	3198D donor
P3198D05	TTACTTGCCGTAGTAAACTTTC	3198D donor
P3198D06	GTTTACTACGGCAAGTAAAGGAGGTTTACAATGACAAAAG	3198D donor
P3198D07	CTGTCAAACATGAGAATTTTAGAGAGCTTTCGTTTTCATG	3198D donor
P3198D08	TTAGAGAGCTTTCGTTTTCATG	3198D donor
P3198D09	AAAACGAAAGCTCTCTAATTTTGTACCGGGCACGTGGT	3198D donor
P3198D10	CTGTCAAACATGAGAATTAAAGTACCGGGAACTGCCTG	3198D donor
P3198D19	TATTATGGTACCCGGGAATGAGGGATTTCAAACCTGAC	3198D donor
P3198D20	CAATTTGTTGCAACGAGAAAGTACCGGGAACTGCCTG	3198D donor
242LAFPR	CGATCCGGGGCGCGCATGCCTGCAGGATTCCATACATTTCATAGTAAGCAAGCTGTGG	1-kb LH of *pyrF* donor
242RAFPF	GCAGGGCCAGGCCAAGCACTGAACGCGTAGTGATGGATTGATTGCTTTCCAGGAAACCTG	1-kb RH of *pyrF* donor
AFPF	CCACAGCTTGCTTACTATGAAATGTATGGAATCCTGCAGGCATGCGCGCCCCGGATCG	*afp* cassette
AFPR	CAGGTTTCCTGGAAAGCAATCAATCCATCACTACGCGTTCAGTGCTTGGCCTGGCCCTGC	*afp* cassette
p1	ATGCAATATGCTTCAATGTTTGATA	Δ*pyrF* and Δ*pyrF/afp*$^{+}$

		mutant identification
p2	ATCCTTCTCCTTCGTAGTGCTTTAT	Δ*pyrF* and Δ*pyrF/afp*$^{+}$ mutant identification
p3	CAGGACCATGAAGGAACATATGAC	Δ*gal* mutant identification
p4	CTGGCAATTTATATCTCTCGGA	Δ*gal* mutant identification
p5	GGTTTCGTTTGGATATCGT	Δ*gal* mutant identification
p6	AAGCAATAGAAGATTTAGGATTTACTG	Δ*pyrF/afp*$^{+}$ mutant identification
p7	CAGTAAATCCTAAATCTTCTATTGCTT	Δ*pyrF/afp*$^{+}$ mutant identification
p8	AGCATCATACCTTCTTTGATGTAGC	*alsS*$^{+}$ mutant identification
p9	CCTCCAGAGTACCAGTTAATTCTGA	*alsS*$^{+}$ mutant identification
p10	GATATCGTGAAATATGCGGAAAGC	*alsS*$^{+}$ mutant identification
p11	GGATAATTGTCCGGTCTCCA	*alsS*$^{+}$ mutant identification
P12	CACTCACACGGGTCTGTACC	Δ*mspI* mutant identification
P13	CGGAGAAACAGGGCTTCGAT	Δ*mspI* mutant identification
P14	ATTAAAGAAGGGTACTCTATAGG	Δ*mspI* mutant identification
RTgRNAF	GGCAGGTTGTACTCACCTGCGT	gRNA semi-qPCR
RTgRNAR	AAGCACCGACTCGGTGCCAC	gRNA semi-qPCR
RTrecAF	GCAAAGAAACTTGGGGTTGA	*recA* semi-qPCR and qPCR
RTrecAR	TGAGACATCAGCCTTGCTTG	*recA* semi-qPCR and qPCR
RTkuF	TACGGCAACGGAAGATAAGG	*ku* qPCR
RTkuR	CCGGGCTCATATTCAAATCC	*ku* qPCR
RTatpF	GCAGCCTAATTCGTTGGTTC	*ATP-dependent DNA ligase* qPCR
RTatpR	CATTGCTCCTGTTCAAGCTG	*ATP-dependent DNA ligase* qPCR
RTligDF	GGAAGCCGGAATTACCAAAC	*DNA polymerase ligD* qPCR
RTligDR	CCCGTGAGGATAACGAATTG	*DNA polymerase ligD* qPCR
RTcel48F	AACAAACCGGCTACATACGC	*cel48F* qPCR
RTcel48R	GGTTCCATCAGCTCTTGCTC	*cel48F* qPCR

Table S2.3 Bioinformatic analysis of targeting space in multiple bacteria.

Name	*Clostridium cellulolyticum* H10 (NC_011898.1)	*Clostridium acetobutylicum* ATCC 824 (NC_003030.1)	*Escherichia coli* K-12 (NC_000913.3)	*Bacillus subtilis* 168 (NC_000964.3)	*Lactobacillus reuteri* DSM 20016 (NC_009513.1)
Genome size (bp)	4068724	3940880	4641652	4215606	1999618
GC content (%)	37.4	30.93	50.79	43.51	38.87
Total N21GG[a]	342148	228988	542072	386029	163816
Usable target sites[b]	256542	159329	464742	316852	120839
Percentage of usable target sites[c]	75%	70%	86%	82%	74%
Mean distance (bp)	15.86	24.73	9.99	13.3	16.55
Median distance (bp)	9	14	6	8	9
No. of UR[d] (>1kb)	91	16	50	15	37
Maximal UR (kb)	21.933	11.827	5.163	5.355	5.736
Gene coverage[e]	95.69%	99.75%	98.48%	99.93%	95.74%

Note: a and b, their definitions can be found in the method; c, the percentage of usable target sites in total N21GG sites; d, the number of unaccessible regions (UR) with the length of > 1kb; e, the percentage of genes that have at least one usable target site inside. All sequences were downloaded from NCBI using their corresponding GenBank accession numbers.

Table S3.1 List of primers used in this study.

Primer	Sequence	Note
ptaF	aaccgagctcggtacccgggCCTTGCTCCTGAATCATTG	Amplify partial pta region
ptaR	gcgatcgttcgactctagagAAGACTTTCAGTTTGATAATG TTAATC	Amplify partial pta region
ackF	aaccgagctcggtacccgggAAAATAACAGCCTCTTCTGA AC	Amplify partial ack region
ackR	gcgatcgttcgactctagagATTTATTATAAATGATGTTAA CTATTATAAGG	Amplify partial ack region
pRNAiF	GTAATCATGGTCATAGCTGTTTCCTG	Amplify pRNAi backbone
pRNAiR	CAGTTGCGCAGCCTGAATGG	Amplify pRNAi backbone
3198upF	acagctatgaccatgattacGGCCGGCCATTCCATCACCT	Amplify upper homologous region
3198upR	tccttgtacaTTACTTGCCGTAGTAAACTTTCAGGTAC ATGTC	Amplify upper homologous region
mlsRF	cggcaagtaaTGTACAAGGAGGTTTACAATG	Amplify promoter-less mlsR
mlsRR	gcgcgaattcCCATGGTTACTTATTAAATAATTTATA GC	Amplify promoter-less mlsR
asRNAF	gtaaccatggGAATTCGCGCCCCGGATCGA	Amplify asRNA cassette
asRNAR	tttcctcgagTTTTATAGGGCGTGTTTGTGGCTTAGAG	Amplify asRNA cassette
3198downF	ccctataaaaCTCGAGGAAACATTTGGTTC	Amplify lower homologous region
3198downR	ccattcaggctgcgcaactgAAGCTTTTTGCATTTTTTACTT C	Amplify lower homologous region
afpF	gtaaccatggGAATTCGAGCTCGGTACCCG	Amplify Fd::afp cassette
afpR	tttcctcgagTTTTATAGGGCGTGTTTGTGGC	Amplify Fd::afp cassette
ldhF	TATACTTTTGCACCCAGAATGTTTT	Identify Δldh mutants
ldhR	TGACTGATACGGGTTTTATCAATTT	Identify Δldh mutants
mdhF	GGGATTTTAATGGGTTTTAAAGTTG	Identify Δmdh mutants
mdhR	TCCAGGTGAATAAGCTAAAGAAAGA	Identify Δmdh mutants
InF	CAGGCAACTAAGAACATTTTTGAAT	Identify asRNA integrants
InR	CCTCCAGAGTACCAGTTAATTCTGA	Identify asRNA integrants
3P4F	gcaagtaagaaacatttggcGCTTCACGTGATCCATGGCA	Amplify 3P4 promoter cluster
3P4R	GATTCAGGAGCAAGGATCCATGGAAGCTTCAAC	Amplify 3P4 promoter cluster
3P4ptaF	CCATGGATCCTTGCTCCTGAATCATTGCAGC	Amplify partial pta region for 3P4::pta asRNA construct
3P4ptaR	gcccggtacaaaaccggaacAAGACTTTCAGTTTGATAAT GTTAATCGG	Amplify partial pta region for 3P4::pta asRNA construct
3198LF	gaattttattatggtacccgggGTGCTGAGGCAATGAGGGAT	Amplify 1-kb right arm of 3198 donor
3198LR	ggaaCCATGGccaaatgtttcTTACTTGCCGTAG	Amplify 1-kb right arm of 3198 donor
3198RF	GGCAAGTAAgaaacatttggCCATGGttccggttttgtaccgggc acgtgg	Amplify 1.4-kb left arm of 3198 donor
3198RR	catcaatttgttgcaacgagACGGCATTCTTTGTAGCCCA	Amplify 1.4-kb left arm of 3198 donor
RTafpF	ATCAGGACGCCCGTTTTCTT	qRTPCR
RTafpR	ACCGGTGTGATGGACAACTC	qRTPCR
RTrecAF	GCAAAGAAACTTGGGGTTGA	qRTPCR
RTrecAR	TGAGACATCAGCCTTGCTTG	qRTPCR

Table S3.2 Measurement and comparison of product titers and molar ratios in the fermentation broth of *C. cellulolyticum* strains grown on cellulose and xylan.

Carbon	Strain	Lactate[a]	Acetate	Ethanol	Molar ratio of lactate : acetate : ethanol[d]	Carbon in ethanol (%)[e]
10 g/L Cellulose	WT-P-pta/	1.159±0.154[b]/	1.122±0.071/	0.639±0.021/	0.93 : 1.37 : 1/	26.56/
	WT-P	2.358±0.199 (0.49)**[c]	1.319±0.027 (0.85)*	0.691±0.036	1.75 : 1.49 : 1	19.55
	LM-P-pta/	0.002±0.008/	0.503±0.077/	1.963±0.203/	0.001 : 0.2 : 1/	83.23/
	LM-P	0.078±0.002 (0.03)**	0.747±0.040 (0.67)**	1.055±0.162 (1.86)**	0.04 : 0.55 : 1	62.11
	LM-G-pta/	0.024±0.004/	0.812±0.079/	2.074±0.308/	0.01 : 0.31 : 1/	75.47/
	LM-G	0.110±0.011 (0.22)**	0.914±0.026	1.211±0.283 (1.71)**	0.05 : 0.59 : 1	60.06
10 g/L Xylan	WT-P-pta/	0.251±0.132/	0.884±0.209/	0.168±0.038/	0.76 : 4.11 : 1/	16.00/
	WT-P	0.357±0.085	1.854±0.193 (0.48)**	0.241±0.042	0.76 : 6 : 1	12.29
	LM-P-pta/	0.019±0.004/	0.588±0.085/	0.906±0.144/	0.01 : 0.51 : 1/	65.57/
	LM-P	0.104±0.011 (0.18)**	0.894±0.027(0.66)**	0.844±0.145 (1.07)**	0.06 : 0.83 : 1	52.08
	LM-G-pta/	0.059±0.034/	0.902±0.115/	0.835±0.172/	0.04 : 0.84 : 1/	52.63/
	LM-G	0.126±0.019 (0.47)*	1.091±0.054 (0.83)*	0.754±0.091	0.09 : 1.13 : 1	44.15

[a]Product titers (g/L) and molar ratios of the engineered strain and its control are divided by a forward slash.
[b]All data is presented as mean± standard deviation (n=3). Student's t test was applied to test the significance of a difference (*P<0.05;**P<0.01).
[c]The numbers in parentheses represent the titer ratio if the difference is statistically significant.
[d]The molar ratio is calculated by normalizing to ethanol.
[e]The percentage of ethanol-containing carbon in three major metabolites including lactate, acetate and ethanol.

Table S3.3 Measurement of alcohol dehydrogenase activity in crude extracts.

Cellobiose-grown strain	ADH activity (U/mg protein)
LM-G	0.053±0.003
LM-G-pta	0.062±0.003
WT	0.008±0.001
LM	0.037±0.001
LM3P	0.052±0.001

Values are presented as mean± standard deviation.

Table S4.1 Mass spectrometry analysis of gel slices from SDS-PAGE.

Hit	Accession	Mascot Score	Mass (Da)	Avg. Intensity	Gene number	Predicted functions
B1	gi|220928179	2608	80608	5.611e+4	Ccel_0729	Cel48F
B2	gi|220928182	3301	97127	9.375e+4	Ccel_0732	Cel9E
B3	gi|220928185	1271	85039	1.659e+4	Ccel_0735	Cel9J
B4	gi|220928187	1058	58027	2.347e+4	Ccel_0737	Cel9M

Table S4.2 Oligonucleotide primers in this study.

Primer name	Sequence	Note
EBSu	CGAAATTAGAAACTTGCGTTCAGTAAAC	Intron modification
Dpi-171|172a-IBS	AAAACCCGGGATAATTATCCTTAGACCTCA TTATCGTGCGCCCAGATAGGGTG	Intron modification
Dpi-171|172a-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATA AGTCATTATCACTAACTTACCTTTCTTTGT	Intron modification
Dpi-171|172a-EBS2	TGAACGCAAGTTTCTAATTTCGATTAGGTCT CGATAGAGGAAAGTGTCT	Intron modification
Cel48F-764a-IBS	AAAACCCGGGATAATTATCCTTACTGGACC ATTTGGTGCGCCCAGATAGGGTG	Intron modification
Cel48F-764a-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATA AGTCCATTTGTATAACTTACCTTTCTTTGT	Intron modification
Cel48F-764a-EBS2	TGAACGCAAGTTTCTAATTTCGGTTTCCAGT CGATAGAGGAAAGTGTCT	Intron modification
Cel9E-653a-IBS	AAAACCCGGGATAATTATCCTTACTACCCG ATTCAGTGCGCCCAGATAGGGTG	Intron modification
Cel9E-653a-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATA AGTCGATTCAAATAACTTACCTTTCTTTGT	Intron modification
Cel9E-653a-EBS2	TGAACGCAAGTTTCTAATTTCGGTTGGTAG TCGATAGAGGAAAGTGTCT	Intron modification
Dpi-171 F	TTGCTCCGGCAAAAGTAAAC	Mutant identification
Dpi-171 R	CACTGATAGCCCGTTGATCC	Mutant identification
Cel48F F	GATGAACATAAATTTGGTGGACAGT	Mutant identification
Cel48F R	TGCATAGTACCATGAAAGCAGATAA	Mutant identification
Cel9E F	CTGGAATTACAGGCTAATACTCCAA	Mutant identification
Cel9E R	TGCAATACCACCATTAACAACATAC	Mutant identification
Intron F1	CCTATGGGAACGAAACGAAA	Mutant identification
Intron R1	CGAGTACTCCGTACCCTTGC	Mutant identification
NtDpi F(NdeI)	GGAATTCCATATGGTGGTAGGAAGTTATAC ACTTTTCGG	Construct pET28a(+)-Dpi expression vector
NtDpi R(Not I)	ATAGTTTAGCGGCCGCTTAAATTACATTTAT TTCACATTGG	Construct pET28a(+)-Dpi expression vector
Dpi-over F	CGCGGATCCCCCGGGATGGAAAAGAATTAC ACACCAA	Construct pClostron3-Dpiover vector
Dpi-over R	TTATTTCGATCGTTAAATTACATTTATTTCA CA	Construct pClostron3-Dpiover vector
RTrecA F	GCAAAGAAACTTGGGGTTGA	*recA* qPCR
RTrecA R	TGAGACATCAGCCTTGCTTG	*recA* qPCR
RTcipC F	TACTGGCGTCGTATCAGTGC	*cipC* qPCR
RTcipC R	TGTCCGCATCCTGAGTGTAA	*cipC* qPCR
RTcel48F F	AACAAACCGGCTACATACGC	*cel48F* qPCR
RTcel48F R	GGTTCCATCAGCTCTTGCTC	*cel48F* qPCR
RTcel9E F	ACCTGGACCGTAATGAATGC	*cel9E* qPCR
RTcel9E R	TCATGAGCTTTGTGGTGAGC	*cel9E* qPCR
RTcel8C F	GGATACGGTTTGCTGCTTTC	cel8C qPCR
RTcel8C R	AGCAAACACAAGGGATACCG	cel8C qPCR
RTorfX F	AAGCAGCAACAGTGGTAAGG	*orfX* qPCR
RTorfX R	AATGCACCGGAAGTACCTTG	*orfX* qPCR

Table S5.1 Primers used to study carbon catabolite regulation.

