The Expansion of Available Bioparts and the Development of CRISPR/Cas9 for Genetic Engineering of *Parageobacillus thermoglucosidasius*

Matthew Simon Harold Lau MSci

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

September 2018



Abstract

Growing public concern over the environmental and health impacts caused by the use of fossil fuels, coupled with diminishing reserves, is now compelling the pursuit for an alternative source. The search for alternative methods to manufacture products conventionally derived from fossil fuels, has led to research into microbial conversion of lignocellulosic biomass into analogous derivatives. Though many well established renewable energy methods currently exist, research into the production of petroleum-based products, such as transportation fuel and platform chemicals, is currently insufficient. Second generation biofuels have successfully overcome numerous problems associated with original biofuel production; however, many challenges are prevalent for established mesophilic organisms such as *Escherichia coli* or *Saccharomyces cerevisiae*. Thermophilic species have been presented as a potential platform strain to circumvent these problems. Typically non-fastidious and catabolically flexible organisms, thermophiles can ferment both hexose and pentose sugars, facilitate lower overall fermentation costs and are generally more robust organisms, a trait valuable for industrial processes. Unfortunately, current genetic tools are not well established to achieve the necessary yields to replace both the fossil fuel industry and the currently used organisms.

This study details the identification and subsequent testing of additional genetic tools, to expand the ability for genetic engineering in the

thermophile *Parageobacillus thermoglucosidasius*. The capacity to express exogenous DNA is often predicated on the availability of necessary gene expression elements for complete optimisation of synthetic systems. The identification and expression of the first two thermostable fluorescent proteins in *P. thermoglucosidasius* allowed the characterisation of several important gene expression elements, which included a collection of promoters, ribosomal binding sites and transcriptional terminators. Subsequently, these tools can now be utilised for precise expression of both endogenous and exogenous DNA in *P. thermoglucosidasius*.

The initial subset of libraries were further successfully utilised to enable the first expression of a CRISPR/Cas9 complex, isolated from *Streptococcus thermophilus*, in *P. thermoglucosidasius*. This bacterial adaptive immunity system was then repurposed to facilitate successful genome editing in *P. thermoglucosidasius*, verified by the removal of the endogenous acetate kinase gene. The efficacy of CRISPR/Cas9 for genome engineering was further demonstrated by its ability to integrate exogenous DNA into both the chromosome and the extrachromosomal megaplasmid present in *P. thermoglucosidasius*. Cas9 was also validated as an effective method of curing plasmids, through the removal of the megaplasmids, pNCI001 and pNCI002, present in *P. thermoglucosidasius*. The curing displayed no essential function of either megaplasmid under the conditions tested, but implied the presence of a carbon utilisation pathway, for degradation of aromatic hydrocarbons. The curing of the megaplasmids indicated no beneficial effects and therefore displayed their

potential as a vehicle for heterologous expression, due to their stability and increased copy number.

Validation of the Cas9-based technology was further achieved through the deletion of the native R-M systems, which revealed an increase in transformation efficiency. This increase can be co-opted to offset the reduction in transformation efficiency observed through the introduction of Cas9 and furthermore, facilitate more efficient allele exchange.

The ability to increase succinate production was a further demonstration of the applicability of these developed tools for genome engineering of *P. thermoglucosidasius*. The production of this important platform chemical, typically produced from petroleum, was achieved through the deletion of several genes from within the central carbon utilisation pathways, using the Cas9 technology exploited in this study. Deletion of succinate dehydrogenase resulted in a 3.5-fold increase compared to WT and produced a base-line strain for further deletions, which will enable the diversion of carbon flux towards succinate production.

Both the tools and strains presented in this study can be built upon to synthetically produce an industrially useful strain of *P. thermoglucosidasius*. This demonstrates significant potential for the future production of biofuels and platform chemicals from lignocellulosic biomass.

Acknowledgements

I would like to begin by thanking my academic supervisors Prof. Nigel Minton, Dr. Ying Zhang, and Dr. Vinod Kumar, for this opportunity and all the help you have provided throughout. Thank you also to my bench supervisor, Dr. Lili Sheng for the lab assistance and photography conversations you supplied along the way. I would also like to acknowledge the funding provided by the BBSRC and ATUM (formerly DNA 2.0), my industrial sponsor, which allowed me to complete this work.

I would like to thank everyone from the SBRC Nottingham, both past and present, it has been a pleasure to get to know every one of you. Special thanks to Evangelos and Magda, it was always amazing to see your smiling faces each day and thank you Evangelos for all the help with the baffled flasks over the years!

On a personal note, I would like to say a special thank you to Robert Harry Habgood. I would never have imagined that I would have met a lifelong friend through this process. You have definitely helped throughout with your humour, support, hot lava java and by introducing me to the wonders of MNF. Thank you also to Gary, Adele, Emma and Shil for all the care and love you have shown towards me over the years and for accepting me into your family. We're All Wednesday Aren't We!

Thank you to my family, Simon, Debbie, Hayden, Brandon and Cherisse Lau, whom without I would not be the person I am today. I could honestly not ask for a better family, who support, encourage, drive and care for me. You all mean so much to me, there aren't enough words to describe. On another note, Dim Sum for celebration?

And finally, thank you to my amazing wife, Nichola; you deserve a chapter dedicated to you. I can't describe how much you have helped, encouraged and inspired me throughout the whole time I have known you. You are the most caring person I have ever known, without which I wouldn't have made it to the end. Thank you for getting me through all the stress and the thesis writing and please, don't ever stop making me smile.

Contents

Abstract ... **ii**
Acknowledgements ... **v**
Contents ... **vii**
List of Figures ... **xii**
List of Tables ... **xvi**
Abbreviations Table ... **xviii**
1. Introduction ... **1**
1.1 The Need for Alternative Fuels ... **2**
1.1.1 The Global Energy Problem ... 2
1.1.2 Non-Renewable Energy Sources ... 2
1.1.3 The Environment, Health and Energy Security ... 4
1.1.4 Alternative Renewable Energy Sources ... 6
1.2 Biofuels and Platform Chemicals ... **7**
1.2.1 First Generation Biofuel Production ... 7
1.2.2 Second Generation ... 9
1.2.3 Platform Chemicals ... 12
1.2.4 Use of Mesophiles ... 14
1.2.5 A Move towards Thermophiles ... 16
1.3 The *Parageobacillus* and *Geobacillus* Genus ... **18**
1.3.1 *Parageobacillus thermoglucosidasius* ... 20
1.4 Genetic Engineering Tools ... **22**
1.4.1 Transformation ... 22
1.4.2 Gene Expression Elements ... 27
1.4.3 Reporter Systems ... 31
1.4.4 Inducible Systems ... 32
1.4.5 Manipulation of Bacterial Genomes ... 34
1.4.6 Allele Coupled Exchange ... 36
1.5 CRISPR/Cas9 ... **39**
1.5.1 Prokaryote Adaptive Immunity ... 39
1.5.2 Application in Genetic Engineering ... 43
1.6 Succinic Acid ... **45**
1.6.1 Succinate Production Pathways ... 45
1.6.2 Metabolic Engineering of Microorganisms to produce Succinate ... 49
1.7 The Overall Aims of this Thesis ... **50**

2. **Materials and Methods** **53**
2.1 Bioinformatics **54**
2.1.1 Analysis of DNA Sequences 54
2.1.2 Metabolic Pathway Information 54
2.1.3 Management of DNA Sequence Data and Plasmid Map Design 54
2.1.4 Oligonucleotide Design, Analysis and Synthesis 55
2.1.5 DNA Fragment Synthesis 55
2.1.6 Prediction of Bacterial Promoters 56
2.1.7 Statistical Analysis 56
2.1.8 NMR Data Analysis 56
2.2 Media **57**
2.2.1 Luria-Bertani medium (LB) 57
2.2.2 Soy Peptone Yeast Extract No Glycerol (2SPYNG) 57
2.2.3 Soy Peptone Yeast Extract (2SPY) 57
2.2.4 Tryptone Soya Agar (TSA) 58
2.2.5 Clostridium Basal Medium (CBM) 58
2.2.6 USYE 58
2.2.7 Modified ASYE 59
2.3 Antibiotics and Supplements **60**
2.4 Solution, Buffers and Gels **61**
2.4.1 Agarose Gel 61
2.4.2 Annealing Buffer (1X) 61
2.4.3 Protein Extraction Buffer (1X) 61
2.4.4 Electroporation buffer (EPB) 62
2.5 Bacterial Growth Conditions **62**
2.5.1 Escherichia coli 62
2.5.2 *Parageobacillus thermoglucosidasius* 62
2.5.3 Bacterial Cell Density 63
2.6 Bacterial Strains **64**
2.7 Plasmids **67**
2.8 Oligonucleotide Primers **74**
2.9 DNA Manipulation, Extraction and Sequencing **74**
2.9.1 Extraction and purification of *P. thermoglucosidasius* Chromosomal DNA 74
2.9.2 Extraction and Purification of Plasmid DNA 75
2.9.3 Amplification of DNA by Polymerase Chain Reaction (PCR) 76
2.9.4 Restriction Endonuclease Digestion of DNA 79
2.9.5 DNA Blunting 79
2.9.6 Dephosphorylation of DNA 79
2.9.7 TA Cloning 80

2.9.8 DNA analysis by Agarose Gel Electrophoresis 80
2.9.9 Agarose gel extraction and purification of DNA 80
2.9.10 Ligation of DNA fragments 81
2.9.11 Site-Directed Mutagenesis (SDM) 81
2.9.12 Sequencing of DNA 82
2.10 Transfer of Plasmid DNA into Bacterial Cells 82
2.10.1 Preparation of Electro-Competent *E. coli* Cells 82
2.10.2 Preparation of Chemical Competent *E. coli* Cells 83
2.10.3 Transformation into *E.coli* using Electroporation 83
2.10.4 Transformation into *E. coli* using Heat Shock 84
2.10.5 Preparation of Electro-Competent *P. thermoglucosidasius* Cells 84
2.10.6 Transformation into *P. thermoglucosidasius* using Electroporation 85
2.11 Protein Extraction and Visualisation 86
2.11.1 Protein Extraction through Chemical Lysis of Cells 86
2.11.2 Western Blot of HIS-tagged Protein 86
2.11.3 Protein Concentration Determination by Bicinchoninic Acid (BCA) assay 87
2.11.4 Analysis of Fluorescent Reporter Proteins 87
2.11.5 sGFP Assay 88
2.12 Enzymatic Assays 89
2.12.1 PheB assay 89
2.13 Analysis of Metabolites 89
2.12.1 HPLC-UV Analysis 89
3. Development of New and Existing Genetic Tools for *P. thermoglucosidasius* 91
3.1 Introduction 92
3.1.1 Genetic Tools Available for Engineering in *P. thermoglucosidasius* 92
3.1.2 Chapter Aims 98
3.2 Results 99
3.2.1 Identification and Testing of a Fluorescent Reporter Gene 99
3.2.2 Development of a Promoter Library for *P. thermoglucosidasius* 105
3.2.3 Validation of a Ribosomal Binding Site Library 110
3.2.4 Validation of Transcriptional Terminator Strengths 114
3.2.5 Codon Optimisation in *P. thermoglucosidasius* 117
3.2.6 Initial development of an Inducible Expression System for *P. thermoglucosidasius* 121
3.3 Discussion 126
3.4 Conclusions and Future Work 149

3.5 Key Outcomes 151

4. Development of CRISPR/Cas-based Knock-In and Knock-Out Technology 152

4.1 Introduction 153

4.1.1 Manipulation of Genes and Genomes 153

4.1.2 CRISPR/Cas9 System 154

4.1.3 Streptococcus thermophilus Cas9 155

4.1.4 Plasmids 156

4.1.5 Chapter Aims 159

4.2 Results 160

4.2.1 Development of a Thermostable CRISPR/Cas9 System in *P. thermoglucosidasius* 160

4.2.2 Elimination of Megaplasmids and Expansion of Tools 169

4.3 Discussion 188

4.4 Conclusions and Future Work 204

4.5 Key Outcomes 208

5. Increasing Transformation Efficiency through the Knock-out of R-M Systems 209

5.1 Introduction 210

5.1.1 Restriction Modification Systems in *P. thermoglucosidasius* 210

5.1.2 Chapter Aims 213

5.2 Results 214

5.2.1 Single Knock-out of Restriction Modification Systems 214

5.2.2 Creation of a Restriction Modification Deficient Strain 220

5.3 Discussion 224

5.4 Conclusions and Future Work 230

5.5 Key Outcomes 232

6. A View Towards Metabolic Engineering of *P. thermoglucosidasius* for Improved Succinic Acid Production 233

6.1 Introduction 234

6.1.1 Mixed Acid Fermentation 234

6.1.2 Succinic Acid Production 236

6.1.3 Chapter Aims 237

6.2 Results 238

6.2.1 Using CRISPR/Cas9 for Genetic Engineering for Strain Development 238

6.2.2 Effect of Cytochrome C Oxidase Deletion on Mixed Acid Fermentation 243

6.2.3 Effect of Succinate Dehydrogenase Deletion on Succinic Acid Production 248

6.3 Discussion 252

6.4 Conclusions and Future Work 262

6.5 Key Outcomes 265

7. General Discussion 267

7.1 Development of New and Existing Genetic Tools for *P. thermoglucosidasius* 269

7.2 Development of CRISPR/Cas-based Knock-In and Knock-Out Technology 272

7.3 Increasing Transformation Efficiency through KO of R-M Systems 274

7.4 A View towards Metabolic Engineering of *P. thermoglucosidasius* for Improved Succinic Acid Production 276

7.5 Conclusions 278

8. Bibliography 279

9. Supplementary Data 305

List of Figures

Chapter 1

Figure 1.1 An overview of renewable energy consumption **4**

Figure 1.2 An overview of the substrates used for biofuel production . . **12**

Figure 1.3 Production of platform chemicals from both renewable biomass and non-renewable fossil fuels **14**