Primer name	Sequence	Note
hprK_RNAiF	CGAGCTCGGTACCCCGATTATCTGTATTCTA	Amplify the partial *hprK* gene for

	TCTGT	pRNA-hprK construction
hprK_RNAiR	ATATCGATCGATACGTTATAATATAATAAG ATATG	Amplify the partial *hprK* gene for pRNA-hprK construction
0806GRF	GTTGGTTACAATAAACTGCCCTGCGTTTTA GAGCTAGAAATAGCAAGT	Modify the gRNA to target Ccel_0806 (*crh*)
0806GRR	CTAAAACGCAGGGCAGTTTATTGTAACCAA CTTAATTTTAACTTTAAAAAATAAATTTGT	Modify the gRNA to target Ccel_0806 (*crh*)
0806LF	gaattttattatggtacccgggTTGGCGGAGACCTCGTT TCGG	Amplify the left homologous arm
0806LR	CAGCTTTGGAATCCAACCCTGCAtaagcttGGG CAGTTTATTGTAACTTTAGTAG	Amplify the left homologous arm
0806RF	GTTACAATAAACTGCCCaagcttaTGCAGGGTT GGATTCCAAAGCTGCGGCATTGC	Amplify the right homologous arm
0806RR	ctcatcaatttgttgcaacgagCGTCATACTGCCCGCCT AAA	Amplify the right homologous arm
ID0806R	ctgctctgccgcagtttccatttgt	Identify the Δ*crh* mutant by PCR amplification
1005GRF	GTTGAAGGATGTTGCCAGCAAGTCGTTTTA GAGCTAGAAATAGCAAGT	Modify the gRNA to target Ccel_1005 (*ccpA*)
1005GRR	CTAAAACGACTTGCTGGCAACATCCTTCAA CTTAATTTTAACTTTAAAAAATAAATTTGT	Modify the gRNA to target Ccel_1005 (*ccpA*)
1005LF	gaattttattatggtacccgggacggagactatgaaggcgg	Amplify the left homologous arm
1005LR	CCTGACtgaattcTTGCTGGCAACATCCTTTATT GTTACAG	Amplify the left homologous arm
1005RF	AAGGATGTTGCCAGCAAgaattcaGTCAGGTCT CTCAATTGCAAC	Amplify the right homologous arm
1005RR	ctcatcaatttgttgcaacgagccacggctgtctccatttctga	Amplify the right homologous arm
ID1005R	CCGGTGTTCCTATTGCCAGT	Identify the Δ*ccpA* mutant by PCR amplification
2999GRF	GTTGTAATATAGGAGTTATCATCCGTTTTAG AGCTAGAAATAGCAAGT	Modify the gRNA to target Ccel_2999 (*lfpC2*)
2999GRR	CTAAAACggatgataactcctatattaCAACTTAATTTT AACTTTAAAAAATAAATTTGT	Modify the gRNA to target Ccel_2999 (*lfpC2*)
2999LF	gaattttattatggtacccgggTCAGGCGGATATGCTGC TACG	Amplify the left homologous arm
2999LR	GAAAATGTTCCGGGATGAATTCTGATAACT CCTATATTATATGACAAACCATC	Amplify the left homologous arm
2999RF	AGGAGTTATCAGAATTCATCCCGGAACATT TTCTGGGGGATGAC	Amplify the right homologous arm
2999RR	CtcatcaatttgttgcaacgagCCGTCCGATTCAAGAGG CTATAC	Amplify the right homologous arm
ID2999R	ccttatggctcaccccggcttagcc	Identify the Δ*lfpC2* mutant by PCR amplification
3000GRF	GTTGCTAGCACTAAGAGGCATGCCGTTTTA GAGCTAGAAATAGCAAGT	Modify the gRNA to target Ccel_3000 (*lfpC3*)
3000GRR	CTAAAACggcatgcctcttagtgctagCAACTTAATTT TAACTTTAAAAAATAAATTTGT	Modify the gRNA to target Ccel_3000 (*lfpC3*)
3000LF	gaattttattatggtacccgggGGTGTATGACCTGTATC GCTGGGTC	Amplify the left homologous arm
3000LR	CTTGTACTTTCGCTTATACCAGGCTGAATTC ATGCCTCTTAGTGCTAGAGATACTG	Amplify the left homologous arm
3000RF	GGCATGAATTCAGCCTGGTATAAGCGAAAG TACAAGAAAAGTAATTGAGCA	Amplify the right homologous arm
3000RR	ctcatcaatttgttgcaacgagTCAGCACGAGCCATGAA	Amplify the right homologous

	ACGCCA	arm
ID3000F	cagtgaacaagatcgggactctcgct	Identify the Δ*lfpC3* mutant by PCR amplification
c_crhF	ataagcgcgccccggatcgagatagCCAAGCGGTTCCTATACTGC	Amplify the *crh* gene for complementation
c_crhR	ttgtactggtgcattcctgcaggccACCCTTTTAAAGCGGGAGCAG	Amplify the *crh* gene for complementation
c_ccpAF	ataagcgcgccccggatcgagatagATTAACCTGGTCCGACAGCG	Amplify the *ccpA* gene for complementation
c/ov_ccpAR	ttgtactggtgcattcctgcaggccTTGCTCAACATAGCCCGGAG	Amplify the *ccpA* gene for complementation
c_lfpC23F	ataagcgcgccccggatcgagatagAGCCGAGAGGTTTATCTTGCc	Amplify the *lfpC2* or *lfpC3* gene for complementation
c_lfpC23R	ttgtactggtgcattcctgcaggccTGTCCATTGTGGACCCTCCT	Amplify the *lfpC2* or *lfpC3* gene for complementation
pErR	CTATCTCGATCCGGGGCGCGCT	Amplify the plasmid backbone for complementation
pErF	GGCCTGCAGGAATGCACCAGT	Amplify the plasmid backbone for complementation
0413F1	TGTGCTATTGGAACGGGAGG	qPCR
0413R1	GGCTCCCAAACCCATCATGT	qPCR
1767F1	AAGGACTGCCAGAGCAAAGG	qPCR
1767R1	TCCGGGAAGTGAAAGACCAA	qPCR
1982F1	AGGAAGACGGGAACTCGGAT	qPCR
1982R1	TCTGCCGTAAACACGGTCAA	qPCR
2660F2	TCCGCATTCAAAGGGCAGAT	qPCR
2660R2	TTGCCTTTGGCCTCCAGAAA	qPCR
0967F2	TCGACTATGGTTTCGTGCTGT	qPCR
0967R2	GAGATATGCTGGAGCTTGCC	qPCR
3115F1	TCAATCTGGTGTACCTATTGGG	qPCR
3115R1	ACTGCAATCCAAAACTTTAGCCA	qPCR
1076F1	CGGAGAAAATGGGGCAGGAA	qPCR
1076R1	CCATACCCAGGCCGCAATTA	qPCR
1178F2	AGCTTGAGAAAGTGCCTGCT	qPCR
1178R2	GCCGCTTCCATACTGCTTTG	qPCR
2587F2	AGCCGAGGGAATGACAATGG	qPCR
2587R2	TTCAATGGGTGGCGCATCTT	qPCR
1074F1	TTGAGGCAGTATTGGCGGTT	qPCR
1074R1	TCCCATTCACCATTCCATGTGT	qPCR
RTrecAF	GCAAAGAAACTTGGGGTTGA	qPCR
RTrecAR	TGAGACATCAGCCTTGCTTG	qPCR

Table S5.2 Bacterial strains and plasmids constructed to study carbon catabolite regulation.

Strain or plasmid	Relevant characteristics	Reference
Escherichia coli Top10	F⁻ *mcr*A Δ(*mrr-hsd*RMS- *mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1	Invitrogen
C. cellulolyticum H10	Wildtype	Petitdemnge, *et al.*,1984
pRNAi-hprK strain	H10 wildtype with pRNAi-hprK plasmid (ErR)	This study
pRNAi strain	H10 wildtype with pRNAi empty plasmid (ErR)	This study
H10 control strain	ΔmspI; it allows the transformation of unmethylated DNA	(Xu *et al.*, 2015)
Δ*crh*	Δcrh in the H10 control background	This study
Δ*ccpA*	ΔccpA in the H10 control background	This study
Δ*lfpC2*	ΔlfpC2 in the H10 control background	This study
Δ*lfpC3*	ΔlfpC3 in the H10 control background	This study
Δ*lfpC2&3*	ΔlfpC2 and ΔlfpC3 in the H10 control background	This study
crh/com	Δ*crh* mutant with pEr-Pm::crh plasmid	This study
ccpA/com	Δ*ccpA* mutant with pEr-Pm::ccpA plasmid	This study
lfpC3/com	Δ*lfpC3* mutant with pEr-Pm::lfpC3 plasmid	This study
ck/com	H10 control strain with pEr-Fd::empty plasmid	This study
Plasmids		
pRNAi	CMP^r in E. coli; TMP^r in H10; Fd::empty cassette	Xu *et al.*, 2015
pRNAi-hprK	pRNAi derivative with a Fd::hprK asRNA cassette	This study
pCas9n-pyrF	CMP^r in E. coli; TMP^r in H10; a Fd::cas9n cassette; a P4::pyrF gRNA cassette	(Xu *et al.*, 2015)
pCas9n-crh-donor	pCas9n derivative with a P4::crh gRNA cassette and a homologous donor template	This study
pCas9n-ccpA-donor	pCas9n derivative with a P4::ccpA gRNA cassette and a homologous donor	This study
pCas9n-lfpC2-donor	pCas9n derivative with a P4::lfpC2 gRNA cassette and a homologous donor	This study
pCas9n-lfpC3-donor	pCas9n derivative with a P4::lfpC3 gRNA cassette and a homologous donor	This study
pCas9n-lfpC2&3-donor2&3	pCas9n derivative with both P4::lfpC2 gRNA cassette and P4::lfpC3 gRNA cassette, and two homologous donors	This study
pEr-Fd::empty	Kan^r in E. coli; Er^r in H10; Fd::empty cassette	This study
pEr-Pm::crh	pEr derivative with a native promoter-driven crh gene	This study
pEr-Pm::ccpA	pEr derivative with a native promoter-driven ccpA gene	This study
pEr-Pm::lfpC3	pEr derivative with a native promoter-driven lfpC3 gene	This study

Table S5.3 List of genes and their associated gene ontology (GO) terms in mutants at the exponential phase.

Strain	GO term (ID)	Locus	Log2 fold-change	Description
Δ***crh***	signal transduction (0007165)	Ccel_2390	1.27	amino acid adenylation domain protein
		Ccel_2270	1.34	two component transcriptional regulator, LytTR family
		Ccel_1227	1.96	histidine kinase
		Ccel_2526	2.20	methyl-accepting chemotaxis sensory transducer
		Ccel_2425	-1.20	response regulator receiver protein
		Ccel_0219	-1.32	methyl-accepting chemotaxis sensory transducer
		Ccel_2269	1.39	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_1894	1.59	sporulation transcriptional activator Spo0A
		Ccel_1982	1.86	two component transcriptional regulator, AraC family
		Ccel_3234	2.13	methyl-accepting chemotaxis sensory transducer
	regulation of sporulation (0042173)	Ccel_1894	1.59	sporulation transcriptional activator Spo0A
		Ccel_2425	-1.20	response regulator receiver protein
	DNA methylation (0006306)	Ccel_2841	-1.30	DNA-cytosine methyltransferase
		Ccel_0834	1.18	DNA adenine methylase
		Ccel_2827	-1.23	DNA-cytosine methyltransferase
	defense response (0006952)	Ccel_3120	1.34	CRISPR-associated protein, Csn1 family
		Ccel_2841	-1.30	DNA-cytosine methyltransferase
		Ccel_2726	1.19	type I site-specific deoxyribonuclease, HsdR family
	organonitrogen compound biosynthetic process (1901566)	Ccel_2098	1.27	arginine biosynthesis bifunctional protein ArgJ
		Ccel_3332	2.73	O-acetylhomoserine/O-acetylserine sulfhydrylase
	calcium ion binding (0005509)	Ccel_2425	-1.20	response regulator receiver protein
		Ccel_1894	1.59	sporulation transcriptional activator Spo0A
		Ccel_1543	1.37	cellulosome anchoring protein cohesin region
	DNA-methyltransferase activity (0009008)	Ccel_2827	-1.23	DNA-cytosine methyltransferase
		Ccel_2841	-1.30	DNA-cytosine methyltransferase
		Ccel_0834	1.18	DNA adenine methylase
Δ***lfpC2***	cellular nitrogen compound metabolic process (0034641)	Ccel_2761	-1.92	Domain containing protein
		Ccel_1490	-1.96	RNA polymerase subunit sigma24
	primary metabolic	Ccel_2516	1.19	AMP-dependent synthetase and ligase
		Ccel_1490	-1.96	hypothetical protein

	process (0044238)	Ccel_1972	1.15	glycoside hydrolase family 43
		Ccel_2761	-1.44	Domain containing protein
		Ccel_0994	-1.35	putative sensor with HAMP domain
	Signal transducer activity (0004871)	Ccel_0994	-1.35	putative sensor with HAMP domain
		Ccel_2125	-1.70	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_0219	-1.18	methyl-accepting chemotaxis sensory transducer
		Ccel_2526	1.74	methyl-accepting chemotaxis sensory transducer
	protein binding (0005515)	Ccel_2526	1.74	methyl-accepting chemotaxis sensory transducer
		Ccel_0390	-1.11	Tetratricopeptide TPR_2 repeat
		Ccel_1493	-1.49	Ankyrin
		Ccel_2974	1.28	hypothetical protein
Δ*lfpC 3*	fatty acid metabolic process (0006631)	Ccel_0854	-1.23	PfaD family protein
		Ccel_2516	1.18	AMP-dependent synthetase and ligase
		Ccel_2887	-1.21	3-Oxoacyl-(acyl-carrier-protein (ACP)) synthase III domain protein
	protein binding (0005515)	Ccel_3209	-1.35	Tetratricopeptide TPR_2 repeat protein
		Ccel_2526	1.62	methyl-accepting chemotaxis sensory transducer
		Ccel_0390	-1.34	Tetratricopeptide TPR_2 repeat
		Ccel_2974	1.40	hypothetical protein
	organic cyclic compound binding (0097159)	Ccel_3075	-1.12	transcriptional regulator, XRE family
		Ccel_1490	-1.71	RNA polymerase subunit sigma24
		Ccel_2587	-1.33	ABC transporter related
		Ccel_0164	-1.22	transcriptional regulator, AbrB family
Δ*lfpC 2&3*	signal transduction (0007165)	Ccel_0994	-1.39	putative sensor with HAMP domain
		Ccel_2526	1.15	methyl-accepting chemotaxis sensory transducer
		Ccel_0944	-1.06	two component transcriptional regulator, AraC family
		Ccel_0219	-1.55	methyl-accepting chemotaxis sensory transducer
		Ccel_2886	1.45	methyl-accepting chemotaxis sensory transducer
		Ccel_2125	-1.91	signal transduction histidine kinase regulating citrate/malate metabolism
	nitrogen compound metabolic process (0006807)	Ccel_0267	1.76	ATP synthase F0, C subunit
		Ccel_0834	1.23	DNA adenine methylase
		Ccel_0944	-1.06	two component transcriptional regulator, AraC family
		Ccel_1490	-1.83	RNA polymerase subunit sigma24
		Ccel_0383	-1.54	histidinol-phosphate aminotransferase
		Ccel_2761	-1.52	Domain containing protein
	signal transducer activity (0004871)	Ccel_0994	-1.39	putative sensor with HAMP domain
		Ccel_2526	1.15	methyl-accepting chemotaxis sensory transducer
		Ccel_0219	-1.55	methyl-accepting chemotaxis sensory transducer

		Ccel_2886	1.45	methyl-accepting chemotaxis sensory transducer
		Ccel_2125	-1.91	signal transduction histidine kinase regulating citrate/malate metabolism

Table S5.4 List of genes and their associated gene ontology (GO) terms in mutants at the stationary phase.

Strain	GO term (ID)	Locus	log2 fold-change	Description
Δ***crh***	response to stimulus (0050896)	Ccel_2270	1.63	two component transcriptional regulator, LytTR family
		Ccel_3120	1.16	CRISPR-associated protein, Csn1 family
		Ccel_3265	1.76	excinuclease ABC, A subunit
		Ccel_2552	1.16	hypothetical protein
		Ccel_2114	1.41	putative sensor with HAMP domain
		Ccel_1982	1.65	two component transcriptional regulator, AraC family
		Ccel_2313	-1.16	methyl-accepting chemotaxis sensory transducer
		Ccel_1983	1.14	putative sensor with HAMP domain
		Ccel_0944	-1.25	two component transcriptional regulator, AraC family
		Ccel_2657	1.43	histidine kinase
		Ccel_1614	-1.52	oxidoreductase/nitrogenase component 1
		Ccel_1227	2.17	histidine kinase
		Ccel_2726	1.62	type I site-specific deoxyribonuclease, HsdR family
		Ccel_0676	1.11	recA protein
		Ccel_2808	-1.37	DNA polymerase beta domain protein region
		Ccel_2100	1.82	methyl-accepting chemotaxis sensory transducer
		Ccel_2841	-1.30	DNA-cytosine methyltransferase
		Ccel_2269	1.62	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_3117	1.75	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_2469	1.21	DNA mismatch repair protein MutS domain protein
		Ccel_0219	-2.16	methyl-accepting chemotaxis sensory transducer
		Ccel_2526	2.37	methyl-accepting chemotaxis sensory transducer
		Ccel_0049	1.69	methyl-accepting chemotaxis sensory transducer
		Ccel_0324	-1.49	transcriptional repressor, CtsR
		Ccel_1797	1.08	chaperone protein DnaJ
		Ccel_1794	1.55	response regulator receiver protein
	cobalamin biosynthetic process (0009236)	Ccel_0645	1.28	cobalamin 5'-phosphate synthase
		Ccel_1285	1.59	Precorrin-8X methylmutase CbiC/CobH
		Ccel_1274	2.46	Tetrapyrrole biosynthesis, glutamyl-tRNA reductase-like protein
		Ccel_1271	1.09	precorrin-2 C20-methyltransferase

		Ccel_1281	1.23	cobalamin (vitamin B12) biosynthesis CbiX protein
		Ccel_1270	1.50	cobalamin biosynthesis protein CbiD
	cellular component organization or biogenesis (0071840)	Ccel_3094	-2.78	SpoVG family protein
		Ccel_0305	-1.29	ribosomal protein L33
		Ccel_0712	-1.06	16S rRNA processing protein RimM
		Ccel_0081	1.29	ribosomal protein L9
		Ccel_3120	1.16	CRISPR-associated protein, Csn1 family
	heme biosynthetic process (0006783)	Ccel_1274	2.46	Tetrapyrrole biosynthesis, glutamyl-tRNA reductase-like protein
		Ccel_1280	1.23	glutamate-1-semialdehyde-2,1-aminomutase
		Ccel_1281	1.23	cobalamin (vitamin B12) biosynthesis CbiX protein
	cell communication (0007154)	Ccel_2270	1.63	two component transcriptional regulator, LytTR family
		Ccel_2100	1.82	methyl-accepting chemotaxis sensory transducer
		Ccel_2114	1.41	putative sensor with HAMP domain
		Ccel_1982	1.65	two component transcriptional regulator, AraC family
		Ccel_2313	-1.16	methyl-accepting chemotaxis sensory transducer
		Ccel_1983	1.14	putative sensor with HAMP domain
		Ccel_2269	1.62	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_3117	1.75	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_0219	-2.16	methyl-accepting chemotaxis sensory transducer
		Ccel_0944	-1.25	two component transcriptional regulator, AraC family
		Ccel_2657	1.43	histidine kinase
		Ccel_2526	2.37	methyl-accepting chemotaxis sensory transducer
		Ccel_1227	2.17	histidine kinase
		Ccel_0049	1.69	methyl-accepting chemotaxis sensory transducer
		Ccel_0676	1.11	recA protein
		Ccel_1794	1.55	response regulator receiver protein
	signal transducer activity (0004871)	Ccel_1227	2.17	histidine kinase
		Ccel_0049	1.69	methyl-accepting chemotaxis sensory transducer
		Ccel_2100	1.82	methyl-accepting chemotaxis sensory transducer
		Ccel_2114	1.41	putative sensor with HAMP domain
		Ccel_2313	-1.16	methyl-accepting chemotaxis sensory transducer
		Ccel_1983	1.14	putative sensor with HAMP domain
		Ccel_2269	1.62	signal transduction histidine kinase regulating citrate/malate metabolism

		Ccel_3117	1.75	signal transduction histidine kinase regulating citrate/malate metabolism
		Ccel_0219	-2.16	methyl-accepting chemotaxis sensory transducer
		Ccel_2657	1.43	histidine kinase
		Ccel_2526	2.37	methyl-accepting chemotaxis sensory transducer
Δ***lfpC2 &3***	protoporphyrinogen IX biosynthetic process (0006782)	Ccel_1274	1.44	Tetrapyrrole biosynthesis, glutamyl-tRNA reductase-like protein
		Ccel_1280	1.31	glutamate-1-semialdehyde-2,1-aminomutase
	chlorophyll metabolic process (0015994)	Ccel_1285	1.33	Precorrin-8X methylmutase CbiC/CobH
		Ccel_1272	-1.19	precorrin-4 C11-methyltransferase
		Ccel_1280	1.31	glutamate-1-semialdehyde-2,1-aminomutase
	precorrin-2 dehydrogenase activity (0043115)	Ccel_1274	1.44	Tetrapyrrole biosynthesis, glutamyl-tRNA reductase-like protein
		Ccel_1272	-1.19	precorrin-4 C11-methyltransferase
	Binding (0005488)	Ccel_2250	-1.60	two component transcriptional regulator, winged helix family
		Ccel_3075	1.26	transcriptional regulator, XRE family
		Ccel_1708	1.72	ribosomal protein S15
		Ccel_2901	-1.54	cold-shock DNA-binding domain protein
		Ccel_2945	1.42	transcriptional modulator of MazE/toxin, MazF
		Ccel_0887	-1.28	ABC transporter related
		Ccel_0534	-1.25	Radical SAM domain protein
		Ccel_1269	-1.23	cobalt ABC transporter, ATPase subunit
		Ccel_0685	1.28	acyl carrier protein
		Ccel_1774	-1.68	recombination helicase AddA
		Ccel_2609	-1.35	transcriptional regulator, GntR family
		Ccel_1274	1.44	Tetrapyrrole biosynthesis, glutamyl-tRNA reductase-like protein
		Ccel_0680	-1.54	regulatory protein DeoR
		Ccel_1280	1.31	glutamate-1-semialdehyde-2,1-aminomutase

Table S5.5 List of selected DEGs related to two-component system (TCS), transporter and transcriptional regulators (TF).

Strain	Locus	Annotation	Log2 fold change	TCS	TF	Transporter
Δ*crh*	Ccel_0164	transcriptional regulator, AbrB family	-2.88		Y	
	Ccel_0410	transcriptional regulator, MarR family	1.32		Y	
	Ccel_0502	hypothetical protein	-1.24	Y		
	Ccel_0803	RNA polymerase, sigma-24 subunit, ECF subfamily	1.16		Y	
	Ccel_0841	transcriptional regulator, AraC family	1.39		Y	
	Ccel_1223	periplasmic binding protein/LacI transcriptional regulator	-1.14			xylose binding ABC
	Ccel_1227	histidine kinase	1.96	Y		
	Ccel_1631	polar amino acid ABC transporter, inner membrane subunit	-1.14			amino acid membrane ABC
	Ccel_1894	sporulation transcriptional activator Spo0A	1.59	Y	Y	
	Ccel_1982	two component transcriptional regulator, AraC family	1.86	Y	Y	
	Ccel_1987	putative solute-binding component of ABC transporter	2.65			xylose binding ABC
	Ccel_2102	phosphate ABC transporter, inner membrane subunit PstA	1.34			phosphate membrane ABC
	Ccel_2145	histidine kinase	1.67	Y		
	Ccel_2269	signal transduction histidine kinase regulating citrate/malate metabolism	1.39	Y		
	Ccel_2270	two component transcriptional regulator, LytTR family	1.34	Y	Y	
	Ccel_2425	response regulator receiver protein	-1.20	Y	Y	
	Ccel_2465	ATPase, P-type (transporting), HAD superfamily, subfamily IC	1.45			Ca2+/Mg2+ P-ATPase
	Ccel_2522	ABC transporter related	1.36			daunorubicin binding ABC
	Ccel_2587	ABC transporter related	-1.63			amino acid binding ABC
	Ccel_2686	ABC transporter related	-1.35			xylose ABC
	Ccel_3075	transcriptional regulator, XRE family	-2.04		Y	
	Ccel_3464	transcriptional regulator, LacI family	1.27		Y	
Δ*lfpC2*	Ccel_0145	extracellular solute-binding protein family 1	-1.10			sugar binding ABC
	Ccel_0994	putative sensor with HAMP domain	-1.35	Y		
	Ccel_2125	signal transduction histidine kinase regulating citrate/malate metabolism	-1.70	Y		
	Ccel_2587	ABC transporter related	-1.40			amino acid binding ABC
	Ccel_2686	ABC transporter related	-1.54			xylose ABC
	Ccel_3075	transcriptional regulator, XRE family	-1.14		Y	
Δ*lfpC3*	Ccel_0164	transcriptional regulator, AbrB family	-1.22		Y	

	Ccel_0267	ATP synthase F0, C subunit	1.36			protons F-ATPase
	Ccel_0967	ABC-2 type transporter	1.61			daunorubicin membrane ABC
	Ccel_0994	putative sensor with HAMP domain	-1.24	Y		
	Ccel_2125	signal transduction histidine kinase regulating citrate/malate metabolism	-1.61	Y		
	Ccel_2587	ABC transporter related	-1.33			amino acid binding ABC
	Ccel_3075	transcriptional regulator, XRE family	-1.12		Y	
	Ccel_3333	Substrate-binding region of ABC-type glycine betaine transport system	-1.28			glycine betaine binding ABC
Δ*lfpC2 &3*	Ccel_0164	transcriptional regulator, AbrB family	-1.59		Y	
	Ccel_0267	ATP synthase F0, C subunit	1.76			protons F-ATPase
	Ccel_0944	two component transcriptional regulator, AraC family	-1.06	Y	Y	
	Ccel_0994	putative sensor with HAMP domain	-1.39	Y		
	Ccel_1177	ABC transporter related	-1.15			daunorubicin binding ABC
	Ccel_1178	ABC-type Na+ efflux pump permease component-like protein	-2.42			Na+ efflux pump permease ABC
	Ccel_1987	putative solute-binding component of ABC transporter	2.30			xylose binding ABC
	Ccel_2125	signal transduction histidine kinase regulating citrate/malate metabolism	-1.91	Y		
	Ccel_2528	drug resistance transporter, EmrB/QacA subfamily	-1.34			Major Facilitator Superfamily
	Ccel_2531	hypothetical protein	-1.54			
	Ccel_2587	ABC transporter related	-1.63			amino acid binding ABC
	Ccel_2793	transcriptional regulator, XRE family	-1.09		Y	
	Ccel_3075	transcriptional regulator, XRE family	-1.77		Y	

Appendix B: Supplementary Figures

Figure S2.1 Schemes for *S. pyogenes* CRISPR-Cas9 system and vectors.

Figure S2.2 Growth profiling of mutants under selection conditions.

Figure S2.3 NHEJ in *C. cellulolyticum*.

Figure S2.4 Generation and identification of Δ*mspI* mutant.

Figure S2.5 Pairwise Sequence Alignment between X21- and X22-containing fragments.

Figure S2.6 A reduced phylogenetic tree showing the distribution of *ku* gene.

Figure S3.1 RNA structures of targeted transcript regions of *pta* and *ack* predicted by RNAfold web server.

Figure S3.2 Cell growth and product measurement on 5 g/L cellobiose.

Figure S3.3 Comparison of cellulose utilization.

Figure S4.1 Diagram of intron insertion in *dpi* gene and strains identification.

Figure S4.2 Identification of *cel48F* and *cel9E* anti-sense mutants by PCR.

Figure S5.1 Alignment of HprK proteins.

Figure S5.2 Alignment of HPr and Crh proteins.

Figure S5.3 Growth profiling of wildtype *C. cellulolyticum* grown in the defined VM medium supplemented with different sugars.

Figure S5.4 Comparison of cellulose hydrolysis, released sugars, and fermentative products between pRNAi and pRNAi-hprK strains.

Figure S5.5 Identification of knockout mutants generated by the one-step Cas9 nickase-based genome editing tool.

Figure S5.6 Cell growth in the defined VM medium with 5 g/L cellobiose.

Figure S5.7 Measurement of residual cellulose after 15 days fermentation.

Figure S5.8 Correlation between qPCR results and microarray data.

Figure S5.9 Venn diagram of DEGs that exhibit between different mutants.

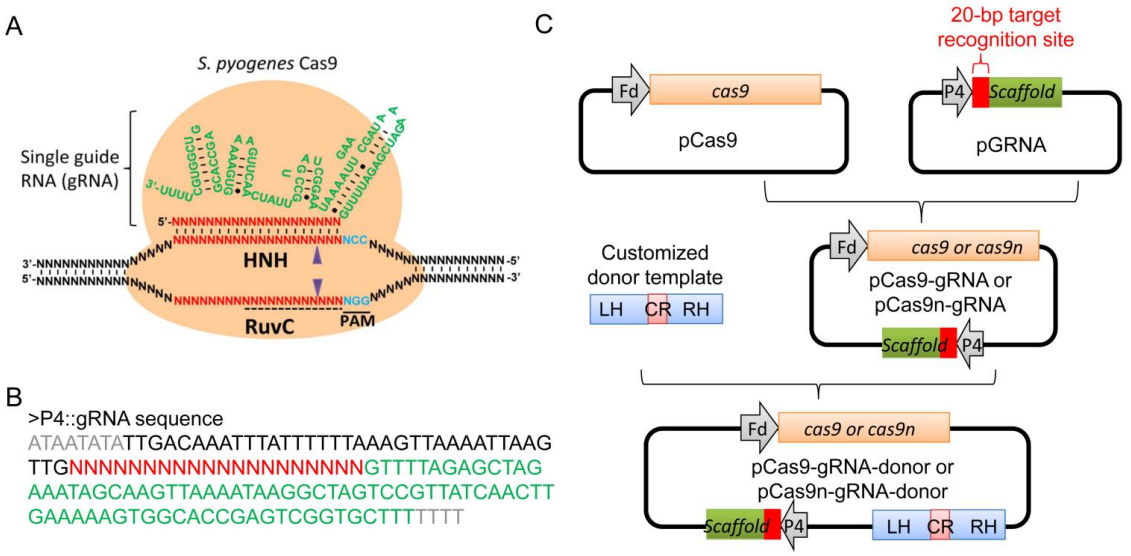


Figure S2.1 Schemes for *S. pyogenes* CRISPR-Cas9 system and vectors. (**A**) Specific targeting of *S. pyogenes* Cas9-gRNA complex. Any 23-bp target site contains a 20-bp target recognition sequence (red), which is also known as protospacer, and a NGG protospacer-adjacent motif (PAM) (light blue). Cas9 is guided by gRNA to bind the target site through base pairing and then cleaves the double strands of target DNA by two domains, HNH and RuvC, respectively (Jinek *et al.*, 2012). The cleavage site is 3-bp upstream of the PAM, indicated by purple triangles. The 12-bp PAM-proximal region labeled with dashed line is critical for targeting precision, so called seed region (Jinek *et al.*, 2012; Jiang *et al.*, 2013). In this study, Cas9 nickase (D10A) only has catalytic HNH domain to cut one strand (Jinek *et al.*, 2012). (**B**) P4::gRNA cassette consists of P4 promoter (black), 20-bp target recognition sequence (red) and gRNA scaffoldin (green). (**C**) pCas9 and pGRNA contain Fd::*cas9* and P4::*gRNA* constructs only expressing Cas9 and gRNA, respectively; however, pCas9-gRNA and pCas9n-gRNA containing both cassettes can simultaneously express both components. For each targeting the 20-bp target recognition site will be customized. The donor editing

template consisting of left homologous (LH) and right homologous (RH) arms and customized region (CR) between LH and RH, can be inserted into pCas9-gRNA and pCas9n-gRNA, generating all-in-one vectors, pCas9-gRNA-donor and pCas9n-gRNA-donor.