Figure 1.4 A schematic demonstrating the mechanism of restriction modification systems **27**

Figure 1.5 A schematic demonstrating uracil biosynthesis and the conversion of 5-fluoroorotic acid to 5-fluorouridine monophosphate **37**

Figure 1.6 Gene integration at the *pyrE* locus using allele couple exchange **38**

Figure 1.7 A schematic demonstrating the mechanism of CRISPR/Cas9 . . . . . **42**

Figure 1.8 A schematic demonstrating the use of Cas9 for genetic engineering **44**

Figure 1.9 A schematic detailing the TCA cycle, mixed acid fermentation pathway and the glyoxylate pathway as routes toward succinate production **48**

Chapter 3

Figure 3.1 Images of the three commercially available fluorescent reports DasherGFP, PaprikaRFP and CayenneRFP from ATUM **100**

Figure 3.2 Emission spectra of sGFP and eCGP123 . . . . . **101**

Figure 3.3 Fluorescent and bright field images of sGFP fluorescing *P. thermoglucosidasius* **102**

Figure 3.4 A western blot of sGFP, GAF, eCGP123 and YtvA LOV **103**

Figure 3.5 Thermostability assay of sGFP and eCGP123 . . . . . **104**

Figure 3.6 A schematic representation of the application module **107**

Figure 3.7 Promoter strengths in both *P. thermoglucosidasius* and *E. coli* **109**

Figure 3.8 Ribosomal binding site strengths in both *P. thermoglucosidasius* and *E. coli* **113**

Figure 3.9 Terminator strengths in both *P. thermoglucosidasius* and *E. coli* **116**

Figure 3.10 Average codon usage for each amino acid within a gene . . . **118**

Figure 3.11 Specific protein activities of *pheB* gene variants for increased protein expression **120**

Figure 3.12 The effect of theophylline concentration on the growth of *P. thermoglucosidasius* **122**

Figure 3.13 The effect of theophylline concentration used for induction of Riboswitch E on sGFP expression in *P. thermoglucosidasius* **123**

Figure 3.14 Expression profiles of Riboswitch E, Riboswitch F and Riboswitch G **125**

Chapter 4

Figure 4.1 A schematic representation of the vectors pCas91_Ack and pCas93_Ack **164**

Figure 4.2 Theoretical and actual images demonstrating the activity of Cas9 . . . . . **166**

Figure 4.3 A schematic representation of the vectors pCas93_Ack_hr and pCas93_Ack_hrT **168**

Figure 4.4 A colony PCR of acetate kinase deletion using both stCas9-1 and stCas9-3 . . . . . **169**

Figure 4.5 A colony PCR of RepA deletion representing the loss of pNCI002 . . . . . **171**

Figure 4.6 A colony PCR of the Toxin deletion representing the loss of pNCI001 . . . . . **172**

Figure 4.7 A colony PCR of RepA and the Toxin deletion representing the loss of both pNCI002 and pNCI001 . . . . . **173**

Figure 4.8 Growth curves of the megaplasmid deficient strains **175**

Figure 4.9 Concentration of metabolites produced and glycerol utilised, by the megaplasmid deficient strains **177**

Figure 4.10 Nuclear magnetic resonance spectra of the megaplasmid deficient strains **180**

Figure 4.11 A schematic representation of the vector pCas93_eC_KI . . . **182**

Figure 4.12 A colony PCR and emission spectra demonstrating the integration of eCGP123 into the chromosome **183**

Figure 4.13 A colony PCR demonstrating the integration of eCGP123 into pNCI001 . . . . . **184**

Figure 4.14 Growth of *P. thermoglucosidasius* strains containing the *pyrE* integration on the pNCI001 . . . . . **186**

Chapter 5

Figure 5.1 A schematic representation of the genome location of the restriction modifications systems in *P. thermoglucosidasius* **212**

Figure 5.2 The transformation efficiency of wild-type *P. thermoglucosidasius* **215**

Figure 5.3 A colony PCR showing the single mutant deletion of each restriction modification system **217**

Figure 5.4 A colony PCR showing the double mutant deletion of each restriction modification system **220**

Figure 5.5 A colony PCR showing the triple mutant deletion of the restriction modification system **221**

Figure 5.6 Transformation efficiencies of all the produced restriction modification mutants **222**

Chapter 6

Figure 6.1 A schematic representation of the mixed acid fermentation pathway **235**

Figure 6.2 A colony PCR demonstrating the deletions of Ldh, AdhE and Cox **239**

Figure 6.3 A colony PCR using internal primers to demonstrating the loss of Ldh, AdhE and Cox **240**

Figure 6.4 Concentration of metabolites produced by the mutants Ldh, AdhE and Cox **241**

Figure 6.5 Concentration of metabolites produced by mutants containing a Cox mutation **245, 246**

Figure 6.6 A colony PCR demonstrating the deletion of Sdh **248**

Figure 6.7 Concentration of metabolites produced by Sdh mutants . . . **249**

Figure 6.8 Concentration of succinate produced by all the mutants tested **251**

List of Tables

Chapter 2

Table 2.1 Working concentrations of antibiotics and other supplements **60**

Table 2.2 Bacterial strains used and produced in this study **64-66**

Table 2.3 Plasmids used and produced in this study **67-73**

Table 2.4 Components for a colony PCR using DreamTaq Green PCR Master Mix **77**

Table 2.5 Thermal cycling conditions used for DreamTaq Green PCR Master Mix **77**

Table 2.6 Components for PCR using Phusion Master Mix **78**

Table 2.7 Thermal cycling conditions used for Phusion Master Mix . . . **78**

Chapter 3

Table 3.1 Excitation and emission wavelengths of fluorescent reporter proteins **101**

Table 3.2 Shine-Dalgarno sequences of the RBS library **111**

Table 3.3 Rho-independent terminator sequences of the terminator library **114, 115**

Table 3.4 Anti-SD sequences of 16S ribosomes of several prokaryotes **136**

Chapter 4

Table 4.1 Spacer and PAM analysis of stcas9-1 and stcas9-3 for the deletion of the ack gene **162, 163**

Table 4.2 Summary of growth of strains with a megaplasmid integration of *pyrE* **187**

Chapter 5

Table 5.1 SNPs identified from the genome sequencing of ΔRM1, ΔRM2 and ΔRM3 . . . . . **218, 219**

Supplementary

Table S.1 Primers used within this study **306-333**

Abbreviations Table

ACE	Allele Coupled Exchange
Ack	Acetate Kinase
AdhE	Aldehyde-alcohol Dehydrogenase
Ae	Aerobic
AE	Allelic Exchange
ATP	Adenosine triphosphate
BCA assay	Bicinchoninic Acid Assay
CAI	Codon Adaptive Index
cAMP	Cyclic Adenosine Monophosphate
Cas	CRISPR-associated proteins
CAT	chloramphenicol acetyltransferase
CBM	Clostridium Basal Medium
cDNA	Complimentary DNA
CFU	Colony Forming Unit
cGMP	Cyclic Guanosine Monophosphate
Cox	Cytochrome C Oxidase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
Cts	Citrate Synthase
dCas9	Catalytically Dead *cas9*
ds	Double-stranded
dsDNA	Double Stranded DNA
eCGP123	Enhanced Consensus Green Protein Variant 123
EDTA	Ethylenediaminetetraacetic acid
EPB	Electroporation buffer
Fbpa	Fructose-bisphosphate Aldolase
focA	Formate Transporter
FRET	Förster Resonance Energy Transfer
GAF	cGMP Phosphodiesterase/adenyl cyclase/FhlA
Gapd	Glyceraldehyde 3-phosphate Dehydrogenase
GFP	Green Fluorescent Protein
GHGs	Greenhouse Gases

Gpd	Glycerol-3-phosphate Dehydrogenase
Gpi2	Glucose-6-phosphate Isomerase 2
HA	Homology Arm
HIS-tag	Histidine-tag
HMSA	2-hydroxymuconic semialdehyde
HPLC-UV	High-performance liquid chromatography
HR	Homologous Recombination
IclR	Isocitrate Lyase Regulator
Idh	Isocitrate Dehydrogenase
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kan	Kanamycin Adenyltransferase
KI	Knock-in
km	Kanamycin Antibiotic
km^{R}	Kanamycin Resistance
KO	Knock-out
LB	Luria-Bertani medium
LDH	Lactate Dehydrogenase
LHA	Left Homology Arm
MA	Micro-aerobic
mRNA	Messenger RNA
MTase	Methyltransferase
NAD	Nicotinamide Adenine Dinucleotide
NHEJ	Non-homologous end-joining
NMR	Nuclear Magnetic Resonance
NOP	No Promoter
NTPase	Nucleoside-triphosphatase
O/N	Overnight
OMP	Orotidine 5'-monophosphate
ORF	Open Reading Frame
PAM	Protospacer Adjacent Motif
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
Pde	Phosphodiesterase
Pdh	Pyruvate Dehydrogenase
pdh^{up}	Pyruvate Dehydrogenase Upregulated
PEP	Phosphoenolpyruvate

Pfk	6-phosphofrucokinase
PFL	Pyruvate Formate Lyase
Pgk	Phosphoglycerate Kinase
PKC	Palm Kernel Cake
Ppc	Phosphoenolpyruvate Carboxylase
PSK	Post-segregational Killing
Pta	Phosphate Acetyltransferase
PyrE	Orotate Phosphoribosyltransferase
RACE	Rapid Amplification of cDNA Ends
RBS	Ribosomal Binding Sites
Rbx	Riboswitch
REase	Restriction Endonuclease
RepA	pNCI002 Replication Gene
RepB	pNCI001 Replication Gene
RFU	Relative Fluorescent Units
RHA	Right Homology Arm
R-M	Restriction Modification
RNase	Ribonuclease
rRNA	Ribosomal RNA
SD	Shine-Dalgarno
Sdh	Succinate Dehydrogenase
SDM	Site-Directed Mutagenesis
sGFP	Superfolder GFP
sgRNA	Synthetic Guide RNA
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SOE	Splicing by Overlap Extension
2SPY	Soy Peptone Yeast Extract
2SPYNG	Soy Peptone Yeast Extract No Glycerol
tAI	tRNA Adaptive Index
TALEN	Transcription Activator-like Effector Nuclease
TBS	Tris-buffered Saline
TCA	Tricarboxylic Acid
tet	Tetracycline
tetR	Tetracycline Repressor
TMB	3,3’,5,5’ tetramethylbenzidine

tracrRNA	Trans-activating CRISPR RNA
TRE	Tet Response Element
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	Transfer RNA
TSA	Tryptone Soya Agar
tTA	Tetracycline Transactivator
UMP	Uridine 5'-monophosphate
UTR	Untranslated Region
WT	Wild-type
X-gal	5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside
XylA	Xylose Isomerase
ZFN	Zinc-finger Nuclease
3-HP	3-Hydroxypropionic acid
5-FOA	5-fluoroorotic acid
5-FOA^R	5-fluoroorotic acid resistance
5-FUMP	5-fluorouridine monophosphate

Species Nomenclature

The species *Parageobacillus thermoglucosidasius* is synonymous with *Geobacillus thermoglucosidasius*, and is referred to as both within the scientific literature. The former is used throughout this thesis.

Chapter 1

Introduction

1.1 The Need for Alternative Fuels

1.1.1 The Global Energy Problem

The global population over the past two decades has drastically increased, reaching over a staggering 7.5 billion in 2017 (1) and is set to increase to over 9 billion by 2050. Furthermore this increase in population is leading to a global increase in food production, which is currently at 72 billion metric tonnes per year (2). This, together with an unprecedented pace of modernisation, has brought an increase in the demand for energy. As a result, the total world energy usage is continually rising at an alarming rate and has now brought the global energy demand to an unsustainable level.