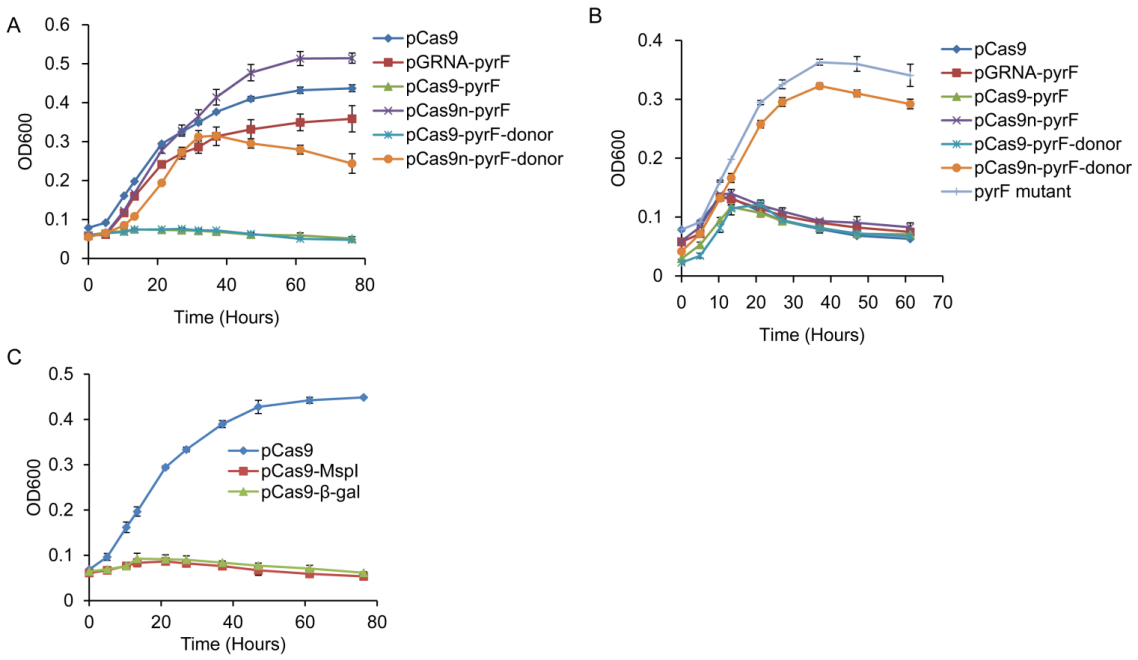


Figure S2.2 Growth profiling of mutants under selection conditions. Cells after transformed with corresponding constructs are grown under TMP antibiotic selection (A, C) and then under 5-FOA counter-selection (B). 5-FOA-resistant *pyrF* mutant generated by group II retrotransposition is used as positive control.

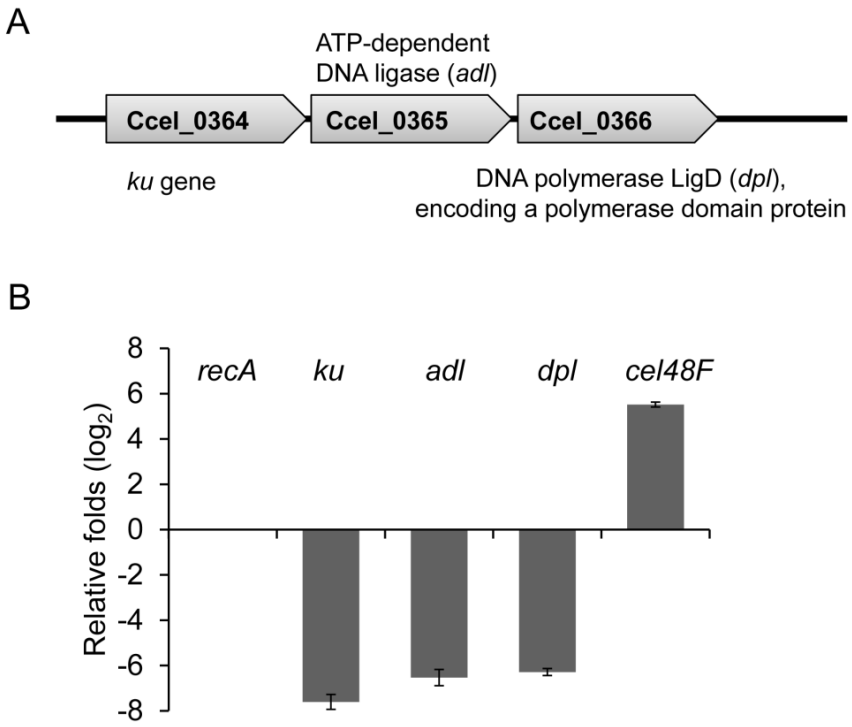


Figure S2.3 NHEJ in *C. cellulolyticum*. (**A**) A predicted operon in *C. cellulolyticum* contains genes encoding major NHEJ components including Ku, ATP-dependent DNA ligase and DNA polymerase LigD. (**B**) Quantitative PCR analysis of their relative transcript amounts using the *recA* as an internal calibrator and *cel48F* as a control.

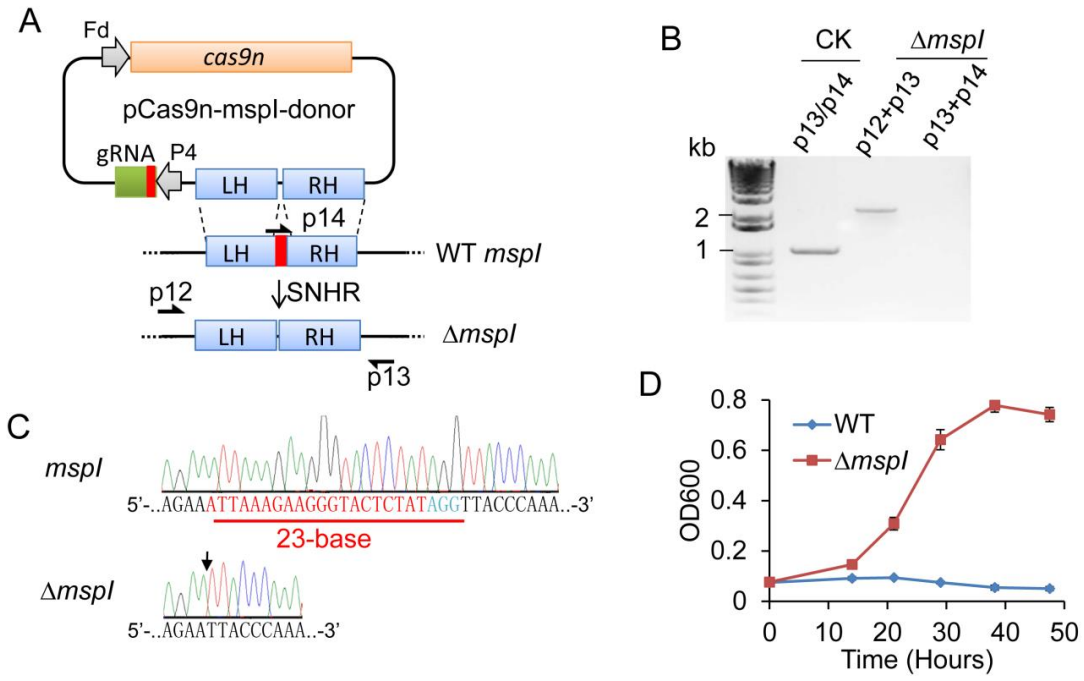


Figure S2.4 Generation and identification of Δ*mspI* mutant. (**A**) Schematic all-in-one vector for *mspI* disruption by SNHR. It consists of an Fd-driven *cas9n* gene, P4-driven gRNA targeting *mspI* gene (Ccel_2866) and donor template with 23-bp deletion flanked by 1-kb left homologous (LH) and right homologous (RH) arms. (**B**) PCR identification of Δ*mspI* mutants in TMP-selected population. Whole genomes from transformants of empty vector (CK) and pCas9n-mspI-donor (Δ*mspI*) are used as templates for PCR with primers as indicated and schematized in **A**. (**C**) Deletion of 23-bp target site in the *mspI* gene confirmed by DNA sequencing of the p12/13 PCR amplicon. Junction site is indicated by a downward black arrow. (**D**) Growth profile of transformants of WT and plasmid-cured Δ*mspI* mutant using non-methylated plasmids. Under antibiotic selection, only Δ*mspI* transformants grow out, suggesting MspI-dependent restriction-modification system is disrupted, which allows non-methylated plasmids for transformation.



Figure S2.5 Pairwise Sequence Alignment between X21- and X22-containing fragments. Colorized regions are 23-bp target sites. X21 and X22 differ by two bases in 5' region preceding the seed region so they have a very low specificity.

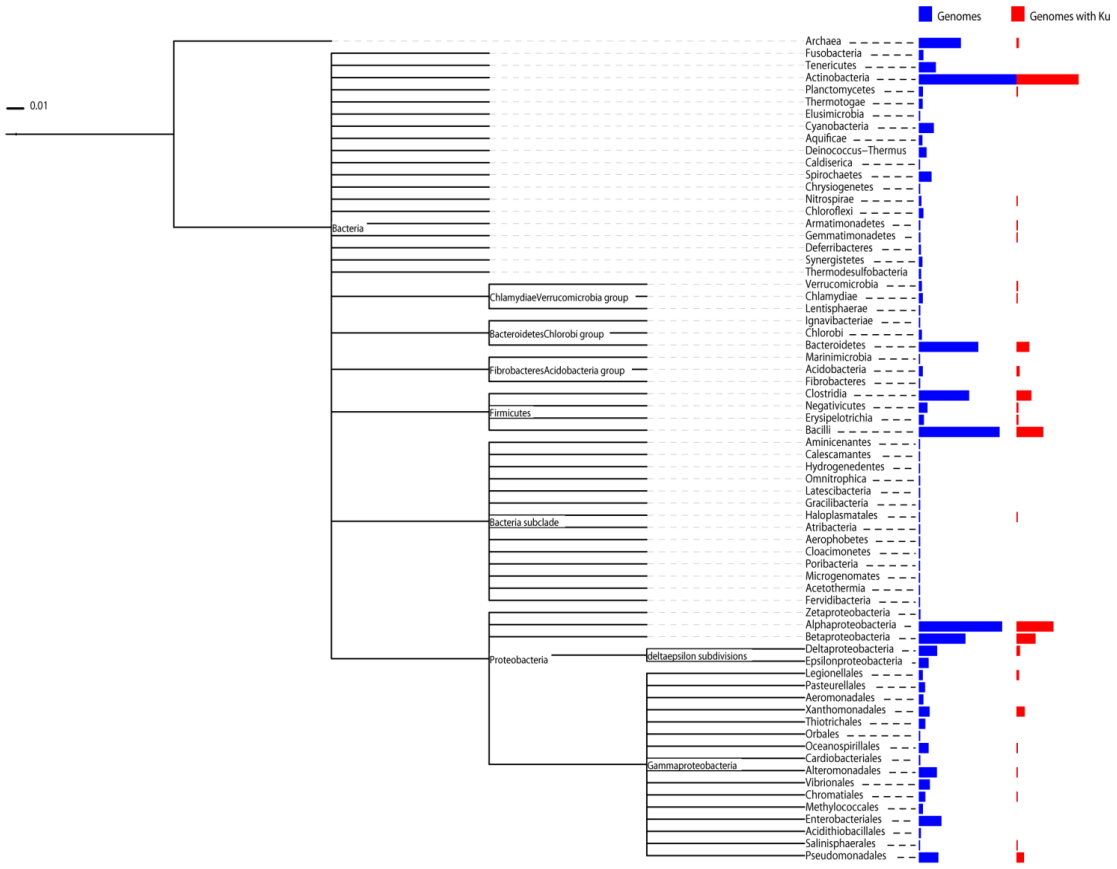


Figure S2.6 A reduced phylogenetic tree showing the distribution of *ku* gene. 4596 sequenced genomes including archaea and bacteria were analyzed. Bacterial and archaeal genomes encoding Ku homologs (COG1237) were identified in IMG (Markowitz *et al.*, 2012). The IMG taxon identifiers for 4596 sequenced genomes were extracted and converted to NCBI taxonomy identifiers (Federhen, 2012). The taxonomy identifier list was trimmed at the strain level based on whether or not the genome had at least one *ku* gene, and then this trimmed list was used for 16S rDNA tree construction using PhyloT. The resulting tree was reduced and visualized using iTOL (Letunic *et al.*, 2011). Green bars show the relative amount of sequenced genomes in each branch to the maximal one; red bars show the relative amount of sequenced genomes with *ku* encoding genes to the total number of genomes in that branch. There is no obvious

distribution pattern in the tree. Totally, *ku*-containing genomes just account for 27.5% (1264/4596).

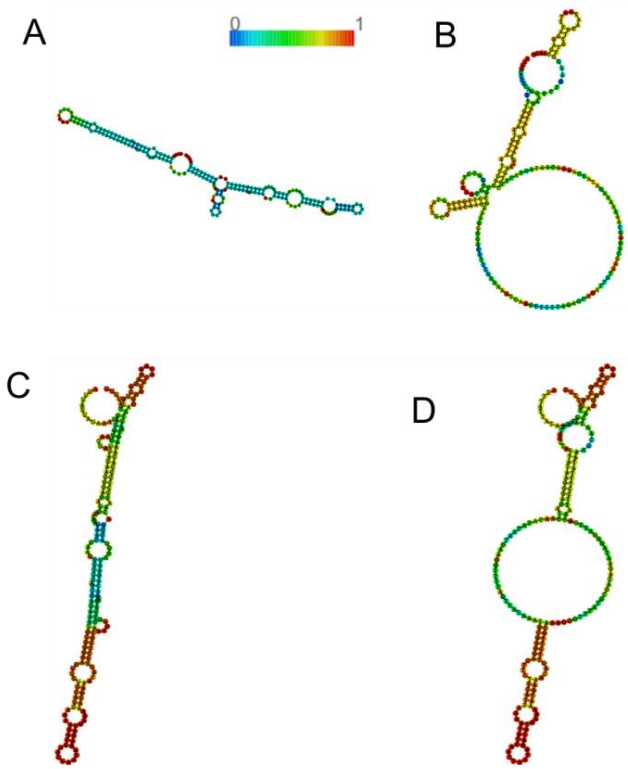


Figure S3.1 RNA structures of targeted transcript regions of *pta* (A, B) and *ack* (C, D) predicted by RNAfold web server. The color represents base-pair probabilities. A and C, secondary structure with minimal free energy; B and D, centroid secondary structure.

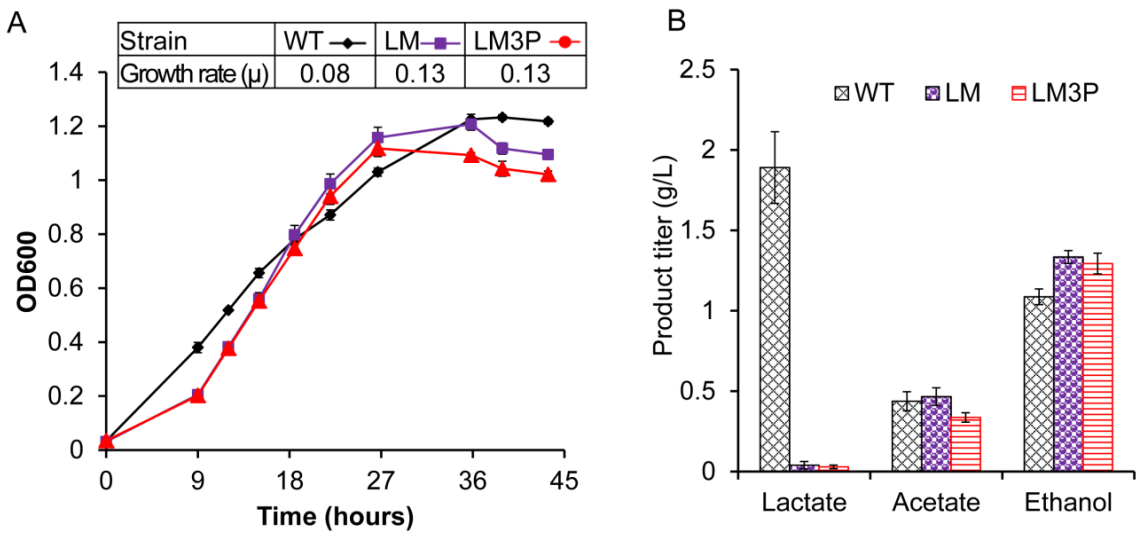


Figure S3.2 Cell growth and product measurement on 5 g/L cellobiose. (A) Cell growth was profiled with an insert table showing the growth rate of each strain under the tested condition. (B) Product titers in the fermentation broth produced from 5 g/L cellobiose. Values are presented as mean± standard deviation.