1.1.2 Non-Renewable Energy Sources

The primary method of energy production is through the burning of fossil fuels (coal, oil, and natural gas), which accounts for over 75% of the total global energy consumption (Figure 1.1). The heat produced, through combustion, is primarily used for conversion into electrical power, mechanical energy for vehicles, industrial processes, or used directly for heating. Furthermore, many products that are heavily relied upon by everyday society, such as solvents, polymers, food ingredients and pharmaceuticals, are all produced through the petrochemical industry and

are heavily reliant upon oil (3,4). The formation of these fossil fuels occurs through the natural process of anaerobic decomposition of organic matter. This typically transpires over millions of years and therefore is a limited commodity. The availability of oil, coal and gas is now becoming limited with only approximately 26, 98 and 28 years (5), respectively left of these fossil fuels, which adds an additional time pressure to find alternative sources of energy.

Industrialised nations are resource intensive societies and are heavily reliant on fossil fuels for their energy, chemical and material needs. The increase in global population has been most significant in the east, particularly in China, where increasing industrialism and urbanisation has led to the country becoming the second largest consumer of energy after the United States (6). This is leading to a rapid depletion of fossil fuels, which is yielding drastic effects on the Earth's environment, through a phenomenon termed the 'green-house' effect. This is resulting in global warming and significant climate change.

Due to both the short life span remaining for fossil fuels and the rising public concern over global warming, finding renewable and carbon-neutral alternatives is essential. More recently there has been a push towards renewable sources of energy, significantly for the production of electricity. However, established renewable energy sources, such as wind or solar power, are not fully reliable alternatives, as their output is often determined by meteorological factors (7). Furthermore, the need for an

alternative is especially critical for sectors such as transportation and the petrochemical industry, which currently rely almost entirely on petroleum based fuels, with only 0.8% of transportation consumption coming from renewable biofuel (8) (Figure 1.1). It has therefore become a major industrial goal to produce liquid biofuels and platform chemicals, such as ethanol and succinate, from lignocellulosic biomass. If accomplished, liquid biofuels and other associated chemicals would reduce dependency on fossil fuels, in particular oil, and provide a reliable and sustainable, carbon-neutral alternative.

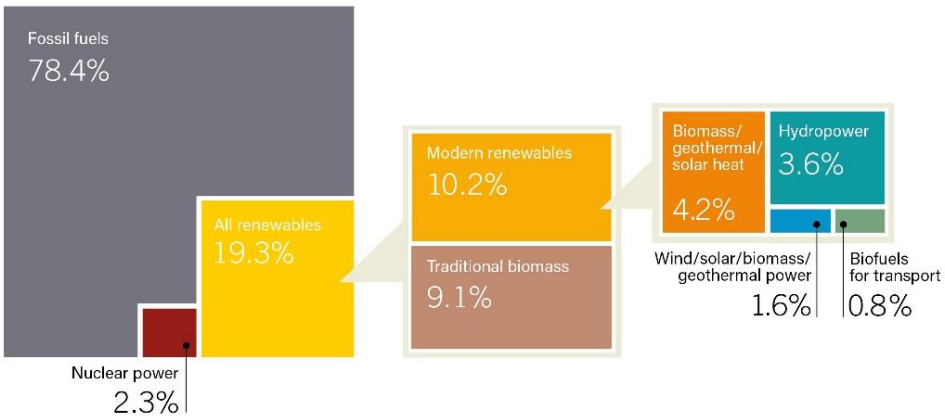


Figure 1.1 – A schematic representation of renewable energy consumption (REN21, 2017, Renewables 2017 Global Status Report).

1.1.3 The Environment, Health and Energy Security

As previously mentioned, the burning of fossil fuels is having a drastic effect on the Earth's climate. This is due to the release of vast amounts of carbon dioxide (CO_2), which is causing the so-called 'green-house' effect,

causing global warming and worldwide climate change. This global increase in temperature, projected to be between 2 °C – 6 °C (5), is having a huge impact on the planet, causing ocean acidification, desertification, melting of polar ice sheets, glacial retreat, and extreme weather conditions. These impacts are occurring over a relatively short space of time, thereby resulting in disastrous consequences to ecosystems and the wildlife within them.

In addition to the environmental impact, significant negative health problems are also associated with the burning of fossil fuels, both directly and indirectly. Alterations in weather patterns can affect communities through fluctuations in temperature and rainfall, causing droughts, floods and other such extreme weather phenomena. Furthermore, this general increase in global temperature may also increase the distribution of disease vectors, such as mosquitos, and escalate the incidence rate of their associated diseases, such as malaria (9). Sea level rise is another key feature associated with this increase in temperature, creating a risk for low lying coastal communities. Direct harm is also caused through air pollution, resulting from the burning of fossil fuels. Air pollutants present potential hazard to health, causing serious illness and leading to an increase in mortality rate (10). These pollutants are present in a variety of forms, such as persistent organic matter, gaseous pollutants, heavy metals and particular matter (11). The affect to health occurs at both an entire system level, through breathing problems and skin irritation, and a molecular level to cause diseases such as cancer.

In addition to the environmental and health difficulties associated with fossil fuels, their finite nature and their geographical location pose serious global political issues, with the energy market responsible for both political alliances and unrest. The foreign policies of key political powers continually reflect a drive for energy security, for example, China and Russia have made their presence felt in key energy producing areas. Meanwhile, in the past, Japan has demonstrated its concern by strengthening the presence of its air force and navy surrounding energy imports, which it is heavily reliant on due to limited domestic energy reserves (12). Political issues can also exaggerate the lack of energy security felt by the world, with the price of oil continually fluctuating due to the location of major reserves in the Middle East, an area which frequently experiences political turmoil and instability.

1.1.4 Alternative Renewable Energy Sources

Renewable energy is collected from a source that is naturally replenished over a human time scale. Wind, solar, geothermal and hydroelectric power are all well-established methods of producing energy and maintain zero or near-zero emissions of both air pollutants and greenhouse gases (GHGs) (13). These forms of energy are primarily aimed at producing electricity and could be used to produce the majority of energy for consumption. However, the most commonly used energy source is oil, which is utilised for chemical production and the transportation sector. It has therefore

become a major industrial goal to produce liquid biofuels and platform chemicals, from lignocellulosic biomass. If accomplished this would reduce dependency on fossil fuels, in particular oil, and provide renewable carbon-neutral alternatives. One option available is to utilise the natural processes of bacteria to generate these much needed fuels and chemicals through fermentation of biomass.

1.2 Biofuels and Platform Chemicals

1.2.1 First Generation Biofuel Production

Biofuels are liquid fuels that are produced from replenishable organic matter, as opposed to non-renewable fossil fuels. This plant matter is produced through photosynthesis using light, CO_2 and water, producing a carbon neutral energy source. The energy from the sun is captured and stored in the form of chemical bonds inside biologically produced materials. This is carbon fixation over a short period of geological time, as opposed to the production of fossil fuels which occur through slow anaerobic decomposition of buried organic matter exposed to high heat and pressure over millions of years. First generation biofuels are produced from food crops and, depending on the starting material, can be produced in the form of biodiesel or bioethanol. Biodiesel is a mix of fatty acid esters produced through a transesterification process involving fats and oils, such as rapeseed oil or palm oil. Whereas, bioethanol is produced

through a fermentation process, using crops such as sugar cane and maize, which are high in sugar and starch (14). Success has already been achieved for bioethanol production by the exploitation of organisms, such as yeast, through the fermentation process, utilising sugar or starch as the feedstock. This fermentation process is based on using the traditional method of brewing and uses variants of the yeast *Saccharomyces cerevisiae*. Currently, the majority of the world's bioethanol supply, 85%, is produced in Brazil and the USA, using a feedstock of sugar cane and maize respectively (15-17).

First generation biofuels were initially an attractive target, as they offered an achievable approach to sustainably produce transport fuels from biomass (18). Ethanol can be blended with conventional fuels without the need for major modifications to conventional car engines (19). However, in their current state, these so called first generation biofuels also pose many problems. Firstly, they are surrounded by a significant humanitarian debate, 'food versus fuel'. The fact that food sources are being used to produce a transportation fuel, when malnutrition is rife in major areas of the world, is a strongly contested issue. In addition, the substrate of these biofuels requires growth on fertile arable land. This reduces the land available for food crops, increasing the negative impact on human sustenance and supporting the 'food vs fuel' debate (20,21). Another consequence of this competition is an increase in food prices (22). There are now an increasing number of agricultural farmers making the switch to biofuel crops, in order to exploit governmental grants and subsidies, as

without this funding production is an expensive alternative to fossil fuels. This conversion of food to fuel farmers, limits the production of food, thereby driving up prices by natural supply and demand mechanisms. Another disadvantage associated with first generation biofuels is the acceleration of deforestation, which is occurring both legally and illegally in countries like Brazil and Malaysia. This deforestation is occurring to provide more land for the growth of biofuel crops, such as sugar cane and palm oil. This deforestation has a negative impact on biodiversity, as habitats like tropical rainforests are removed. Furthermore, competition for water in regions with a scarcity can also cause further issues (16). Due to the pitfalls associated with first generation biofuels, research has now focused on the development of second generation biofuels.

1.2.2 Second Generation

Second generation biofuels differ from first generation by the ability to utilise any part of the plant, including the lignocellulosic biomass, as opposed to just converting the oil-rich or sugar-dominant parts (14). Therefore, municipal, wood process and forest waste can all be utilised as a lignocellulosic feedstock, as well as low cost crops, short rotation forest crops and vegetative grasses (16). The ability to utilise lignocellulosic biomass provides a huge advantage over first generation biofuels, as it avoids the depletion of human and animal feedstock, in addition to being the most abundantly available raw material on Earth. Furthermore,

second generation crops grown on equal amounts of land will produce a higher yield of energy than first generation crops, due to the increased amount of substrate utilised. Second generation crops are also capable of growing on land which is of poorer quality than that needed for food production, meaning competition with agriculturally farmed land is unnecessary (16). Moreover, food waste is becoming a newly utilised source of biomass. Currently 1.3 billion tonnes of food produced for human consumption is wasted each year. Not only is this an ethical nightmare, it also produced an estimated 3.3 billion tonnes of CO_2 in addition to the $750 billion financial loss (2). Therefore, if this waste source can be utilised, it becomes an additional source of biomass for fuel and chemical production, but also helps reduce the carbon footprint and tackles this unethical waste.

Lignocellulosic biomass is mainly comprised of lignin, cellulose and hemicellulose. Lignin is a biopolymer and is used to maintain the structural integrity of the plant and is typically not utilised in biofuel generation. Cellulose and hemicellulose are the abundant source of carbohydrate that is utilised as the sugar for second generation biofuel production (23). The pre-treatment of the lignocellulose biomass is often the first step of the fermentation process and is typically achieved through heating of the biomass in the presence of an acid or a base. This pre-treatment step can hydrolyse hemicellulose, breakdown cellulose and remove the lignin. This allows the specific enzymes that digest the feedstock to gain access to the individual parts for fermentation (18).

Second generation biofuels, however, have previously been restricted by the absence of microorganisms that can sufficiently metabolise the hydrolysate completely. This is a significant problem, as pentose sugars are abundant in the hemicellulose portion of lignocellulose and are substantially produced when hydrolysed (24). Xylose, a pentose sugar, is the second most abundant carbohydrate found in lignocellulosic material. Therefore, the ability for an industrial strain to grow on xylose is essential to allow the production of biofuel from lignocellulosic biomass. The organism of choice for first generation biofuel production, *S. cerevisiae*, has a narrow range of fermentable sugars and cannot naturally utilise pentose sugars. Therefore, this presents a need to identify alternative strains that are capable of utilising a wide range of sugars, whist still offering suitability for industrial scale up (18). In addition to the currently pursued first and second generation biofuels, research into third and, more recently, fourth generation biofuels has now been initiated (25) (Figure 1.2).

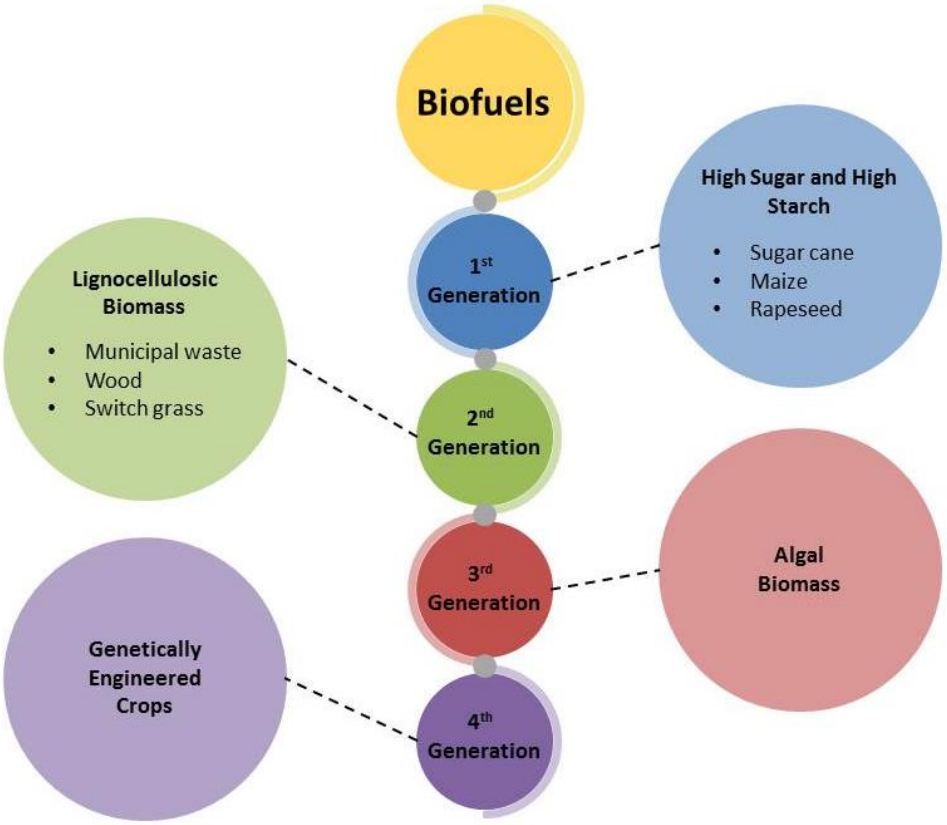


Figure 1.2 – An overview of the feedstock used for first, second, third and fourth generation biofuels.