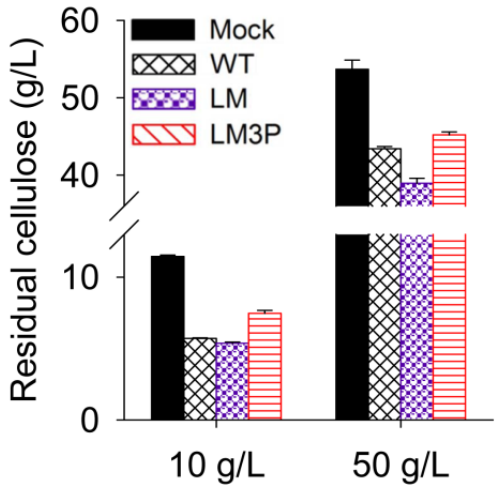


Figure S3.3 Comparison of cellulose utilization. All mutants were tested on 10 g/L and 50 g/L cellulose by reference to the mock without bacterial inoculation. Values are presented as mean± standard deviation.

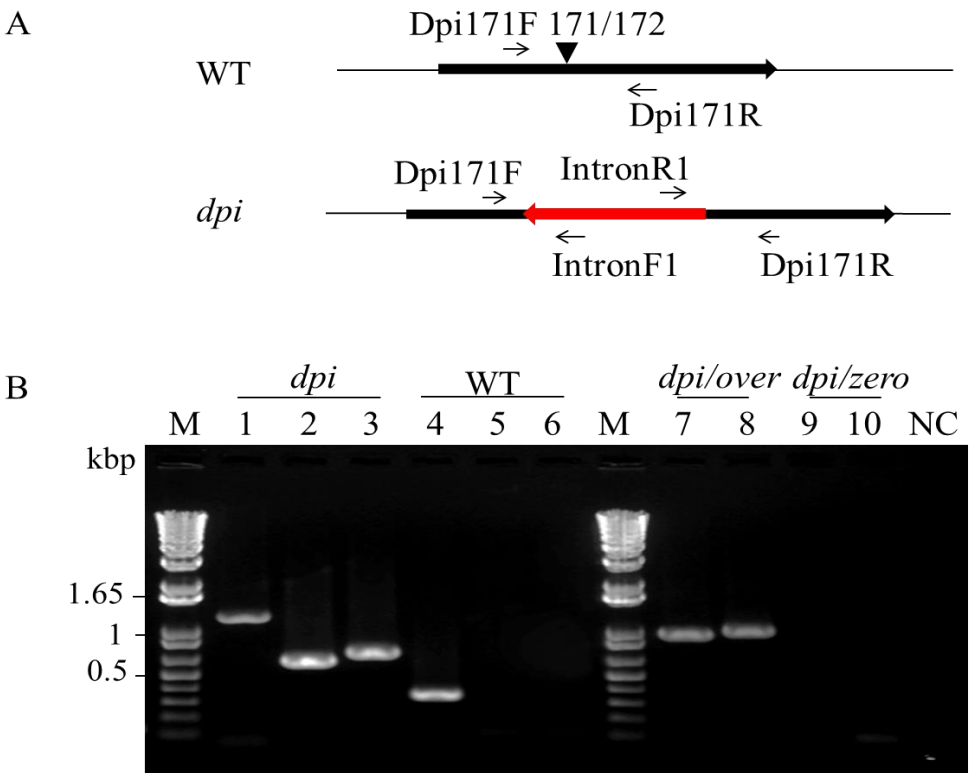


Figure S4.1 Diagram of intron insertion in *dpi* gene and strains identification. A. The group II intron (red arrow) potentially inserted into the *dpi* ORF (bold black arrow) at 171/172nt in the anti-sense direction. Small arrows indicated locations of four primers (Dpi171F, Dpi171R, IntronF1 and Intron R1) that were applied to identify the anticipated intron insertion. B. PCR identification. Primers used in each PCR reaction are as follows: Dpi171F-Dpi171R (lane 1 and 4); Dpi171F-IntroF1 (lane 2 and 3), IntronR1-Dpi171R (lane 5 and 6); pClostron3RBSF-Dpi overexpR (lane 7 and 9); pClostron3RBSF-pClostron3seqR (lane 8 and 10). NC indicates negative control without any templates in the PCR system.

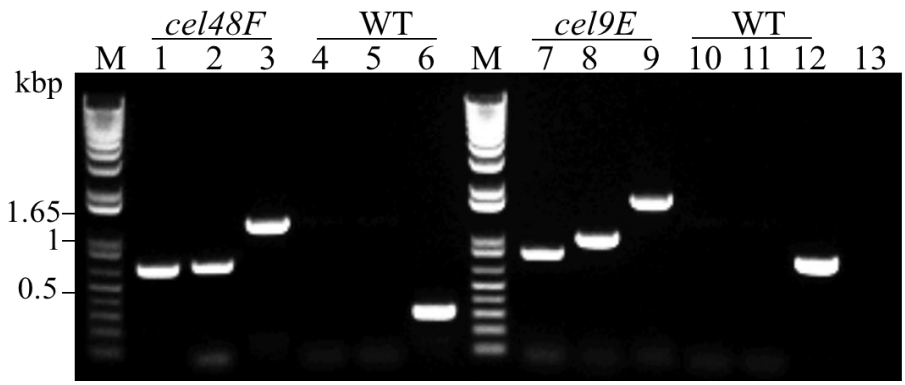


Figure S4.2 Identification of *cel48F* and *cel9E* anti-sense mutants by PCR. Primers for each PCR reaction are as follows: 1, Cel48FF- intronF1; 2, intronR1-Cel48FR; 3, Cel48FF-Cel48FR; 4, Cel48FF-intronF1; 5, intronR1-Cel48FR; 6, Cel48FF-Cel48FR; 7, Cel9EF-intronF1; 8, intronR1-Cel9ER; 9, Cel9EF-Cel9ER; 10, Cel9EF-intronF1; 11, intronR1-Cel9ER; 12, Cel9EF-Cel9ER.

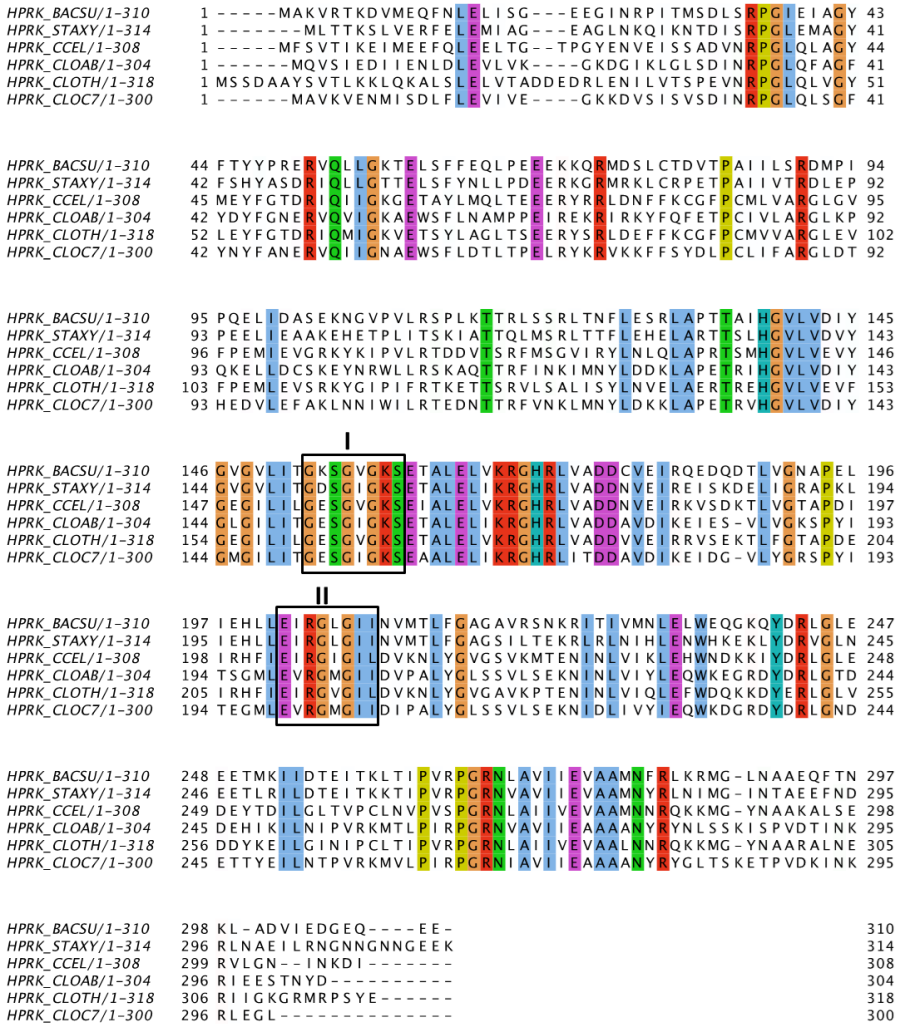


Figure S5.1 Alignment of HprK proteins. Colorized amino acids presented 100% conservation. The Walker A box nucleotide-binding motif (box I) and the signature sequence of HprK proteins (box II) were indicated respectively. Abbreviations are as follows: BACSU = *Bacillus subtilis* (O34483); STAXY = *Staphylococcus xylosus* (Q9S1H5); CCEL = *C. cellulolyticum* (B8I4X6); CLOAB = *Clostridium acetobutylicum* (Q97K32); CLOTH = *Clostridium thermocellum* (A3DBM2); CLOC7 = *Clostridium cellulovorans* (D9SKB9). The alignment was performed with ClustalX-2.1and then polished with clustalX module of Jalview 2.9.0b2.

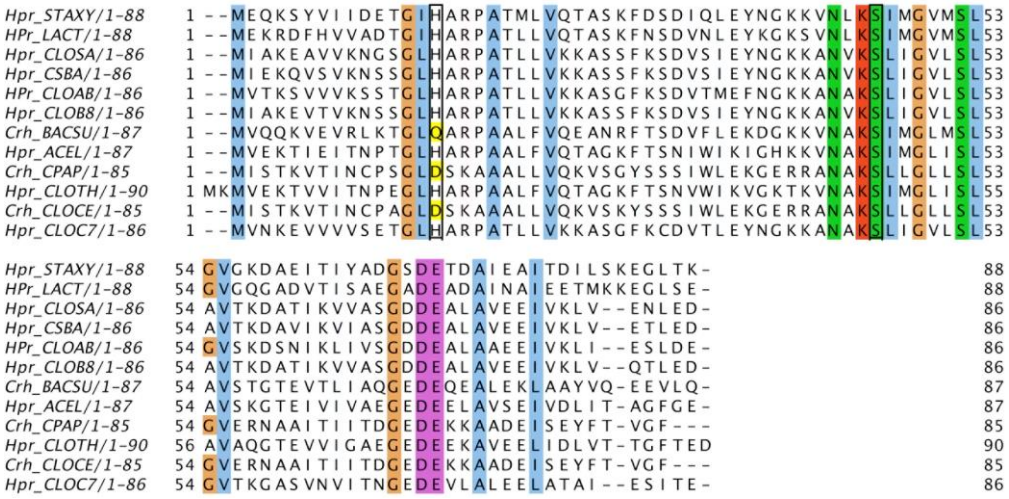


Figure S5.2 Alignment of HPr and Crh proteins. Colorized amino acids presented 100% conservation. Two highly conserved amino acids for phosphorylation in HPr (His15 and Ser46) were indicated in the boxes. However, only Ser46 was retained in Crh proteins. Abbreviations are as follows: STAXY = *Staphylococcus xylosus* (Q9EYQ9.1); LACT = *Lactobacillus (*WP_004265632*);* CLOSA = *Clostridium acetobutylicum* (WP_010965126); CSBA = *Clostridium saccharoperbutylacetonicum* (WP_015391452); CLOAB = *Clostridium acetobutylicum* (WP_010965126); CLOB8 = *Clostridium beijerinckii* (ABR33401); BACSU = *Bacillus subtilis* (NP_391354); ACEL = *Acetivibrio cellulolyticus* (WP_010251505); CPAP = *Clostridium papyrosolvens* (WP_004620766); CLOTH = *Clostridium thermocellum* (ABN51358); CCEL = *C. cellulolyticum* (WP_015924346); CLOC7 = *Clostridium cellulovorans* (WP_010076980). The alignment was performed with ClustalX-2.1and then polished with clustalX module of Jalview 2.9.0b2.

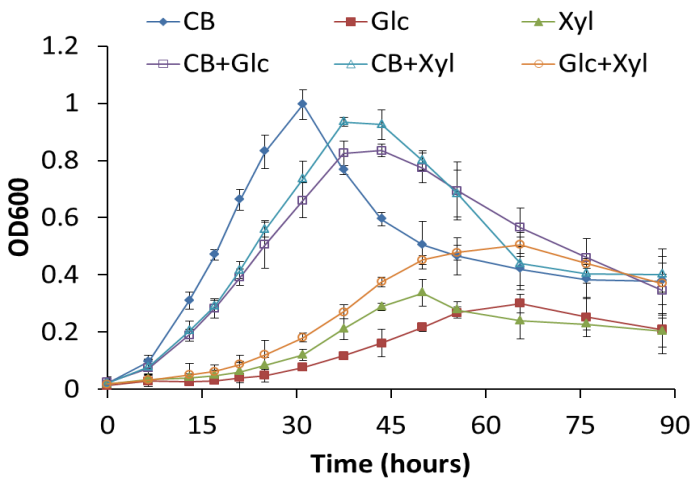


Figure S5.3 Growth profiling of wildtype *C. cellulolyticum* grown in the defined VM medium supplemented with different sugars. Irrespective of a sole sugar or dual sugars used as the carbon source, cellobiose (CB) was added at the concentration of 4 g/L but other sugars, D-glucose (Glc) and D-xylose (Xyl), were at 2 g/L. Culture optical density was measured and expressed as mean ± standard deviation (three biological replicates).

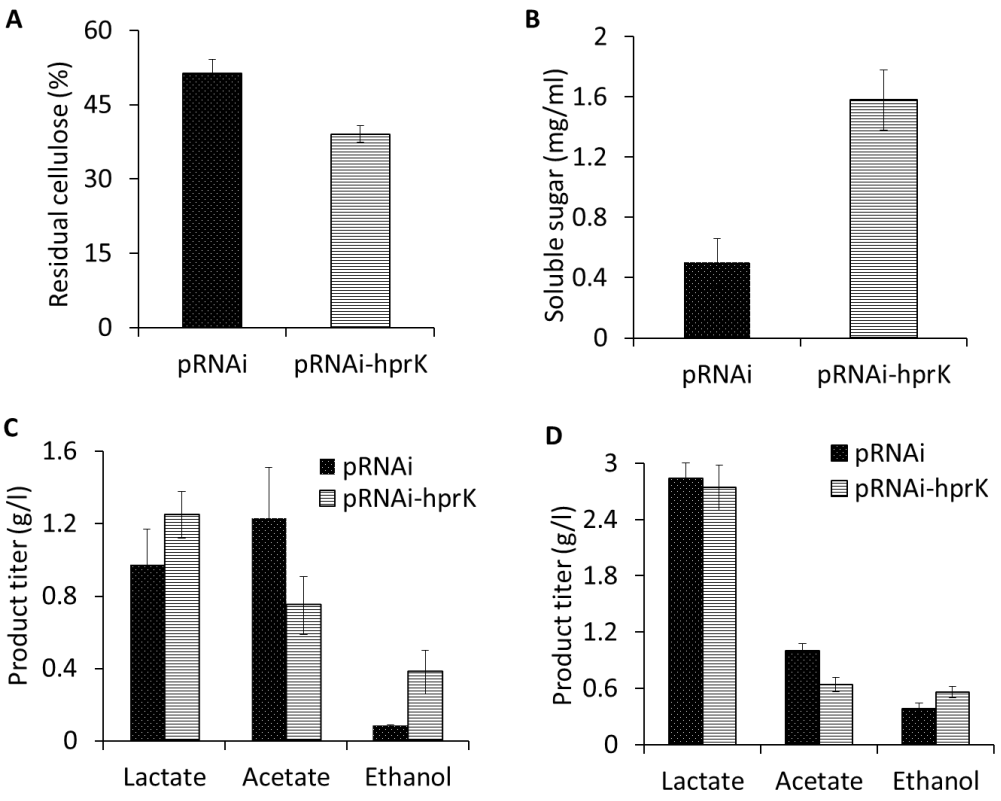


Figure S5.4 Comparison of cellulose hydrolysis, released sugars, and fermentative products between pRNAi and pRNAi-hprK strains. Cells were cultivated in the defined VM medium with 15 g/L Avicel cellulose (A, B and C), or 15 g/L cellobiose (D). The concentration of residual cellulose (A), released total sugars (B) and metabolites (C and D) in the endpoint fermentation broth were measured and displayed separately. The mean and standard deviation are shown for three biological replicates at each time point.