1.2.3 Platform Chemicals

Platform chemicals are often carbon derivatives, with a chain length between 2 and 6, which contain attached functional groups that can be flexibly exploited. These compounds, such as lactic acid, succinic acid, fumaric acid and 3-hydroxypropionic acid, are important precursors that can be utilised to produce a wide range of chemicals and materials. Platform chemicals are therefore exceedingly useful for multiple industries and are used in a multitude of areas, such as for fuel and fuel

additives, solvents, polymers, food additives, pharmaceuticals and cosmetic ingredients (26,27). The importance of these compounds is demonstrated by a market value of $15 billion (28) for C4 dicarboxylic acids and their derivatives alone.

As previously mentioned, these platform chemicals are overwhelmingly produced through the petrochemical industry and therefore are non-renewable. Originally, the chemical industry was developed using coal as a feedstock. The existing coal reserves are higher compared to oil, meaning that production could theoretically revert back; however, this would not solve the problem of carbon emissions on health and the environment (29). Fortunately, like biofuels, these platform chemicals are being increasingly produced from renewable biomass resources (30,31). Many of these chemicals are naturally occurring metabolic intermediates and can be enhanced or synthetically produced through the metabolic engineering of microorganisms (Figure 1.3).

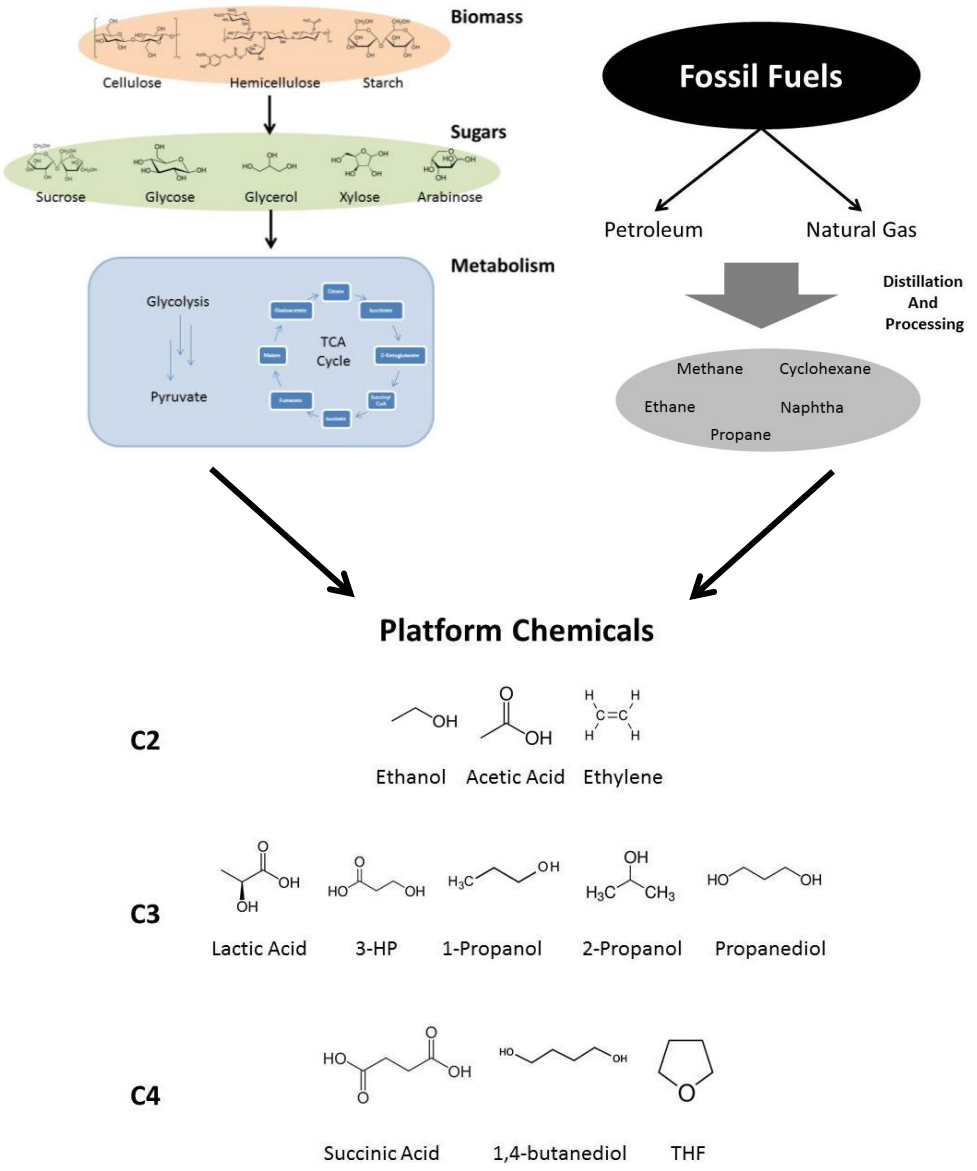


Figure 1.3 – An overview of the production of platform chemicals from renewable biomass (left) and from non-renewable fossil fuels (right).

1.2.4 Use of Mesophiles

Production of many of these useful products occurs naturally within many microorganisms. Metabolic engineering of these microorganisms can allow the enhanced production through the genetic modifications of

existing metabolic pathways. Production of fuels and chemicals using microorganisms is not a recent idea, with the successful construction of several previous *Escherichia coli* strains previously achieved to selectively produce ethanol (32). Original strain development was accomplished through multiple rounds of random mutagenesis. However, the expansion of genetic engineering has enabled simple gene deletions and gene insertions, which can facilitate metabolic engineering. Engineering on several Gram-negative mesophiles, such as *E. coli*, *Klebsiella oxytoca*, and *Zymomonas mobilis* (35) has been effective and strains have been developed to successfully utilise pentose sugars. *E. coli* is a popular industrial organism of choice, offering the advantage of not requiring complex growth factors and more importantly, the ability to ferment a wide array of sugars. However, the major disadvantage surrounding this organism is its narrow, neutral pH range necessary for growth (35), a negative attribute considering the process of fermentation produces an array of acids. *Z. mobilis* is another organism that is highly attractive to industry, due to its high yields and tolerance of ethanol, which is capable of producing a higher yield than even *S. cerevisiae* (36). Although *Z. mobilis* doesn't ferment pentose sugars naturally, it has been genetically modified to metabolise xylose through the expression of exogenous *E. coli* genes (37). However, the use of this organism has been unsuccessful to date as no strain has been developed to be used at an industrial level. The most desirable bacteria strains are those that are able to withstand low pH and high ethanol concentrations, as these conditions prevent the culture from being susceptible to easy contamination (18). Typically, for microbial

fermentation to occur, the lignocellulose, which is comprised of cellulose and hemicellulose, must be pre-treated (38). *E. coli* is able grow directly on the pre-treated lignocellulose (39,40); however, this is with a reduced final yield. This is due to an inhibitory effect caused by compounds released from the pre-treatment of lignocellulose (41). In addition, cellulases and hemicellulases, necessary for the breakdown of lignocellulose into sugars, are often enzymatically slow at temperatures below 50 °C, leading to incomplete hydrolysis.

1.2.5 A Move towards Thermophiles

Despite the wealth of research dedicated to model organisms such as *E. coli*, there are still clear issues. Consequently, focus is now beginning to shift towards thermophilic bacteria, which offer an attainable alternative to mesophiles and possess distinct advantages, such as catabolic flexibility and other beneficial physiological characteristics. Thermophiles, unlike mesophiles such as *S. cerevisiae*, are capable of fermenting both pentose and hexose sugars of biomass hydrolysate (17). Furthermore, organisms that are able to metabolise both pentose and hexose sugars are often subjected to carbon catabolite repression, whereby xylose uptake is repressed until glucose is depleted (42-44). However, these regulation pathways have been demonstrated to be independent of one another in *Thermoanaerobacter* (45) which also may be the case in other thermophiles. Generally, thermophiles are also more robust

microorganisms; they are able to withstand fluctuations in temperature, pH and environmental change, traits that are often deemed essential for a commercially viable process. Furthermore, an increase in temperature causes a decrease in gas solubility and this inverse relationship has two advantages; it helps maintain a near anaerobic environment, which is ideal for the fermentation process, and also reduces the ability for obligatory aerobic contaminants to establish (17). Additionally, the high temperatures optimum for the growth of thermophiles also facilitates the effective recovery of volatile chemicals, such as ethanol, through continuous distillation of the product, which helps to cut running costs (17). This also aids in avoiding the issue of end-product inhibition. Further reductions in cost can be achieved by keeping a constant temperature level above 50 ˚C, as it removes the need for an energy input to continually cool the reactor between successive cycles of the traditional fermentation process. Microbial contamination is another substantial problem faced by mesophilic biofuel production. Competition for the available carbon source from contaminating bacteria reduces the ethanol yield of the process (17,46,47). To prevent contamination antibiotics are necessary, which drives the cost of production higher. Thermophilic strains, however, offer a significant advantage, as their optimum growth temperatures are too high for common feedstock derived contaminants to grow (17,48). Furthermore, the optimum temperature of most cellulases is between 50 ˚C and 55 ˚C. This high temperature range also facilitates the physical degradation of cellulose, therefore suggesting cellulose as an ideal feedstock for thermophiles. Indeed, some thermophiles already

possess the ability to degrade cellulose (49), demonstrating that genetic engineering for feedstock utilisation would be minimal. Therefore, by using temperatures in this region, simultaneous fermentation, saccharification and cellulose hydrolysis can be achieved (7). Yet as attractive as thermophiles appear for the production of biofuels, they are not without their own caveats. Despite their strengths, the major downfall for the use of thermophiles in biofuel production is the typically low ethanol yield and tolerance. Previously, this has caused major issues, as the genetic tools needed to engineer these thermophilic strains were not available. However, recent progress has facilitated the arrival of techniques that allow genetic manipulation of these species. For example, *Thermoanaerobacter ethanolicus* has been engineered to tolerate high ethanol concentrations, whilst still being able to replicate (17). Nevertheless, working with thermophiles is technically challenging. The lack of tools available, the higher working temperatures, potentials of different codon usage systems and generally less knowledge available, all cause a possible barrier to entry for these systems (49-51).

1.3 The *Parageobacillus* and *Geobacillus* Genus

Thermophiles, such as species belonging to the *Parageobacillus* and *Geobacillus* genus are a source of many important thermostable enzymes,

which are significant in field of biotechnology and have changed the face of microbiology with processes such as PCR. The metabolic versatility of thermophiles enables exploitation of these bacteria as whole cell biocatalysts, maintaining the ability to both hydrolyse and ferment biomass, allowing consolidated bioprocessing. *Parageobacillus* species are not fastidious organisms and have the ability to utilise many different carbon sources and subsequently convert them into desirable metabolites. This allows a reduction in the number of steps to convert lignocellulosic biomass into biofuels (52). *Parageobacillus* species also exhibit a fast growth rate, which facilitates reaching a high cell density rapidly and therefore the ability to produce the required products swiftly. However, in order to fully exploit these species, reliable and reproducible genetic engineering is essential. Genetic tools for these species are limited, but the necessary genetic manipulation tools and protocols are developing rapidly (52). Even though genetic manipulation of *Parageobacillus* species is in initial stages of development, there are already several strains with established transformation protocols. To date, the most commonly used strains are *Geobacillus stearothermophilus* NUB36, *Parageobacillus thermoglucosidasius* NCIMB 11955 and *Geobacillus kaustophilus* HTA426 (52).

1.3.1 *Parageobacillus thermoglucosidasius*

P. thermoglucosidasius is a thermophilic, Gram-positive bacterium in the phylum *Firmicutes*. First isolated from soil samples in Japan, *P. thermoglucosidasius* was originally classified as *Bacillus thermoglucosidasius* (53). The genus *Bacillus* is a large and diverse collection of bacteria, containing both aerobes and facultative anaerobes that are Gram-positive, rod-shaped and capable of forming endospores. The extreme diversity of this group meant that it contained species that had very broad requirements and metabolic variety. Ash *et al.* (54) aimed to determine the phylogenetic structure of the genus, through the comparison of the 16S ribosomal RNA (rRNA) sequences of over 50 bacilli species by reverse transcription (54). This sequence data sub-divided the genus *Bacillus* into five distinct groups, placing *B. thermoglucosidasius* into group five. Simultaneously, at oilfields in China, Kazakhstan and Russia, several hydrocarbon-oxidising strains had been isolated and analysed, and the initial characterisation of these species placed these too into group five of the genus *Bacillus*. After further in-depth characterisation of the isolates, including fatty acid composition, physiological and cultural features and DNA comparison of other thermophiles of group five, it was decided to establish this group as a new genus called *Geobacillus*, which contained *B. thermoglucosidasius* amongst others (55). *P. thermoglucosidasius* is a facultative anaerobic bacterium that grows optimally at 60 °C (53). However, it has been show that strains can grow at temperatures anywhere between 37 °C and 68 °C (55).

P. thermoglucosidasius is one of the few thermophiles that can be genetically manipulated and is currently being targeted as an industrial microorganism capable of producing biofuels from lignocellulose. Like many *Geobacillus* species, *P. thermoglucosidasius* has both a fast growth rate and is able to ferment a wide variety of sugars. However, it also suffers the disadvantage of being unable to grow on crystalline and amorphous cellulose and therefore at the moment, to utilise this strain, initial enzymatic pre-treatment of the feedstock is necessary (56). In a push to use these advantageous thermophilic strains, *P. thermoglucosidasius* was initially genetically engineered to produce ethanol as a primary product during fermentation, instead of the mixed acid profile that is normally produced (24). This strain, TM242, was produced by elimination of two pathways, lactate dehydrogenase (Ldh) and pyruvate formate lyase (Pfl), by disruption of the genes lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*) respectively. In addition, the pyruvate dehydrogenase (Pdh) gene (*pdh*) was upregulated to produce the ethanol producing strain. As lactate is the predominate product formed during a typical mixed acid fermentation profile, inactivation of *ldh* was necessary to shift to a predominantly ethanol profile, which was successfully observed. However, an increase of pyruvate was also observed, suggesting that the pyruvate formate lyase and the pyruvate dehydrogenase pathways were saturated. Therefore, a double mutant was instead produced by upregulating *pdh*, using a strong *ldh* promotor (P_{ldh}) isolated from *G. stearothermophilus*. This double mutant facilitated an additional increase in ethanol production and a

decrease in pyruvate build up. Finally, the strain was completed by disrupting *pfl*, to eliminate the increased production of formic acid (24). Production of ethanol from the strain TM242 resulted in 95% of theoretical yields, using cellobiose as a carbon source. More recently, the strain developed by Cripps *et al.* (24), TM242, enabled ethanol production in the region 9.9 ± 0.4 g/L (92 % yield), from palm kernel cake (PKC), a mannan-rich by-product from palm oil processing (57). Mannose oligomers, produced by saccharification of PKC, often require degradation by mannosidases prior to fermentation. However, Raita *et al.* (57) demonstrated that TM242 could produce ethanol without mannosidase supplementation. Furthermore, TM242 was able to produce more ethanol than *S. cerevisiae* (35.4 %) utilising the same feedstock. In addition to ethanol production, using *P. thermoglucosidasius*, Lin *et al.* (7) has also demonstrated isobutanol production at 50 ˚C, using both glucose and cellobiose as the feedstock.

1.4 Genetic Engineering Tools

1.4.1 Transformation

The development of genetic tools is an essential requirement, but to enable genetic engineering of an organism it is also essential to be able to deliver exogenous DNA. Frequently, delivery of DNA occurs in the form of plasmids and is often achieved through methods such as transformation,

conjugation, or transduction (58). Bacterial conjugation relies upon direct cell-to-cell contact, in order to transfer the genetic element from the donor cell to the recipient cell; whereas transduction introduces exogenous DNA using viruses. Transformation, however, does not need cell-to-cell contact for the uptake of exogenous DNA. It requires the induction of competence and allows the DNA to enter through the membrane of the cell. Transformation often occurs in one of two ways, electroporation and heat-shock. Heat-shock involves the treatment of cells with calcium chloride before briefly heat-shocking the cells. The mechanism of action is still undetermined. It is proposed that the Ca^{2+} divalent cation may play a role in adhering the DNA to the membrane (59); however, there is some suggestion that the heat-shock step is actually unnecessary (60). Alternatively, electroporation transformation transfers DNA across the cell membrane via an electrical current. The process of electroporation works through perturbing the charged membrane of the bacteria, by initiating a high voltage electric pulse. This charge interaction causes temporary hydrophilic pores in the membrane, which allows the uptake of charged molecules such as DNA (61). Competency for transformation is achieved by subsequent washes with a high osmolarity washing buffer. Natural competency of Gram-positive thermophiles is uncommon, showing high intractability to genetic transformation due to the thick peptidoglycan cell wall and low permeability of the plasma membrane (52). However, early attempts at transformation using protoplasts in *Parageobacillus* species were successful (62,63). In more recent years,

due to advances in the technique, the majority of transformations carried out in thermophiles are achieved through electroporation (24).

Vectors are a tool in molecular biology, used as the vehicle to transfer DNA into a cell. As mentioned, plasmids are one form of vector and are small self-replicating carriers of genetic material. Other forms of vectors include viral vectors and cosmids. Vectors are necessary for introducing exogenous DNA and often contain independent replicons, a multiple cloning site and selection markers for maintenance. Genetic engineering often makes use of shuttle vectors that contain two origins of replication suitable for two different host bacteria. Often construction of a plasmid occurs in an amenable organism, such as *E. coli*, before being transferred to the organism of choice. These origins of replication are often found in native plasmids from the species. Unfortunately for *Parageobacillus* species there are few plasmids that have been isolated and described (52). However, promiscuity is observed between thermophilic replicons of different plasmids, as well as between Gram-positive bacteria (64). This is the basis for the pUB110 replicon used by Cripps *et al.* (24), which was isolated originally from *Staphylococcus aureus*. Often, many parts are required to successfully construct a viable vector. In addition to the origin of replication, selection markers that are suitable for both organisms are also necessary.

1.4.1.1 Restriction Modification Systems

A restriction modification (R-M) system can be viewed as a primitive immune system that protects bacteria from invading foreign DNA (Figure 1.4). These systems are often a barrier to successful transformation and need to be circumvented. These successful defence mechanisms utilise a restriction endonuclease (REase) to digest DNA at a specific recognition sequence. The R-M systems are able to distinguish the invading DNA from its own genome by the action of a methyltransferase (MTase), which is the second enzyme associated with the R-M systems. The MTase epigenetically methylates the DNA at any restriction recognition sequence present within its own genome, in order to protect from the REase (65). There are four different types of R-M systems, which are distinguished by sequence recognition and subunit composition. Type I systems consist of a protein complex comprising of three polypeptides, the MTase, REase and a specificity subunit (66). In contrast, Type II R-M systems contain two distinct polypeptides, one MTase and one REase, acting independently but recognising the same recognition sequence (67). Type III enzymes differ from the previous two types, in that the enzyme complex containing MTase and REase also contains DNA-dependent nucleoside-triphosphatase (NTPase) activities, which consequently results in competition between methylation and restriction (68). Finally, Type IV systems are the most significantly different, as they cleave DNA substrates

containing methylated, hydroxymethylated, or glucosyl-hydroxymethylated bases at specific sequences (69).

Two other defence systems that have been observed to inhibit foreign DNA from entering the cell within prokaryotes are CRISPR/Cas, a nuclease that confers an adaptive immunity, and the RecBCD enzyme, a helicase that also exhibits nuclease activity which can restrict foreign DNA (70). RecBCD can also aid in DNA repair of the host genome through recombination (71).

This coevolutionary arms race between bacteria and phages has caused the development of these defence systems, which inhibit the multiplication and spread of phages. Therefore, phages have developed methods to avoid the R-M systems of bacteria, which include phages that encode MTases (72,73), the attachment of bulky groups by hydroxymethylation (65), and through proteins that inhibit the restriction enzymes employed by the bacteria (74). Furthermore, ssDNA is resistant to REases, which requires double stranded DNA (dsDNA) for recognition and cleavage, and so viruses that contain ssDNA are protected against certain R-M systems (75).

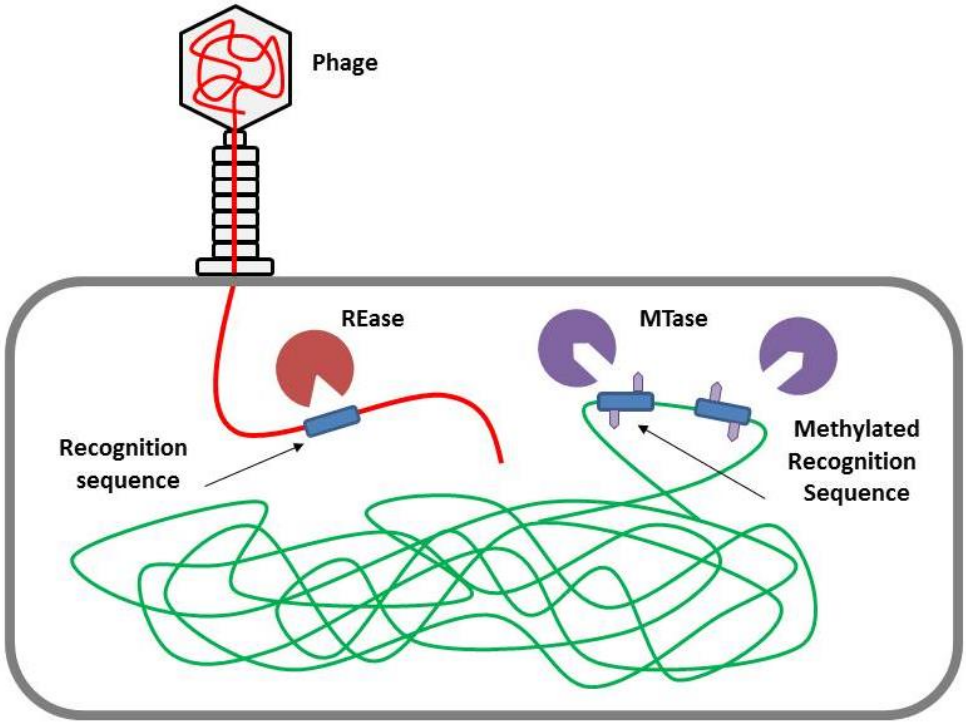


Figure 1.4 – An overview of restriction modification (R-M) systems as a defence mechanism. The recognition sequence on the virus DNA, or any invading DNA, is cut by the restriction endonuclease (REase). Any recognition sequences on the genome are methylated by the methyltransferases (MTase). Methylated recognition sequences are not cut by the REase.

1.4.2 Gene Expression Elements

With the ability to introduce exogenous DNA into the cell, manipulation can occur through the overexpression of genes, both chromosomally and on a plasmid, and chromosomal knock-outs (KOs) and knock-ins (KIs). In order to carry out these functions, gene expression elements are essential to drive and regulate the expression of the required genes and proteins. Through the fine tuning of transcription and translation, the levels of desirable metabolites and other cellular properties can be affected. These

expression elements include promoter, ribosomal binding sites (RBS) and transcriptional terminators.

Promoters are regions of DNA that RNA polymerases bind to, to initiate transcription of a gene. These sequences are located upstream of the gene and are often located near the transcriptional start site. Each promoter sequence is unique, though bacterial promoters often contain two conserved consensus sequences upstream of the transcriptional start site. These are typically at positions -10 and -35 and have the sequences TATAAT and TTGACA respectively. RNA polymerase binds to the promoter to begin transcription to produce mRNA. Often, transcription factors, such as activators and repressors, associate with promoter regions to promote or repress transcription. A library of promoters is usually needed to fine tune expression, using strong, weak, inducible and constitutively expressed promoters. Many promoters are known for Gram-negative bacteria, such as *E. coli*; however, the production of new libraries is required for Gram-positive bacteria (76).

The RBS, or Shine-Dalgarno (SD) sequence in prokaryotes, is a sequence of DNA upstream from the gene that the ribosome binds to, to begin translation of the mRNA. The SD sequence is often purine-rich and is complementary to the anti-SD sequence, UCCU, at the 3'-end of 16S rRNA. The SD consensus sequence does vary, but often follows a similar structure to 5'-AGGAGG-3'. The SD sequence is upstream of the start codon and can vary in both length and nucleotide composition, which in turn

varies the activity of the RBS. Interactions between ribosomal subunits and transfer RNA (tRNA) allow the synthesis of a polypeptide sequence from the mRNA polymer. Therefore, in addition to promoters, the RBS sites are a crucial tool for successful gene expression within an organism.