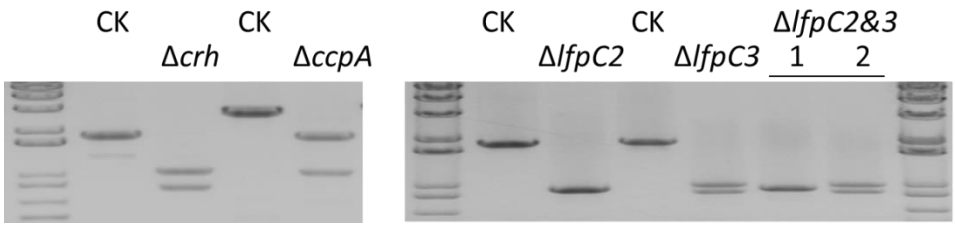


Figure S5.5 Identification of knockout mutants generated by the one-step Cas9 nickase-based genome editing tool. Gel images above shows the restriction enzyme-digested PCR products which were amplified from the chromosome of each mutant and the control strain (CK) with specific primer sets as follows: 0806LF/ID0806R for Δ*crh* by HindIII digestion; 1005LF/ID1005R for Δ*ccpA* by EcoRI digestion; 2999LF/ID2999R for Δ*lfpC2* by EcoRI digestion; ID3000F/3000RR for Δ*lfpC3* by EcoRI digestion; 2999LF/ID2999R for Δ*lfpC2&3* by EcoRI digestion (1); ID3000F/3000RR for Δ*lfpC2&3* by EcoRI digestion (2). Smaller bands or halved hands after enzyme digestion only showed up in the mutant, indicating successful gene editing.

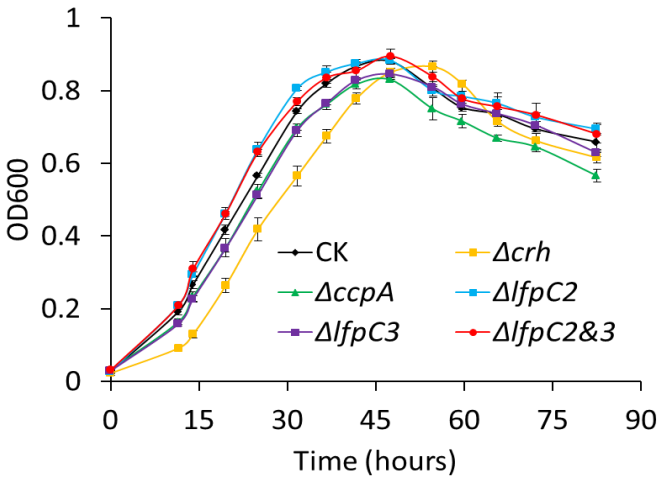


Figure S5.6 Cell growth in the defined VM medium with 5 g/L cellobiose. The mean and standard deviation are shown for three biological replicates at each time point.

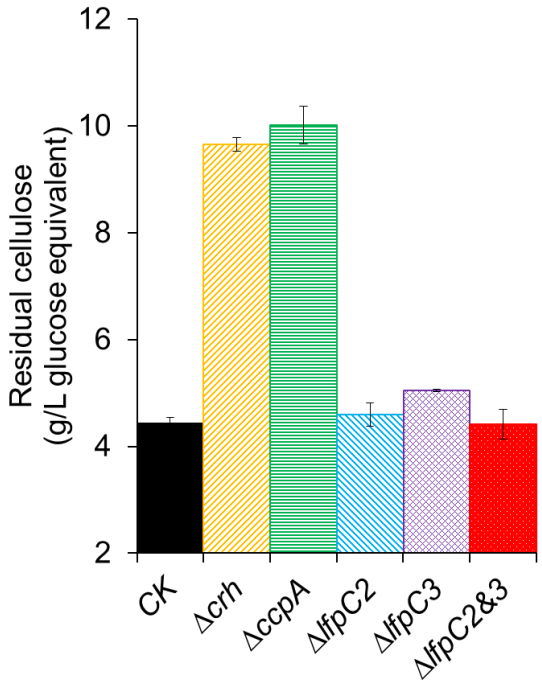


Figure S5.7 Measurement of residual cellulose after 15 days fermentation. The initial Avicel load was 10 g/L Avicel in the defined VM medium.

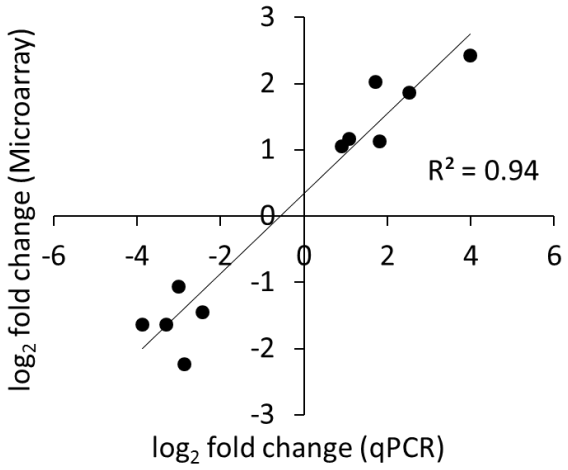


Figure S5.8 Correlation between qPCR results and microarray data. The fold changes in transcript levels of ten selected genes were log2 transformed before plotting. Genes and primers used in this study were listed in Table S5.1.

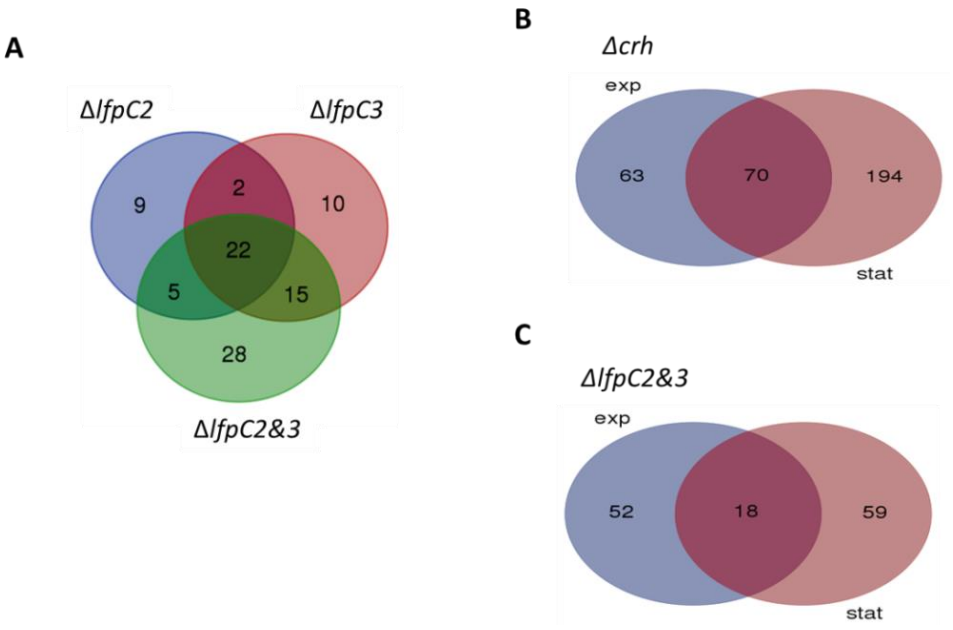


Figure S5.9 Venn diagram of DEGs that exhibit between different mutants at the exponential phase (A), or between the exponential and stationary phase in the Δ*crh* mutant (B) and in the Δ*lfpC2&3* mutant (C). exp, the exponential phase; stat, the stationary phase.

References

Abdou, L., Boileau, C., de Philip, P., Pages, S., Fierobe, H.P., Tardif, C. (2008) Transcriptional regulation of the *Clostridium cellulolyticum cip-cel* operon: A complex mechanism involving a catabolite-responsive element. *Journal of Bacteriology* 190: 1499-1506.

Alexander, J.K. (1968) Purification and specificity of cellobiose phosphorylase from *Clostridium thermocellum*. *J Biol Chem* 243: 2899-2904.

Andersch, W., Bahl, H., Gottschalk, G. (1983) Level of Enzymes Involved in Acetate, Butyrate, Acetone and Butanol Formation by *Clostridium acetobutylicum*. *European Journal of Applied Microbiology and Biotechnology* 18: 327-332.

Argyros, D.A., Tripathi, S.A., Barrett, T.F., Rogers, S.R., Feinberg, L.F., Olson, D.G., *et al.* (2011) High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology* 77: 8288-8294.

Balan, V. (2014) Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN biotechnology* 2014: 463074.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-1712.

Bayer, E.A., Belaich, J.P., Shoham, Y., Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58: 521-554.

Bayer, E.A., Lamed, R. (1986) Ultrastructure of the cell-surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *Journal of Bacteriology* 167: 828-836.

Bayer, E.A., Shimon, L.J., Shoham, Y., Lamed, R. (1998) Cellulosomes-structure and ultrastructure. *J Struct Biol* 124: 221-234.

Bertolla, F., VanGijsegem, F., Nesme, X., Simonet, P. (1997) Conditions for natural transformation of Ralstonia solanacearum. *Applied and Environmental Microbiology* 63: 4965-4968.

Bhaya, D., Davison, M., Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45: 273-297.

Bikard, D., Euler, C.W., Jiang, W., Nussenzweig, P.M., Goldberg, G.W., Duportet, X., *et al.* (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat Biotechnol* 32: 1146-1150.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 41: 7429-7437.

Blair, B.G., Anderson, K.L. (1998) Comparison of staining techniques for scanning electron microscopic detection of ultrastructural protuberances on cellulolytic bacteria. *Biotechnic & Histochemistry* 73: 107-113.

Blencke, H.M., Homuth, G., Ludwig, H., Mader, U., Hecker, M., Stulke, J. (2003) Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. *Metabolic Engineering* 5: 133-149.

Blouzard, J.C., Bourgeois, C., de Philip, P., Valette, O., Belaich, A., Tardif, C., *et al.* (2007) Enzyme diversity of the cellulolytic system produced by *Clostridium cellulolyticum* explored by two-dimensional analysis: Identification of seven genes encoding new dockerin-containing proteins. *Journal of Bacteriology* 189: 2300-2309.

Blouzard, J.C., Coutinho, P.M., Fierobe, H.P., Henrissat, B., Lignon, S., Tardif, C., *et al.* (2010) Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* 10: 541-554.

Boerjan, W., Ralph, J., Baucher, M. (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54: 519-546.

Bokinsky, G., Peralta-Yahya, P.P., George, A., Holmes, B.M., Steen, E.J., Dietrich, J., *et al.* (2011) Synthesis of three advanced biofuels from ionic liquid-pretreated

switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci U S A* 108: 19949-19954.

Bowater, R., Doherty, A.J. (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet* 2: e8.

Brown, S.D., Guss, A.M., Karpinets, T.V., Parks, J.M., Smolin, N., Yang, S., *et al.* (2011) Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. *Proc Natl Acad Sci U S A* 108: 13752-13757.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37: D233-238.

Carroll, D. (2012) A CRISPR approach to gene targeting. *Mol Ther* 20: 1658-1660.

Carvalho, A.L., Dias, F.M., Prates, J.A., Nagy, T., Gilbert, H.J., Davies, G.J., *et al.* (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc Natl Acad Sci U S A* 100: 13809-13814.

Casados-Vazquez, L.E., Lara-Gonzalez, S., Brieba, L.G. (2011) Crystal structure of the cysteine protease inhibitor 2 from Entamoeba histolytica: functional convergence of a common protein fold. *Gene* 471: 45-52.

Celik, H., Blouzard, J.C., Voigt, B., Becher, D., Trotter, V., Fierobe, H.P., *et al.* (2013) A Two-Component System (XydS/R) Controls the Expression of Genes Encoding CBM6-Containing Proteins in Response to Straw in *Clostridium cellulolyticum*. *PLoS One* 8.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., *et al.* (2013) Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res* 23: 465-472.

Chang, V.S., Holtzapple, M.T. (2000) Fundamental factors affecting biomass enzymatic reactivity. *Appl Biochem Biotechnol* 84-86: 5-37.

Cheung, J.K., Awad, M.M., McGowan, S., Rood, J.I. (2009) Functional analysis of the VirSR phosphorelay from *Clostridium perfringens*. *PLoS One* 4: e5849.

Cho, S.W., Kim, S., Kim, J.M., Kim, J.S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31: 230-232.

Chylinski, K., Le Rhun, A., Charpentier, E. (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol* 10.

Chylinski, K., Makarova, K.S., Charpentier, E., Koonin, E.V. (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research* 42: 6091-6105.

Citorik, R.J., Mimee, M., Lu, T.K. (2014) Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nature biotechnology* 32: 1141-1145.

Cobb, R.E., Wang, Y., Zhao, H. (2014) High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS synthetic biology*.

Collier, D.N., Hager, P.W., Phibbs, P.V. (1996) Catabolite repression control in the Pseudomonads. *Research in Microbiology* 147: 551-561.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-823.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome research* 14: 1188-1190.

Cui, G.Z., Hong, W., Zhang, J., Li, W.L., Feng, Y.G., Liu, Y.J., Cui, Q. (2012) Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *Journal of Microbiological Methods* 89: 201-208.

Dashtban, M., Schraft, H., Qin, W. (2009) Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. *Int J Biol Sci* 5: 578-595.

Davis, L., Maizels, N. (2011) DNA nicks promote efficient and safe targeted gene correction. *PLoS One* 6: e23981.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O.P., Kok, J. (2012) PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* 13: 299.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., *et al.* (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471: 602-607.

Desai, R.P., Papoutsakis, E.T. (1999) Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum. *Applied and Environmental Microbiology* 65: 936-945.

Desvaux, M. (2005) *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol Rev* 29: 741-764.

Desvaux, M., Guedon, E., Petitdemange, H. (2000) Cellulose catabolism by *Clostridium cellulolyticum* growing in batch culture on defined medium. *Applied and Environmental Microbiology* 66: 2461-2470.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology* 11: 87-93.

Deutscher, J., Ake, F.M.D., Derkaoui, M., Zebre, A.C., Cao, T.N., Bouraoui, H., *et al.* (2014) The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiology and Molecular Biology Reviews* 78: 231-256.

Deutscher, J., Francke, C., Postma, P.W. (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews* 70: 939-+.

Deutscher, J., Kuster, E., Bergstedt, U., Charrier, V., Hillen, W. (1995) Protein Kinase-Dependent Hpr/Ccpa Interaction Links Glycolytic Activity to Carbon Catabolite Repression in Gram-Positive Bacteria. *Molecular Microbiology* 15: 1049-1053.

Deutscher, J., Saier, M.H. (1983) Atp-Dependent Protein Kinase-Catalyzed Phosphorylation of a Seryl Residue in Hpr, a Phosphate Carrier Protein of the Phosphotransferase System in *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* 80: 6790-6794.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., *et al.* (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190: 1390-1400.

Dianov, G.L., Hubscher, U. (2013) Mammalian Base Excision Repair: the Forgotten Archangel. *Nucleic Acids Research* 41: 3483-3490.

Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M. (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* 41: 4336-4343.

Dickinson, D.J., Ward, J.D., Reiner, D.J., Goldstein, B. (2013) Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat Methods*.

Dillingham, M.S., Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72: 642-671, Table of Contents.

DOE (2012) Replacing the Whole Barrel To Reduce U.S. Dependance on Oil. *US DOE*.

Doi, R.H. (2008) Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers. *Ann N Y Acad Sci* 1125: 267-279.

Doi, R.H., Kosugi, A. (2004) Cellulosomes: Plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology* 2: 541-551.

Doi, R.H., Kosugi, A. (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol* 2: 541-551.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., Smith, F. (1956) Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry* 28: 350-356.

EIA (2016) FAQ: What is U.S. electricity generation by energy source? : U.S. Energy Information Adminstration.

Enyeart, P.J., Chirieleison, S.M., Dao, M.N., Perutka, J., Quandt, E.M., Yao, J., *et al.* (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Molecular Systems Biology* 9.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., Church, G.M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* 10: 1116-1121.

Esvelt, K.M., Wang, H.H. (2013) Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol* 9: 641.

Fan, L.H., Zhang, Z.J., Yu, X.Y., Xue, Y.X., Tan, T.W. (2012) Self-surface assembly of cellulosomes with two miniscaffoldins on Saccharomyces cerevisiae for cellulosic ethanol production. *Proc Natl Acad Sci U S A* 109: 13260-13265.

Federhen, S. (2012) The NCBI taxonomy database. *Nucleic Acids Research* 40: D136-D143.

Ferdinand, P.H., Borne, R., Trotter, V., Pages, S., Tardif, C., Fierobe, H.P., Perret, S. (2013) Are Cellulosome Scaffolding Protein CipC and CBM3-Containing Protein HycP, Involved in Adherence of *Clostridium cellulolyticum* to Cellulose? *PLoS One* 8.

Figueiredo da Silva, A.A., de Carvalho Vieira, L., Krieger, M.A., Goldenberg, S., Zanchin, N.I., Guimaraes, B.G. (2007) Crystal structure of chagasin, the endogenous cysteine-protease inhibitor from Trypanosoma cruzi. *J Struct Biol* 157: 416-423.