Translational efficiency is also affected by the redundancy of the genetic code. The majority of amino acids are coded for by more than one codon. This flexibility offers the opportunity for evolutionary processes to affect the efficiency and accuracy of translation, without affecting the final amino acid sequence. Although each codon is theoretically 'synonymous', the levels of each tRNA and the speed of the ribosome to recognise them differ, offering a selection advantage to certain codons (77-80). It has been shown in many organisms, including *E. coli* and *S. cerevisiae*, that codon bias is correlated to gene expression (81-85). Furthermore, the variation in codon selection means that the nucleotide sequence of a messenger RNA (mRNA) transcript can vary drastically, giving rise to differences in stability and secondary structure, both of which may affect translation (86). This could therefore lead to a codon optimised system within an organism, which likewise may vary from organism to organism. Understanding the codon usage bias is a key element in synthetic biology. When expressing a foreign gene, the protein expression level maybe hampered if the codons selected are underrepresented in the organism.

Another genetic component used in metabolic engineering and gene expression is a transcriptional terminator. A terminator is a sequence of

DNA at the end of a gene or operon that mediates transcriptional termination. These terminators can eliminate interference produced from upstream promoters, which if not resolved can have detrimental 'read-through' effects on downstream promoters. There are three mechanisms by which this 'read-through' can affect downstream transcription, specifically, occlusion, 'sitting-duck' interference and transcription factor inhibition. Occlusion is the blocking of the downstream promoter by the RNA polymerase of the upstream promoter. Similarly, 'sitting-duck' interference refers to the RNA polymerase of the upstream promoter physically dislodging the downstream RNA polymerase (87). In similar fashion to 'sitting-duck' interference, transcription factor inhibition is where the transcription factors involved in initiating transcription of the downstream promoter are removed by the upstream RNA polymerase. Occlusion is expected to produce the greatest effect when the upstream promoter is much stronger than the downstream promoter (88). Furthermore, it has been shown that gene expression may be increased by placing a terminator between promoters, by improving the stability of the mRNA (89), reducing these effects and increasing transcription (90).

Termination can occur both with and without external interaction with a protein. There are two classes of transcriptional terminators common in prokaryotes, Rho-dependent and Rho-independent. Rho-dependent terminators require a protein called Rho factor. Rho disrupts the transcription complex by hydrolysing adenosine triphosphate (ATP), in order to translocate down the mRNA and come into contact with the RNA

polymerase complex. On contact with Rho, the complex is disrupted through allosteric effects. In contrast, Rho-independent terminators, also known as intrinsic terminators, stall the mRNA-DNA-RNA polymerase complex, through the production of a self-annealing, GC rich, hairpin structure followed by a poly-T tail. This hairpin structure disrupts and stalls the RNA polymerase on this reduced stability region, terminating transcription.

1.4.3 Reporter Systems

Reporter systems are non-native genes that are introduced into a cell, eukaryote or prokaryote, to produce a characteristic response which is easily identified or quantified. These reporter systems can vary in their response, from conferring resistance through a selectable marker to a visual response from a fluorescent reporter. Common reporters include, the green fluorescent protein (GFP) gene (*gfp*), β-galactosidase (LacZ), the *lux* operon and the chloramphenicol acetyltransferase (CAT) gene (*cat*). These reporter genes are all detected using different assays, specifically, fluorescence, enzymatic, luminescence and resistance respectively. GFP is a fluorescent protein that emits when exposed to blue and ultraviolet light. This gene was first identified in the jellyfish *Aequorea victoria* (91). GFP has been applied to many areas of research applications (92,93) and can be used not only as a reporter gene to determine gene expression (94,95) but also as a localisation tool to elucidate protein localisation and

hence determine cellular processes via fluorescence microscopy (94,96,97). LacZ is often used as a reporter marker for recombinant clones and works by cleaving the glycosidic bond in the substrate 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside (X-gal). Cleavage of this bond causes the production of 5-bromo-4-chloro-3-hydroxyindole, which dimerizes and oxidizes to 5,5'-dibromo-4,4'-dichloro-indigo and forms a blue colour that can be visually observed (98). However, when *lacZ* is disrupted by integration of a gene, this reaction cannot occur. This forms the basis of blue/white screening of recombinant bacterial colonies. Luciferase reporter genes are similar to fluorescent genes, but are instead enzymes that produce bioluminescence. This is often in the form of a *lux* operon, which encodes genes necessary for self-regulation and for the production of luminescent proteins, which can be used in bioluminescent imaging (99,100). Another class of reporters are resistance genes, such as *cat*. CAT confers resistance to the antibiotic chloramphenicol, through acetylation, and therefore can demonstrate evidence of the uptake of exogenous DNA through selection. Furthermore, by incubation with radioactive [^{14}C]chloramphenicol, expression can be quantified by measuring the acetylation of [^{14}C]chloramphenicol (101).

1.4.4 Inducible Systems

Inducible expression systems regulate gene expression, often by an external factor, allowing selective induction or repression. Examples of

this include the Tetracycline (tet) inducible system or the *lac* inducible system. In the tet system, the Tet response element (TRE) is bound by the tetracycline repressor (tetR), inhibiting transcription. In the presence of tet, tetR binds tetracycline and is released from the TRE, allowing transcription to occur. This system has been modified to produce both the tet-on and tet-off systems. The tet-off system utilises a tetracycline transactivator (tTA), a chimeric protein comprising of tetR and the domain VP16, a viral transcriptional activator. The tTA instead facilitates transcription in the absence of tet and is repressed on addition (102). The tet-on system exhibits the same phenotype as the WT system, using tet as an activator. By mutating tetR, the rTetR was produced, which facilitated induction of expression through the addition of tet. With the fusion of VP16 to the rTetR, a transactivating tet system was produced (103). Alternatively, the *lac* operon is induced by the presence of lactose. In the absence of lactose, the operator is bound by the *lac* repressor, blocking the RNA polymerase from binding. The *lac* repressor is released in the presence of lactose, allowing transcription of the operon. After degradation of the lactose by the *lac* genes, the *lac* repressor can then bind to the operator to stop transcription (104). Repurposing of this system for inducible expression has seen the use of isopropyl-β-D-thiogalactopyranoside (IPTG) as a successful inducer of the *lac* operon. However, this is not recognised by β-galactosidase, allowing a continuous expression response (105). By utilising inducible systems, expression of introduced genes can be efficiently controlled and regulated at will. The ability to control expression is often advantageous within systems utilised

for either overexpression or expression of a toxic compound. Successfully achieving high levels of growth prior to induction can circumvent deleterious growth conditions caused by the physiological shock of overexpression or exposure to a toxic compound.

1.4.5 Manipulation of Bacterial Genomes

The ability to manipulate an organism, by knocking-in and knocking-out genes of interest, is one of the key foundations of genetic engineering. Changing the phenotype of an organism or diverting carbon flux through a specific pathway is essential when engineering a strain to produce a specific product or metabolite. Integrating DNA into a microorganism's genome is vital as, although exogenous DNA can be introduced through replicative plasmids, these are inherently unstable and limited in their use. To preserve these mobile genetic elements, a selection pressure is often applied to maintain the plasmid (106). This is typically in the form of antibiotics, which on an industrial scale will increase costs significantly. However, if integrated into stable DNA, either the chromosome or another stable genetic element, exogenous DNA can be maintained without these external pressures (106). Early approaches to genetic modification involved using bacteriophage λ, utilising its native ability to integrate into the chromosome (107,108). Other classical methods require the use of a replication deficient plasmid which is only maintained through homologous recombination. These plasmids often contain an antibiotic

resistance marker and are therefore obligated to undergo homologous recombination to ensure preservation (109-111). Homologous recombination can also be achieved using linear dsDNA in yeast. *S. cerevisiae* is very proficient at double-strand break repair recombination, which is stimulated by linear dsDNA (112). However, using linear dsDNA within bacteria is much more of a challenge, due to rapid degradation by nucleases. Recombination can be increased through nuclease deficient strains such as RecBCD, which has a mutation in the nuclease domain *recD* but retains recombination activity (113). Additionally, one-step homologous recombination can be achieved using naturally competent bacteria such as *Bacillus subtilis*, through allele exchange or gene replacement. This method utilises a selectable marker, usually an antibiotic resistance marker, alongside the gene of interest within an allele exchange cassette (114). Furthermore, multiple sequences, or sequences that are sufficiently large enough, can be inserted at a single locus through a series of alternating integration steps, using two selectable markers (115). Other possible methods used to achieve gene integration besides homologous recombination include site-specific recombinases, such as Cre-loxP, and transposon-mediated gene transposition.

Currently, the most commonly applied method involves the use of a plasmid with a homologous recombination cassette containing two homologous regions of DNA. The whole plasmid can integrate in a single-crossover event. However, by using two separate homologous regions from upstream and downstream of the targeted DNA, the insert alone can

be integrated via a double-crossover event. This method of KI using homologous recombination is also applicable for KO. Instead of inserting a gene of interest, homology arms containing a selectable marker would replace a gene targeted for deletion.

1.4.6 Allele Coupled Exchange

The isolation of double-crossover clones can be difficult to achieve and requires an extensive screening process, hence counter-selection markers are required to overcome this problem. Allele coupled exchange (ACE) is a recently developed technique, originally demonstrated in *Clostridium difficile*, which facilitates rapid insertion of heterologous DNA into a genome (116). The technique exploits both the inactivation and activation of *pyrE* to produce either a positively or negatively selectable response (117). The *pyrE* gene encodes the enzyme orotate phosphoribosyltransferase (PyrE), which is involved in the uracil biosynthesis pathway (Figure 1.5) and is essential for the production of uracil and consequently cell survival. PyrE is responsible for converting orotic acid, a pyrimidine intermediate, to orotidine 5'-monophosphate (OMP), which is subsequently converted into uridine 5'-monophosphate (UMP). In addition to the uracil biosynthesis pathway, PYRE also has the ability to convert 5-fluoroorotic acid (5-FOA), an analogue of orotic acid, to 5-fluorouridine monophosphate (5-FUMP). This accumulation of 5-

FUMP is highly toxic to the cell and leads to cell death, hence cells lacking in *pyrE* are resistant to 5-FOA (5-FOA^R). It is therefore possible to grow cells lacking in *pyrE* on 5-FOA, in order to counter select against WT cells.

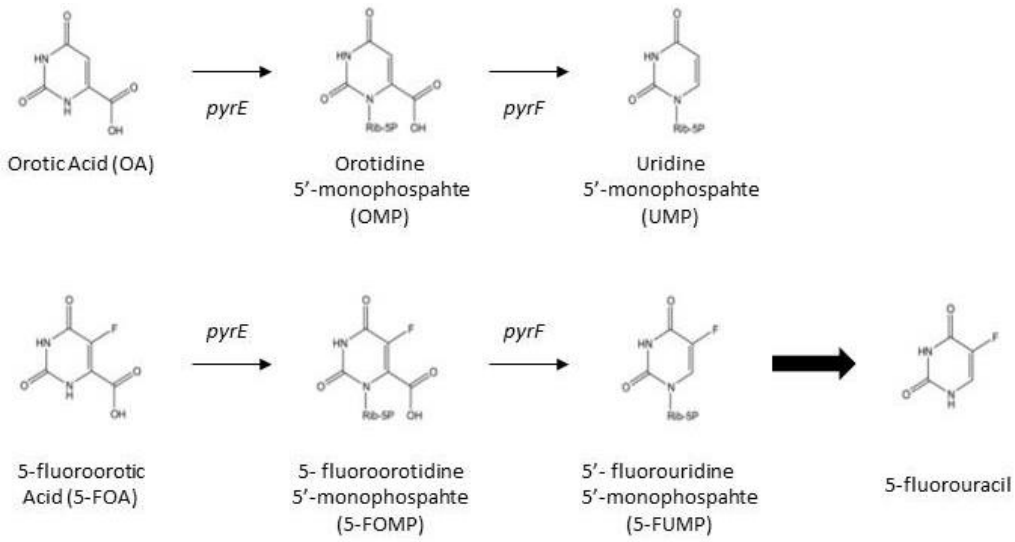


Figure 1.5 – A schematic detailing the biosynthesis of uracil and the subsequent conversion of 5-fluoroorotic acid (5-FOA) to 5-fluorouridine monophosphate (5-FUMP), which is toxic to the cell.

This technology has been successfully implemented within *P. thermoglucosidasius* (118). A *P. thermoglucosidasius* strain, lacking in *pyrE*, was constructed through two sequential recombination events. This used a small left homology arm (LHA), which consisted of the first 300 bp of the *pyrE* gene, and a larger right homology arm (RHA) of 1200 bp, which was downstream, and included the stop codon, of the *pyrE* gene. Through the use of unequal HA lengths, the order of the recombination events was successfully controlled. It was shown that the large HA downstream of the gene was used to encourage single-crossover recombination events in the correct region, without disrupting the *pyrE* gene. Subsequently, the second recombination event then occurred at the 300 bp start of the gene and

created a 5-FOA^R double-crossover mutant, a phenotype which could be selected for through growth on 5-FOA. This *pyrE* negative strain is auxotrophic for uracil and therefore must be supplemented into the media to allow for growth. To integrate DNA into the organism, the exogenous DNA is positioned between the downstream, large homology arm and an upstream homology arm that allows the repair of the *pyrE* gene. Therefore, a successful double recombination event would recreate a fully intact *pyrE* gene, along with the gene of interest successfully integrated downstream, allowing selection for recombinant cells on media without uracil (Figure 1.6). This method of gene integration is a part of the roadmap formulated at the Synthetic Biology Research Centre (SBRC), University of Nottingham (119).