Filho, E.X., Tuohy, M.G., Puls, J., Coughlan, M.P. (1991) The xylan-degrading enzyme systems of *Penicillium capsulatum* and *Talaromyces emersonii*. *Biochem Soc Trans* 19: 25S.

Fontes, C.M., Gilbert, H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem* 79: 655-681.

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiacovo, M.P., Church, G.M., Calarco, J.A. (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods* 10: 741-743.

Friehs, K. (2004) Plasmid copy number and plasmid stability. *Advances in biochemical engineering/biotechnology* 86: 47-82.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*.

Gaj, T., Gersbach, C.A., Barbas, C.F., 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*.

Gal, L., Pages, S., Gaudin, C., Belaich, A., ReverbelLeroy, C., Tardif, C., Belaich, J.P. (1997) Characterization of the cellulolytic complex (cellulosome) produced by *Clostridium cellulolyticum*. *Applied and Environmental Microbiology* 63: 903-909.

Galinier, A., Haiech, J., Kilhoffer, M.C., Jaquinod, M., Stulke, J., Deutscher, J., MartinVerstraete, I. (1997) The *Bacillus subtilis crh* gene encodes a HPr-like protein involved in carbon catabolite repression. *Proceedings of the National Academy of Sciences of the United States of America* 94: 8439-8444.

Galinier, A., Kravanja, M., Engelmann, R., Hengstenberg, W., Kilhoffer, M.C., Deutscher, J., Haiech, J. (1998) New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression. *Proceedings of the National Academy of Sciences of the United States of America* 95: 1823-1828.

Gao, F., Zhang, R.R. (2011) Enzymes are enriched in bacterial essential genes. *PLoS One* 6: e21683.

Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109: E2579-2586.

Gehin, A., Gelhaye, E., Petitdemange, H. (1996) Adhesion of *Clostridium cellulolyticum* spores to filter paper. *Journal of Applied Bacteriology* 80: 187-190.

Georg, J., Hess, W.R. (2011) cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiology and Molecular Biology Reviews* 75: 286-+.

Gilbert, H.J., Stalbrand, H., Brumer, H. (2008) How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol* 11: 338-348.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., *et al.* (2013) CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* 154: 442-451.

Girio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S., Bogel-Lukasik, R. (2010) Hemicelluloses for fuel ethanol: A review. *Bioresour Technol* 101: 4775-4800.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., *et al.* (2006) Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America* 103: 425-430.

Goerke, B., Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology* 6: 613-624.

Gold, N.D., Martin, V.J. (2007) Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J Bacteriol* 189: 6787-6795.

Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R., Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio* 5: e00928-00913.

Goyal, G., Tsai, S.L., Madan, B., DaSilva, N.A., Chen, W. (2011) Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. *Microb Cell Fact* 10: 89.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., *et al.* (2013) Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics*.

Graves, M.C., Rabinowitz, J.C. (1986) In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J Biol Chem* 261: 11409-11415.

Grimmler, C., Held, C., Liebl, W., Ehrenreich, A. (2010) Transcriptional analysis of catabolite repression in *Clostridium acetobutylicum* growing on mixtures of D-glucose and D-xylose. *Journal of Biotechnology* 150: 315-323.

Groth, A.C., Calos, M.P. (2004) Phage integrases: biology and applications. *J Mol Biol* 335: 667-678.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuboock, R., Hofacker, I.L. (2008) The Vienna RNA Websuite. *Nucleic Acids Research* 36: W70-W74.

Gruno, M., Valjamae, P., Pettersson, G., Johansson, G. (2004) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol Bioeng* 86: 503-511.

Guedon, E., Desvaux, M., Petitdemange, H. (2002) Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering. *Appl Environ Microbiol* 68: 53-58.

Guglielmi, G., Beguin, P. (1998) Cellulase and hemicellulase genes of *Clostridium thermocellum* from five independent collections contain few overlaps and are widely scattered across the chromosome. *Fems Microbiology Letters* 161: 209-215.

Han, S.O., Yukawa, H., Inui, M., Doi, R.H. (2005) Effect of carbon source on the cellulosomal subpopulations of *Clostridium cellulovorans*. *Microbiology* 151: 1491-1497.

Hanson, K.G., Steinhauer, K., Reizer, J., Hillen, W., Stulke, J. (2002) HPr kinase/phosphatase of *Bacillus subtilis*: expression of the gene and effects of mutations on enzyme activity, growth and carbon catabolite repression. *Microbiology-Sgm* 148: 1805-1811.

He, Z.L., Zhou, J.Z. (2008) Empirical evaluation of a new method for calculating signal-to-noise ratio for microarray data analysis. *Applied and Environmental Microbiology* 74: 2957-2966.

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

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

Heap, J.T., Kuehne, S.A., Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C., Minton, N.P. (2010) The ClosTron: Mutagenesis in Clostridium refined and streamlined. *J Microbiol Methods* 80: 49-55.

Hemme, C.L., Fields, M.W., He, Q., Deng, Y., Lin, L., Tu, Q.C., *et al.* (2011) Correlation of Genomic and Physiological Traits of Thermoanaerobacter Species with Biofuel Yields. *Applied and Environmental Microbiology* 77: 7998-8008.

Hemme, C.L., Mouttaki, H., Lee, Y.J., Zhang, G., Goodwin, L., Lucas, S., *et al.* (2010) Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production. *J Bacteriol* 192: 6494-6496.

Higashide, W., Li, Y., Yang, Y., Liao, J.C. (2011) Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose. *Appl Environ Microbiol* 77: 2727-2733.

Hong, W., Zhang, J., Feng, Y.G., Mohr, G., Lambowitz, A.M., Cui, G.Z., *et al.* (2014) The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnology for Biofuels* 7.

Horstmann, N., Seidel, G., Aung-Hilbrich, L.M., Hillen, W. (2007) Residues His-15 and Arg-17 of HPr participate differently in catabolite signal processing via CcpA. *Journal of Biological Chemistry* 282: 1175-1182.

Horvath, P., Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327: 167-170.

Huang, H., Zheng, G., Jiang, W., Hu, H., Lu, Y. (2015) One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta biochimica et biophysica Sinica* 47: 231-243.

Hwang, W.Y., Fu, Y.F., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* 31: 227-229.

IEA (2009) Bioenergy-a Sustainable and Reliable Energy Source.

IEA (2011) IEA Statistics and Balance. International Energy Agency.

IEA (2016) Key World Energy Statistics 2016. International Energy Agency. pp. 6.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., *et al.* (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 21: 635-637.

Jault, J.M., Fieulaine, S., Nessler, S., Gonzalo, P., Di Pietro, A., Deutscher, J., Galinier, A. (2000) The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding. *Journal of Biological Chemistry* 275: 1773-1780.

Ji, Y.D., Marra, A., Rosenberg, M., Woodnutt, G. (1999) Regulated antisense RNA eliminates alpha-toxin virulence in *Staphylococcus aureus* infection. *Journal of Bacteriology* 181: 6585-6590.

Jia, H., Liang, T., Wang, Z., He, Z., Liu, Y., Yang, L., *et al.* (2014) Multistage regulator based on tandem promoters and CRISPR/Cas. *ACS synthetic biology* 3: 1007-1010.

Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31: 233-239.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., Weeks, D.P. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res*.

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., Yang, S. (2015) Multigene Editing in the *Escherichia coli* Genome via the CRISPR-Cas9 System. *Applied and Environmental Microbiology* 81: 2506-2514.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E.B., Doudna, J. (2013) RNA-programmed genome editing in human cells. *Elife* 2.

Juhas, M., Eberl, L., Church, G.M. (2012) Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends in Biotechnology* 30: 601-607.

Kang, S., Barak, Y., Lamed, R., Bayer, E.A., Morrison, M. (2006) The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. *Mol Microbiol* 60: 1344-1354.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., Siksnys, V. (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biol* 10.

Kelley, L.A., Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4: 363-371.

Khasanov, F.K., Zvingila, D.J., Zainullin, A.A., Prozorov, A.A., Bashkirov, V.I. (1992) Homologous Recombination between Plasmid and Chromosomal DNA in *Bacillus-Subtilis* Requires Approximately 70 Bp of Homology. *Molecular & General Genetics* 234: 494-497.

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B.A., Blanch, H.W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering* 109: 1083-1087.

Kleman-Leyer, K.M., Siika-Aho, M., Teeri, T.T., Kirk, T.K. (1996) The Cellulases Endoglucanase I and Cellobiohydrolase II of *Trichoderma reesei* Act Synergistically To Solubilize Native Cotton Cellulose but Not To Decrease Its Molecular Size. *Appl Environ Microbiol* 62: 2883-2887.

Kuhad, R.C., Gupta, R., Singh, A. (2011) Microbial cellulases and their industrial applications. *Enzyme Res* 2011: 280696.

Kuhad, R.C., Singh, A. (1993) Lignocellulose Biotechnology - Current and Future-Prospects. *Critical Reviews in Biotechnology* 13: 151-172.

Kung, S.H., Retchless, A.C., Kwan, J.Y., Almeida, R.P. (2013) Effects of DNA size on transformation and recombination efficiencies in Xylella fastidiosa. *Applied and environmental microbiology* 79: 1712-1717.

Lan, E.I., Liao, J.C. (2013) Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. *Bioresour Technol* 135: 339-349.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., Qi, L.S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* 8: 2180-2196.

Lee, J., Jang, Y.S., Choi, S.J., Im, J.A., Song, H., Cho, J.H., *et al.* (2012) Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for isopropanol-butanol-ethanol fermentation. *Appl Environ Microbiol* 78: 1416-1423.

Lee, J.H., Sung, B.H., Kim, M.S., Blattner, F.R., Yoon, B.H., Kim, J.H., Kim, S.C. (2009) Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microbial Cell Factories* 8.

Lee, S.Y., Mermelstein, L.D., Papoutsakis, E.T. (1993) Determination of Plasmid Copy Number and Stability in *Clostridium acetobutylicum* ATCC824. *Fems Microbiology Letters* 108: 319-323.

Letunic, I., Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*.

Letunic, I., Doerks, T., Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 40: D302-D305.

Levine, E., Zhang, Z., Kuhlman, T., Hwa, T. (2008) Quantitative characteristics of gene regulation by small RNA (vol 5, pg e229, 2007). *Plos Biology* 6: 191-191.

Lewis, N.S., Nocera, D.G. (2006) Powering the planet: Chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences of the United States of America* 103: 15729-15735.

Li, F.Y., Papworth, M., Minczuk, M., Rohde, C., Zhang, Y.Y., Ragozin, S., Jeltsch, A. (2007) Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Research* 35: 100-112.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., *et al.* (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* 31: 688-691.

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

Li, Y.C., Tschaplinski, T.J., Engle, N.L., Hamilton, C.Y., Rodriguez, M., Liao, J.C., *et al.* (2012) Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations. *Biotechnology for Biofuels* 5.

Liao, J.C., Mi, L., Pontrelli, S., Luo, S.S. (2016) Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology* 14: 288-304.

Lin, P.P., Mi, L., Morioka, A.H., Yoshino, K.M., Konishi, S., Xu, S.C., *et al.* (2015) Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metabolic engineering* 31: 44-52.

Lin, Y.N., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., *et al.* (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research* 42: 7473-7485.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., *et al.* (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364.

Liu, M.Z., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J., Blattner, F.R. (2005) Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *Journal of Biological Chemistry* 280: 15921-15927.

Loomis, W.F., Magasanik, B. (1967) Glucose-Lactose Diauxie in *Escherichia coli*. *Journal of Bacteriology* 93: 1397-+.

Lutke-Eversloh, T., Bahl, H. (2011) Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Curr Opin Biotechnol* 22: 634-647.

Lynd, L.R., van Zyl, W.H., McBride, J.E., Laser, M. (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Current opinion in biotechnology* 16: 577-583.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506-577, table of contents.

Lytle, B.L., Volkman, B.F., Westler, W.M., Heckman, M.P., Wu, J.H. (2001) Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *J Mol Biol* 307: 745-753.

Maamar, H., Abdou, L., Boileau, C., Valette, O., Tardif, C. (2006) Transcriptional analysis of the cip-cel gene cluster from *Clostridium cellulolyticum*. *J Bacteriol* 188: 2614-2624.

Maamar, H., Valette, O., Fierobe, H.P., Belaich, A., Belaich, J.P., Tardif, C. (2004) Cellulolysis is severely affected in *Clostridium cellulolyticum* strain cipCMut1. *Mol Microbiol* 51: 589-598.

Maida, Y., Kyo, S., Lassmann, T., Hayashizaki, Y., Masutomi, K. (2013) Off-Target Effect of Endogenous siRNA Derived from RMRP in Human Cells. *Int J Mol Sci* 14: 9305-9318.

Makarova, K.S., Aravind, L., Wolf, Y.I., Koonin, E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 6: 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9: 467-477.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., *et al.* (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*.

Mali, P., Esvelt, K.M., Church, G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods* 10: 957-963.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823-826.

Man, S.A., Cheng, R.B., Miao, C.C., Gong, Q.H., Gu, Y.C., Lu, X.Z., *et al.* (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Research* 39.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., *et al.* (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research* 40: D115-D122.

Marraffini, L.A., Sontheimer, E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11: 181-190.

Martin-Verstraete, I., Deutscher, J., Galinier, A. (1999) Phosphorylation of HPr and Crh by HprK, early steps in the catabolite repression signalling pathway for the *Bacillus subtilis* levanase operon. *Journal of Bacteriology* 181: 2966-2969.

Martinez-Alonso, M., Garcia-Fruitos, E., Ferrer-Miralles, N., Rinas, U., Villaverde, A. (2010) Side effects of chaperone gene co-expression in recombinant protein production. *Microbial cell factories* 9: 64.

Masset, J., Calusinska, M., Hamilton, C., Hiligsmann, S., Joris, B., Wilmotte, A., Thonart, P. (2012) Fermentative hydrogen production from glucose and starch using pure strains and artificial co-cultures of Clostridium spp. *Biotechnol Biofuels* 5: 35.

McMillan, J.D., Newman, M.M., Templeton, D.W., Mohagheghi, A. (1999) Simultaneous saccharification and cofermentation of dilute-acid pretreated yellow poplar hardwood to ethanol using xylose-fermenting *Zymomonas mobilis*. *Appl Biochem Biotechnol* 77-79: 649-665.

Meguro, H., Morisaka, H., Kuroda, K., Miyake, H., Tamaru, Y., Ueda, M. (2011) Putative role of cellulosomal protease inhibitors in *Clostridium cellulovorans* based on gene expression and measurement of activities. *J Bacteriol* 193: 5527-5530.

Mercier, E.D.A. (2009) Energy Recovery. In: Mercier EDA, editor. New York: Nova Science Publishers, Inc.

Metzger, M.J., McConnell-Smith, A., Stoddard, B.L., Miller, A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res* 39: 926-935.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., *et al.* (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res*.

Mijakovic, I., Poncet, S., Galinier, A., Monedero, V., Fieulaine, S., Janin, J., *et al.* (2002) Pyrophosphate-producing protein dephosphorylation by HPr kinase/phosphorylase: A relic of early life? *Proceedings of the National Academy of Sciences of the United States of America* 99: 13442-13447.

Miller, J.C., Tan, S.Y., Qiao, G.J., Barlow, K.A., Wang, J.B., Xia, D.F., *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* 29: 143-U149.

Minczuk, M., Papworth, M.A., Kolasinska, P., Murphy, M.P., Klug, A. (2006) Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. *Proceedings of the National Academy of Sciences of the United States of America* 103: 19689-19694.

Mitchell, W.J. (2016) Sugar uptake by the solventogenic clostridia. *World Journal of Microbiology & Biotechnology* 32.

Mohand-Oussaid, O., Payot, S., Guedon, E., Gelhaye, E., Youyou, A., Petitdemange, H. (1999) The extracellular xylan degradative system in *Clostridium cellulolyticum* cultivated on xylan: Evidence for cell-free cellulosome production. *Journal of Bacteriology* 181: 4035-4040.

Moses, V., Prevost, C. (1966) Catabolite Repression of Beta-Galactosidase Synthesis in *Escherichia coli*. *Biochemical Journal* 100: 336-&.

Mussolino, C., Cathomen, T. (2012) TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol* 23: 644-650.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39: 9283-9293.

Na, D., Yoo, S.M., Chung, H., Park, H., Park, J.H., Lee, S.Y. (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology* 31: 170-174.

Naik, S.N., Goud, V.V., Rout, P.K., Dalai, A.K. (2010) Production of first and second generation biofuels: A comprehensive review. *Renewable & Sustainable Energy Reviews* 14: 578-597.