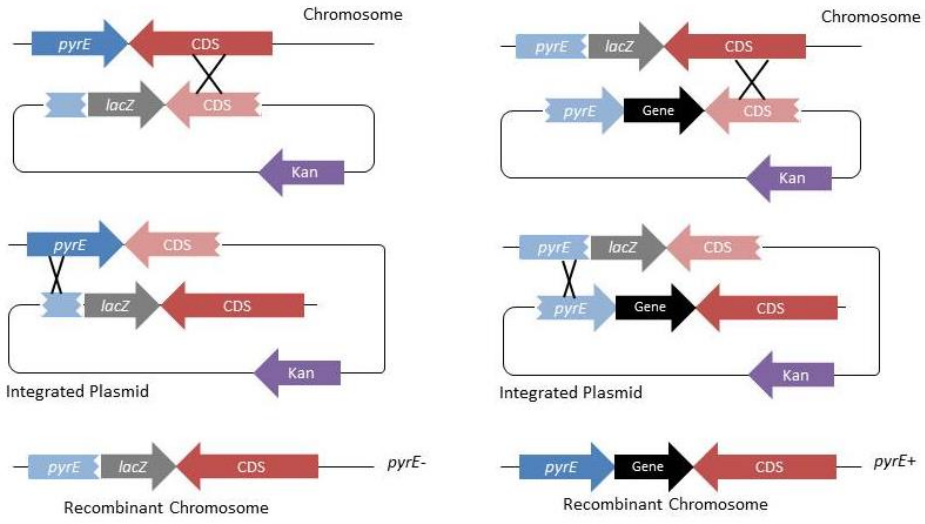


Figure 1.6 – A schematic representation of allele coupled exchange (ACE), utilising the positive and negative selection marker *pyrE*. The left schematic represents the production of a uracil auxotroph and 5-fluoroorotic acid resistant mutant (FOA^R). The right schematic represents the integration and repair of the *pyrE* locus to produce a uracil autotrophic and 5-fluoroorotic acid (FOA) sensitive mutant, in addition to the integration of the gene of interest.

1.5 CRISPR/Cas9

1.5.1 Prokaryote Adaptive Immunity

Clustered regularly interspaced short palindromic repeats, more commonly referred to as CRISPR, and CRISPR-associated proteins (Cas) have revolutionised genome editing. CRISPR are short palindromic repeats followed by a short DNA spacer, first observed by Ishino *et al.* (120) within the genome of *E. coli*. *cas* genes were later identified adjacent to these repeats and were suggested to have a functional relationship (121,122). These repeats were subsequently demonstrated to have homology to invading exogenous DNA, such as phages and plasmids (123,124). The CRISPR locus was later shown to be a prokaryote adaptive immune system that can identify and selectively cut exogenous DNA (125) and is observed in nearly 50 % of the genomes of sequenced bacteria and nearly 90 % of the genomes of sequenced archaea (126).

Within prokaryotes there are two classes of CRISPR/Cas systems, distinguished by the *cas*. Class 1 CRISPR/Cas systems, which encompass Types I, III, and IV, employ a multi-CAS protein complex for interference. In contrast, class 2 systems, which include Types II, V, and VI, only require the action of one multi-functional Cas protein, with the simplest and most widely used being Cas9 from the Type II system (127-129).

Three essential components are necessary to allow Cas9 to efficiently and specifically cut foreign genetic material, specifically CRISPR RNA (crRNA), trans-activating crRNA (tracrRNA) and the Cas9 itself. For Type II CRISPR/Cas systems, there are three phases associated with CRISPR adaptive immunity, specifically acquisition of CRISPRs, processing or biogenesis to produce mature crRNA and the restriction or interference of the invading DNA. On exposure to invading DNA, the CRISPR/Cas system is activated and the nuclease Cas1 and Cas2 processes the DNA into small fragments, which are called protospacers. These short DNA spacers are incorporated into the CRISPR locus of the prokaryote genome. This CRISPR locus is comprised of multiple repeats of short palindromic DNA sequences interspaced by the short sequences acquired from previous exposure to foreign DNA. Protospacers are selected through a specific recognition sequence that is used to distinguish the genome of the bacteria from foreign DNA, called protospacer adjacent motifs (PAMs) (130). These PAM sequences are specific to the CAS9 protein from different species and are not found within the hosts' genome. Subsequently, the CRISPR locus is then transcribed to form pre-crRNA. The tracrRNA then hybridises to the repeat sequences on the pre-crRNA. Ribonuclease (RNase) III cleaves the pre-crRNA/tracrRNA complex to produce the complete crRNA, which interacts with the Cas9 protein (131,132). The Cas9 protein is a large, multi-domain, RNA-dependent endonuclease, which targets and cleaves dsDNA. The mature crRNA forms a complex with the Cas9 protein, to guide it to the complementary sequence of the foreign DNA (133-135). The CAS9 protein then degrades

the DNA through two nuclease domains, HNH and RuvC, which are positioned to enable the protein to cut both the complementary and non-complementary DNA strands respectively, to produce a double stranded DNA break (135) (Figure 1.7).

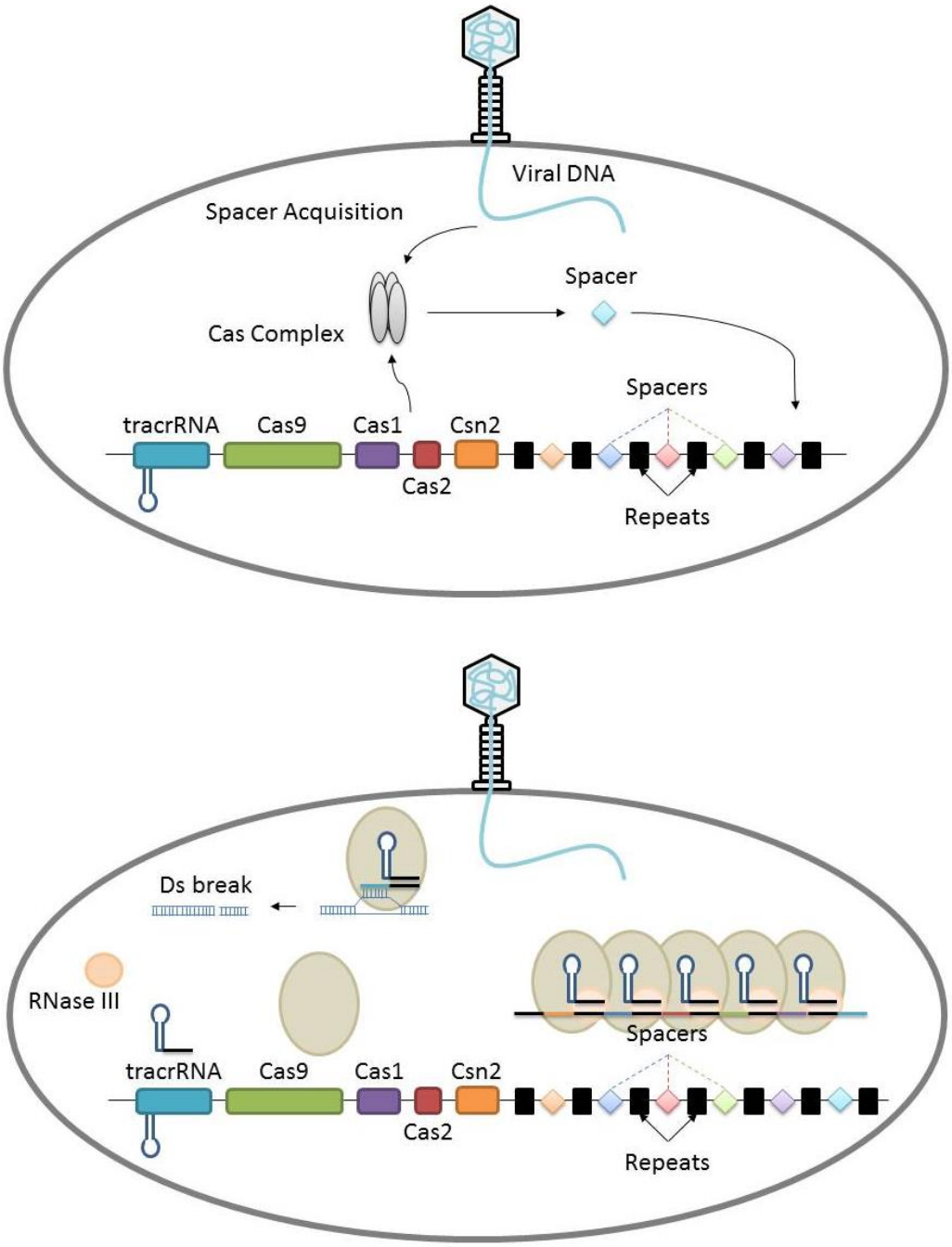


Figure 1.7 – A schematic representation of the adaptive immunity provided by Type II CRISPR/Cas9. The top image represents the immunity acquisition. The exposure to foreign DNA triggers the production of the *cas* genes, Cas1 and Cas2, involved with spacer acquisition which is integrated into the genome. During spacer acquisition within *S. thermophilus*, Csn2 is also used. The bottom image represents the immune response to the foreign DNA, producing a Cas9 complex which recognises and causes a double stranded break in the foreign DNA.

1.5.2 Application in Genetic Engineering

This immune response has now been harnessed as a way to selectively target locations in the genome for genome editing. Through transformation of a plasmid containing a Cas9 nuclease protein, in addition to a synthetic guide RNA (sgRNA), the genome of an organism can be cut at a specific location (136). A sgRNA is designed to be complementary to the target site, upstream of the specific PAM, and is used to guide the Cas9 protein to the targeted sequnce. The *cas9* from *Streptococcus pyogenes* is the most commonly used, with a PAM of 5'-NGG-3', making the sgRNA sequence 5'-N_{20}-NGG-3', where N represents any nucleotide. Genome editing can then be achieved using DNA repair systems, non-homologous end joining or homology directed repair. This system is often delivered on a plasmid containing a homology editing template, making *cas9* an efficient counter-selection marker for cells which have successfully undergone two-step homologous recombination (Figure 1.8). The use of this CRISPR/Cas9 system has advanced many areas of research due to its quick, easy and efficient application, compared to previously used genome editing technologies, such as Zinc-finger nucleases (ZFNs) and Transcription activator-like effector nucleases (TALENs).