Nakayama, T., Fish, M.B., Fisher, M., Oomen-Hajagos, J., Thomsen, G.H., Grainger, R.M. (2013) Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus tropicalis*. *Genesis*.

Naumoff, D.G. (2011) Hierarchical classification of glycoside hydrolases. *Biochemistry (Mosc)* 76: 622-635.

Ng, T.K., Benbassat, A., Zeikus, J.G. (1981) Ethanol-Production by Thermophilic Bacteria - Fermentation of Cellulosic Substrates by Cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. *Applied and Environmental Microbiology* 41: 1337-1343.

Ng, T.K., Zeikus, J.G. (1982) Differential Metabolism of Cellobiose and Glucose by *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. *Journal of Bacteriology* 150: 1391-1399.

O'Neill, B.C., Oppenheimer, M. (2002) Climate change - Dangerous climate impacts and the Kyoto protocol. *Science* 296: 1971-1972.

Oh, J.H., van Pijkeren, J.P. (2014) CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res* 42: e131.

Oh, K.K., Kim, S.W., Jeong, Y.S., Hong, S.I. (2000) Bioconversion of cellulose into ethanol by nonisothermal simultaneous saccharification and fermentation. *Appl Biochem Biotechnol* 89: 15-30.

Oleksy, A., Golonka, E., Banbula, A., Szmyd, G., Moon, J., Kubica, M., *et al.* (2004) Growth phase-dependent production of a cell wall-associated elastinolytic cysteine proteinase by *Staphylococcus epidermidis*. *Biol Chem* 385: 525-535.

Olofsson, K., Bertilsson, M., Liden, G. (2008) A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol Biofuels* 1: 7.

Olofsson, K., Palmqvist, B., Liden, G. (2010) Improving simultaneous saccharification and co-fermentation of pretreated wheat straw using both enzyme and substrate feeding. *Biotechnol Biofuels* 3: 17.

Olson, D.G., Tripathi, S.A., Giannone, R.J., Lo, J., Caiazza, N.C., Hogsett, D.A., *et al.* (2010) Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc Natl Acad Sci U S A* 107: 17727-17732.

Olsson, L., Soerensen, H.R., Dam, B.P., Christensen, H., Krogh, K.M., Meyer, A.S. (2006) Separate and simultaneous enzymatic hydrolysis and fermentation of wheat hemicellulose with recombinant xylose utilizing *Saccharomyces cerevisiae. Appl Biochem Biotechnol* 129-132: 117-129.

Osterberg, S., del Peso-Santos, T., Shingler, V. (2011) Regulation of Alternative Sigma Factor Use. *Annual Review of Microbiology, Vol 65* 65: 37-55.

Pabo, C.O., Peisach, E., Grant, R.A. (2001) Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annual Review of Biochemistry* 70: 313-340.

Pacala, S., Socolow, R. (2004) Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. *Science* 305: 968-972.

Papanek, B., Biswas, R., Rydzak, T., Guss, A.M. (2015) Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. *Metabolic engineering* 32: 49-54.

Parche, S., Beleut, M., Rezzonico, E., Jacobs, D., Arigoni, F., Titgemeyer, F., Jankovic, I. (2006) Lactose-over-glucose preference in *Bifidobacterium longum* NCC2705: glcP, encoding a glucose transporter, is subject to lactose repression. *Journal of Bacteriology* 188: 1260-1265.

Park, E.Y., Naruse, K., Kato, T. (2012) One-pot bioethanol production from cellulose by co-culture of *Acremonium cellulolyticus* and *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 5: 64.

Perez, J., Munoz-Dorado, J., de la Rubia, T., Martinez, J. (2002) Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol* 5: 53-63.

Perret, S., Belaich, A., Fierobe, H.P., Belaich, J.P., Tardif, C. (2004) Towards designer cellulosomes in clostridia: Mannanase enrichment of the cellulosomes produced by *Clostridium cellulolyticum*. *Journal of Bacteriology* 186: 6544-6552.

Perret, S., Maamar, H., Belaich, J.P., Tardif, C. (2004) Use of antisense RNA to modify the composition of cellulosomes produced by *Clostridium cellulolyticum*. *Mol Microbiol* 51: 599-607.

Perret, S., Maamar, H., Belaich, J.P., Tardif, C. (2004) Use of antisense RNA to modify the composition of cellulosomes produced by *Clostridium cellulolyticum*. *Molecular Microbiology* 51: 599-607.

Perutka, J., Wang, W.J., Goerlitz, D., Lambowitz, A.M. (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *Journal of Molecular Biology* 336: 421-439.

Petitdemange, E., Caillet, F., Giallo, J., Gaudin, C. (1984) *Clostridium cellulolyticum* Sp-Nov, a Cellulolytic, Mesophilic Species from Decayed Grass. *International Journal of Systematic Bacteriology* 34: 155-159.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45.

Philippidis, G.P., Smith, T.K., Wyman, C.E. (1993) Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process. *Biotechnol Bioeng* 41: 846-853.

Pinheiro, B.A., Proctor, M.R., Martinez-Fleites, C., Prates, J.A., Money, V.A., Davies, G.J., *et al.* (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J Biol Chem* 283: 18422-18430.

Pitcher, R.S., Brissett, N.C., Doherty, A.J. (2007) Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* 61: 259-282.

Postma, P.W., Lengeler, J.W., Jacobson, G.R. (1993) Phosphoenolpyruvate - Carbohydrate Phosphotransferase Systems of Bacteria. *Microbiological Reviews* 57: 543-594.

Pritchard, L., White, J.A., Birch, P.R., Toth, I.K. (2006) GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* 22: 616-617.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152: 1173-1183.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173-1183.

Raman, B., Pan, C., Hurst, G.B., Rodriguez, M., Jr., McKeown, C.K., Lankford, P.K., *et al.* (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS One* 4: e5271.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., *et al.* (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154: 1380-1389.

Rao, M.B., Tanksale, A.M., Ghatge, M.S., Deshpande, V.V. (1998) Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* 62: 597-635.

Ravcheev, D.A., Khoroshkin, M.S., Laikova, O.N., Tsoy, O.V., Sernova, N.V., Petrova, S.A., *et al.* (2014) Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Frontiers in Microbiology* 5.

Rawlings, N.D., Barrett, A.J., Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 40: D343-350.

Reichenbecher, M., Lottspeich, F., Bronnenmeier, K. (1997) Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium stercorarium*. *Eur J Biochem* 247: 262-267.

Reizer, J., Bergstedt, U., Galinier, A., Kuster, E., Saier, M.H., Hillen, W., *et al.* (1996) Catabolite repression resistance of gnt operon expression in *Bacillus subtilis* conferred by mutation of His-15, the site of phosphoenolpyruvate-dependent phosphorylation of the phosphocarrier protein HPr. *Journal of Bacteriology* 178: 5480-5486.

Ren, C., Gu, Y., Wu, Y., Zhang, W.W., Yang, C., Yang, S., Jiang, W.H. (2012) Pleiotropic functions of catabolite control protein CcpA in Butanol-producing *Clostridium acetobutylicum*. *BMC Genomics* 13.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y.F., Law, C.W., Shi, W., Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43.

Rocha, E.P., Cornet, E., Michel, B. (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* 1: e15.

Rose, I.A. (1955) Acetate kinase of bacteria (acetokinase). *Methods in Enzymology* 1: 591-595.

Saha, B.C., Nichols, N.N., Qureshi, N., Cotta, M.A. (2011) Comparison of separate hydrolysis and fermentation and simultaneous saccharification and fermentation processes for ethanol production from wheat straw by recombinant *Escherichia coli* strain FBR5. *Appl Microbiol Biotechnol* 92: 865-874.

Saier, M.H. (2000) Families of transmembrane sugar transport proteins. *Molecular Microbiology* 35: 699-710.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* 39: 9275-9282.

Schumacher, M.A., Allen, G.S., Diel, M., Seidel, G., Hillen, W., Brennan, R.G. (2004) Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell* 118: 731-741.

Schumacher, M.A., Seidel, G., Hillen, W., Brennan, R.G. (2006) Phosphoprotein Crh-Ser(46)-P displays altered binding to CcpA to effect carbon catabolite regulation. *Journal of Biological Chemistry* 281: 6793-6800.

Schwarz, W.H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol* 56: 634-649.

Schwarz, W.H., Zverlov, V.V. (2006) Protease inhibitors in bacteria: an emerging concept for the regulation of bacterial protein complexes? *Mol Microbiol* 60: 1323-1326.

Selle, K., Barrangou, R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends in microbiology* 23: 225-232.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., *et al.* (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108: 10098-10103.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., *et al.* (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31: 686-688.

Shaw, A.J., Podkaminer, K.K., Desai, S.G., Bardsley, J.S., Rogers, S.R., Thorne, P.G., *et al.* (2008) Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America* 105: 13769-13774.

Shen, P., Huang, H.V. (1986) Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics* 112: 441-457.

Shuman, S., Glickman, M.S. (2007) Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol* 5: 852-861.

Singh, K.D., Schmalisch, M.H., Stulke, J., Gorke, B. (2008) Carbon Catabolite Repression in *Bacillus subtilis*: Quantitative Analysis of Repression Exerted by Different Carbon Sources. *Journal of Bacteriology* 190: 7275-7284.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., Siksnys, V. (2013) In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* 32: 385-394.

Smith, G.R. (2001) Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. *Annu Rev Genet* 35: 243-274.

Solopova, A., van Gestel, J., Weissing, F.J., Bachmann, H., Teusink, B., Kok, J., Kuipers, O.P. (2014) Bet-hedging during bacterial diauxic shift. *Proceedings of the National Academy of Sciences of the United States of America* 111: 7427-7432.

Sorek, R., Lawrence, C.M., Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry* 82: 237-266.

Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., *et al.* (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463: 559-562.

Stevenson, D.M., Weimer, P.J. (2005) Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. *Appl Environ Microbiol* 71: 4672-4678.

Stulke, J., Hillen, W. (1999) Carbon catabolite repression in bacteria. *Current Opinion in Microbiology* 2: 195-201.

Sygmund, C., Kracher, D., Scheiblbrandner, S., Zahma, K., Felice, A.K., Harreither, W., *et al.* (2012) Characterization of the two *Neurospora crassa* cellobiose dehydrogenases and their connection to oxidative cellulose degradation. *Appl Environ Microbiol* 78: 6161-6171.

Tang, Y., Zhao, D., Cristhian, C., Jiang, J. (2011) Simultaneous saccharification and cofermentation of lignocellulosic residues from commercial furfural production and corn kernels using different nutrient media. *Biotechnol Biofuels* 4: 22.

Thomason, M.K., Storz, G. (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annual review of genetics* 44: 167-188.

Tolonen, A.C., Chilaka, A.C., Church, G.M. (2009) Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol Microbiol* 74: 1300-1313.

Tong, Y., Charusanti, P., Zhang, L., Weber, T., Lee, S.Y. (2015) CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS synthetic biology*.

Tripathi, S.A., Olson, D.G., Argyros, D.A., Miller, B.B., Barrett, T.F., Murphy, D.M., *et al.* (2010) Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Appl Environ Microbiol* 76: 6591-6599.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11: 636-646.

van den Bogaard, P.T.C., Kleerebezem, M., Kuipers, O.P., de Vos, W.M. (2000) Control of lactose transport, beta-galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus*: Evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *Journal of Bacteriology* 182: 5982-5989.

van der Ploeg, J.R. (2009) Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155: 1966-1976.

Vandamme, W.S.E. (2009) Biofuels. In: Wim Soetaert EV, editor: Wiley.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., Boerjan, W. (2010) Lignin biosynthesis and structure. *Plant Physiol* 153: 895-905.

Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B.B., Tijsterman, M., van den Heuvel, S., Boxem, M. (2013) CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis elegans. *Genetics*.

Wang, H.Y., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R. (2013) One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* 153: 910-918.

Wang, Q., Tull, D., Meinke, A., Gilkes, N.R., Warren, R.A., Aebersold, R., Withers, S.G. (1993) Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-beta-1,4-glucanase. *J Biol Chem* 268: 14096-14102.

Wang, T., Wei, J.J., Sabatini, D.M., Lander, E.S. (2014) Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* 343: 80-84.

Wang, Y., Yau, Y.Y., Perkins-Balding, D., Thomson, J.G. (2011) Recombinase technology: applications and possibilities. *Plant Cell Rep* 30: 267-285.

Warnes, G.B., Ben (2009) R package version.

Wen, F., Sun, J., Zhao, H. (2010) Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. *Appl Environ Microbiol* 76: 1251-1260.

WHO (2016) Ambient (outdoor) air quality and health. World Health Organization.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379: 466-469.

Wilson, R.C., Doudna, J.A. (2013) Molecular mechanisms of RNA interference. *Annu Rev Biophys* 42: 217-239.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* 32: 670-676.

Wu, Z., Lee, Y.Y. (1998) Nonisothermal simultaneous saccharification and fermentation for direct conversion of lignocellulosic biomass to ethanol. *Appl Biochem Biotechnol* 70-72: 479-492.

Wyman, C., Kanaar, R. (2006) DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* 40: 363-383.

Xiao, H., Gu, Y., Ning, Y.Y., Yang, Y.L., Mitchell, W.J., Jiang, W.H., Yang, S. (2011) Confirmation and Elimination of Xylose Metabolism Bottlenecks in Glucose

Phosphoenolpyruvate-Dependent Phosphotransferase System-Deficient *Clostridium acetobutylicum* for Simultaneous Utilization of Glucose, Xylose, and Arabinose. *Applied and Environmental Microbiology* 77: 7886-7895.

Xie, K., Yang, Y. (2013) RNA-guided Genome Editing in Plants Using A CRISPR-Cas System. *Mol Plant*.

Xu, C., Huang, R., Teng, L., Wang, D., Hemme, C.L., Borovok, I., *et al.* (2013) Structure and regulation of the cellulose degradome in Clostridium cellulolyticum. *Biotechnology for biofuels* 6: 73.

Xu, C.G., Huang, R.R., Teng, L., Jing, X.Y., Hu, J.Q., Cui, G.Z., *et al.* (2015) Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. *Nature communications* 6.

Xu, Q., Resch, M.G., Podkaminer, K., Yang, S.H., Baker, J.O., Donohoe, B.S., *et al.* (2016) Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. *Science Advances* 2.

Xu, T., Li, Y., He, Z., Zhou, J. (2014) Dockerin-containing protease inhibitor protects key cellulosomal cellulases from proteolysis in *Clostridium cellulolyticum*. *Molecular microbiology* 91: 694-705.

Xu, T., Li, Y., Shi, Z., Hemme, C.L., Zhu, Y., Van Nostrand, J.D., *et al.* (2015) Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Applied and Environmental Microbiology* 81: 4423-4431.

Xu, T., Li, Y., Van Nostrand, J.D., He, Z., Zhou, J. (2014) Cas9-based tools for targeted genome editing and transcriptional control. *Appl Environ Microbiol* 80: 1544-1552.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., Jaenisch, R. (2013) One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*.

You, C., Zhang, X.Z., Zhang, Y.H.P. (2012) Mini-scaffoldin enhanced mini-cellulosome hydrolysis performance on low-accessibility cellulose (Avicel) more than on high-accessibility amorphous cellulose. *Biochemical Engineering Journal* 63: 57-65.

Yu, Y., Tangney, M., Aass, H.C., Mitchell, W.J. (2007) Analysis of the mechanism and regulation of lactose transport and metabolism in *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* 73: 1842-1850.

Yue, Z., Bin, W., Baixu, Y., Peiji, G. (2004) Mechanism of cellobiose inhibition in cellulose hydrolysis by cellobiohydrolase. *Sci China C Life Sci* 47: 18-24.

Zhang, D.Y., Liu, J.J., Li, B.J. (2014) Tackling Air Pollution in China-What do We Learn from the Great Smog of 1950s in LONDON. *Sustainability* 6: 5322-5338.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., *et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45: 303-313.

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., *et al.* (2013) Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol Cell* 50: 488-503.

Zhong, J., Karberg, M., Lambowitz, A.M. (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research* 31: 1656-1664.

Zhu, M., Li, P., Gong, X., Wang, J. (2012) A comparison of the production of ethanol between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using unpretreated cassava pulp and enzyme cocktail. *Biosci Biotechnol Biochem* 76: 671-678.

Zhu, Y., Lee, Y.Y., Elander, R.T. (2007) Conversion of aqueous ammonia-treated corn stover to lactic acid by simultaneous saccharification and cofermentation. *Appl Biochem Biotechnol* 137-140: 721-738.

Zhu, Y., Liu, X., Yang, S.T. (2005) Construction and characterization of pta gene-deleted mutant of *Clostridium tyrobutyricum* for enhanced butyric acid fermentation. *Biotechnology and bioengineering* 90: 154-166.