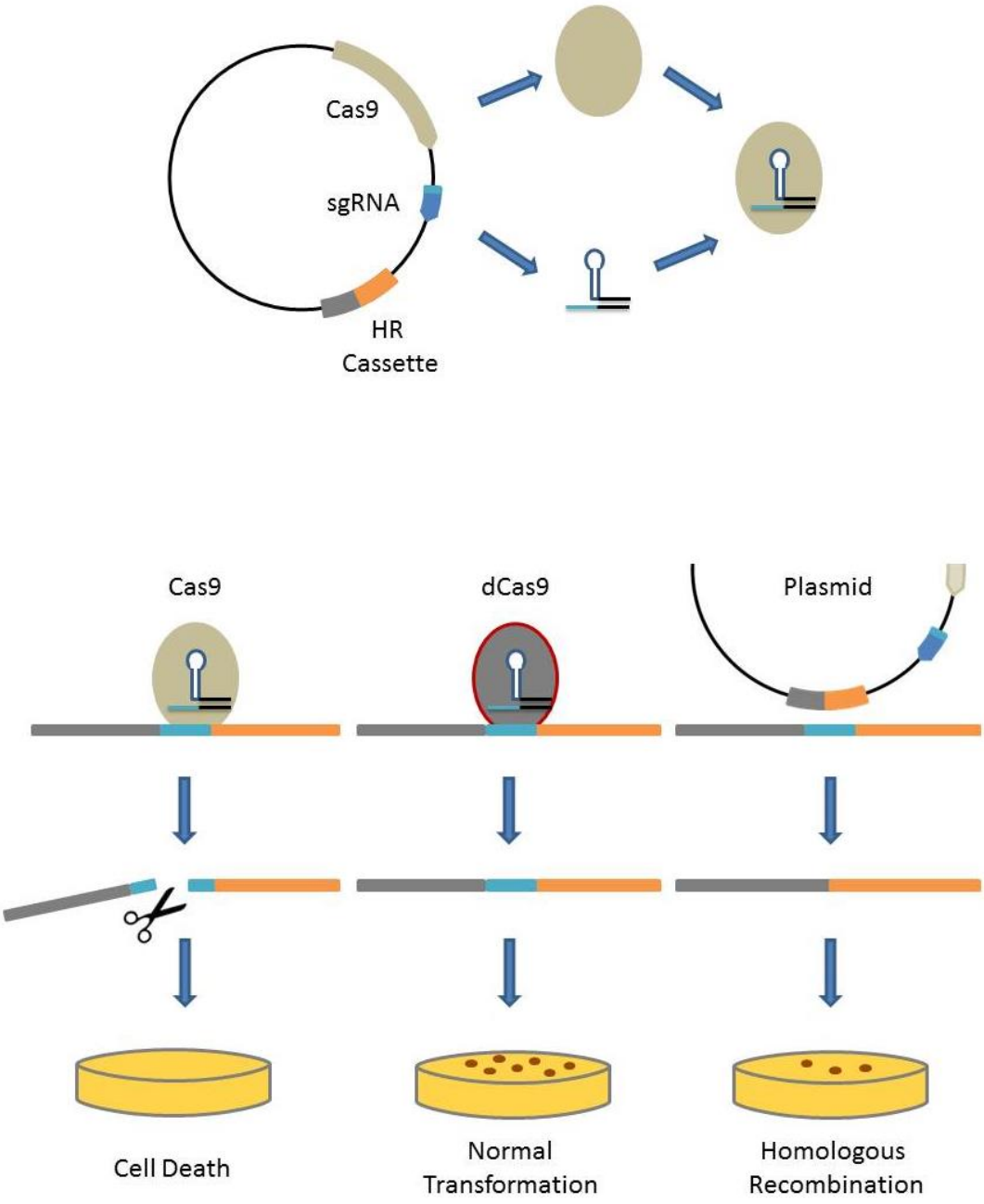


Figure 1.8 – A schematic representation Cas9-based editing in a one plasmid system (top), containing the *cas9*, synthetic guide RNA (sgRNA) and a editing template for allelic exchange. The bottom schematic shows transformation with an active *cas9* and no editing template (left), a catalytically dead *cas9* (dCas9) (middle) and a present editing template (right).

1.6 Succinic Acid

Succinic acid, or succinate, is an organic dicarboxylic acid that is naturally produced in many prokaryotes. Succinic acid is an industrially relevant platform chemical (see section 1.2.3) used in many different industries, such as food, chemical and pharmaceutical. Like many other platform chemicals, the production of succinic acid is mainly achieved through the petrochemical industry. Annually, it has been estimated that 245,000 tonnes of succinic acid is produced per year worldwide (137), through the catalytic hydrogenation of maleic acid (138). The U.S. Department of Energy determined that succinate is one of twelve high value building blocks targeted for production from biomass (137,139). This is due to the amazing versatility of succinate to produce a myriad of industrially important products and many important chemical derivatives, such as 1,4-butanediol and tetrahydrofuran (27,138,140).

1.6.1 Succinate Production Pathways

Succinate production in bacteria can occur through three different pathways, the tricarboxylic acid (TCA) or citric acid cycle, the phosphoenolpyruvate (PEP) pathway or the glyoxylate pathway (Figure 1.9).

The TCA cycle occurs aerobically and converts acetyl-CoA, produced from the carbon source such as glucose, into precursors of some amino acids, as well as generating reduced nicotinamide adenine dinucleotide (NADH). This NADH is then used to generate further ATP through the electron transport chain. Succinate is an intermediate in this pathway, generated from succinyl-CoA by the enzyme succinyl-CoA synthetase, before being oxidised to fumarate by succinate dehydrogenase (SDH).

Under micro-aerobic or anaerobic conditions, succinate is produced through the mixed acid fermentation pathway. In conditions where oxygen is limited or unavailable, it cannot be used as an electron acceptor to regenerate NAD^+ from NADH. Therefore, to allow respiration and to maintain redox balance, it was shown in *E. coli* that pyruvate is instead used as the electron acceptor and in doing so produces a number of waste organic acids such as lactate and succinate (141). The production of succinate during mixed acid fermentation is achieved through conversion of PEP. PEP is converted to oxaloacetate by carboxylation, before subsequent conversion into malate, fumarate and finally succinate, producing one molecule of ATP and oxidising two molecules of NADH in the process. The enzymes involved in this pathway are phosphoenolpyruvate carboxylase (*ppc*), malate dehydrogenase, fumarate hydratase and fumarate reductase.

Aerobically, succinate can also be produced through the glyoxylate pathway. The glyoxylate pathway is a variation of the TCA cycle and

occurs when glucose or other complex carbon sources are limited. The glyoxylate pathway converts isocitrate to malate, producing succinate as a by-product, by-passing the production of succinyl-CoA and fumarate. Instead, an alternative carbon source, such as acetate, is used to produce acetyl-CoA, which combines with 2-ketoglutarate to produce malate (142). Regulation of this pathway is controlled by isocitrate lyase regulator (IclR), which activates the pathway when the only carbon sources available are non-complex, fermentation products, such as acetate and lactate (143). Many different studies have tried to manipulate this pathway to over-produce succinate (144-148).

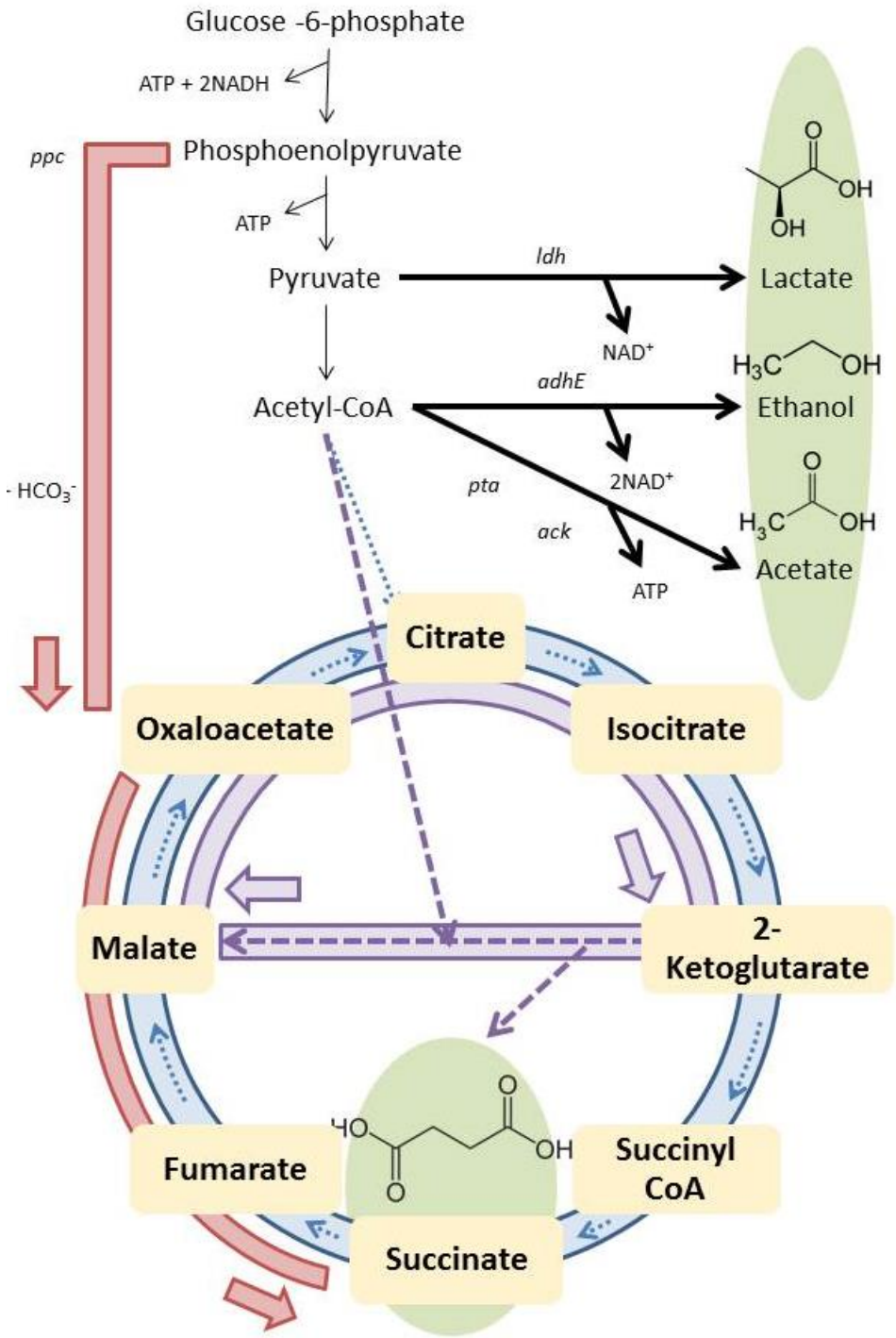


Figure 1.9 – A schematic representation of the production of succinate and other fermentation products. The schematic shows the central carbon metabolism pathway from glucose to pyruvate. Pyruvate can then be used to regenerate NAD^+ through mixed acid formation (Black arrows). The TCA cycle is represented by the blue circle. The phosphoenolpyruvate (PEP) pathway, which produces succinate under anaerobic conditions, is represented in red. The glyoxylate pathway, which is active under limiting carbon conditions, also produces succinate as a by-product, and is represented in purple.

1.6.2 Metabolic Engineering of Microorganisms to produce Succinate

The products of mixed acid fermentation, such as succinate and ethanol, have numerous applications in both biotechnology and industry. For this reason, there has been much research into the genetic engineering of many different microorganisms to produce high yields of these end products. As with microbiology in general, *E. coli* has been extensively studied and genetically engineered to produce succinate at increased concentrations. *E. coli* naturally produces succinate, both aerobically, as an intermediate of the TCA cycle, and anaerobically, through mixed acid fermentation. However, this is at low levels. Under anaerobic conditions, the removal of competing pathways allows carbon flow to proceed through pathways linked with succinate production. However, often this removal unbalances the redox potentials of the cell, as the oxidation of NADH cannot be achieved fast enough. This causes problems with glucose utilisation and growth, as is the case for the strain NZN111. This strain is deficient in the genes *pfl* and *ldh*. However, overexpressing the gene of malate dehydrogenase achieved balancing of the redox potentials and allowed the utilisation of glucose and the production of succinate. This showed an increase of succinate from 1.00 mol/mol of glucose to 1.19 mol/mol (149). Alternatively, a process of evolution and selection was used to successfully improve the growth of a strain deficient in *ldh*, aldehyde-alcohol dehydrogenase (*adhE*), and acetate kinase (*ack*). When

analysed, mutations were identified in the *pfl* and formate transporter (*focA*), which reduced the loss of flux to formate and acetyl-CoA and produced 87 g/L of succinate with a yield of 1.41 mol/mol glucose (150). Aerobic overexpression of succinate has also been successfully achieved through both the TCA cycle and the glyoxylate pathway (143-145). Through inactivation of *sdh*, isocitrate lyase regulator (*iclR*), *ack* and phosphate acetyltransferase (*pta*), the resulting strain was able to produce 58.3 g/L of succinate with a yield of 0.94 mol/mol glucose. This yield can also be increased through the overexpression of *ppc* (151). In addition to *E. coli*, other strains of significance that have been studied are bacteria that naturally produce succinate, such as *Actinobacillus succinogenes*, *Anaerobiospirillum succiniciproducens*, and *Mannheimia succiniciproducens* (30). These bacteria have been shown to naturally produce succinate at levels between 52–106 g/L, with yields of 0.76–0.88 mol/mol glucose (152-154).

1.7 The Overall Aims of this Thesis

The primary focus of this thesis is to expand the number of tools available for efficient genetic engineering of the thermophile *P. thermoglucosidasius*, in order to utilise this species to produce high value chemicals from biomass waste. *P. thermoglucosidasius* has been demonstrated to be a good candidate organism for the production of biofuels and other high value chemicals from lignocellulosic material. However, currently, the

number of tools available for use in *P. thermoglucosidasius* is not diverse enough to allow efficient and simple genetic engineering. Therefore, the initial work of the project will focus on the identification and testing of genetic tools required for gene expression. This work will involve evaluation of a thermostable fluorescent marker, new to *P. thermoglucosidasius*, which can then be utilised to test the relative strengths of several gene expression elements. These will include toosl such as promoters and ribosomal binding sites. The availability of multiple gene expression elements is essential to allow tailored expression of the genes necessary to produce a specific desired product. The second part of the project will aim to successfully implement, using the gene expression elements tested initially, an active CRISPR/Cas9 system into *P. thermoglucosidasius*, to allow the addition and removal of genetic information into the genome. The ability to efficiently and precisely edit the genome is an essential aspect of metabolic engineering and using current methods is laborious and location dependent. Therefore, the implementation of a CRISPR/Cas9 system will allow simple, efficient and location independent genome manipulation for the metabolic engineering of *P. thermoglucosidasius*. The final part of this project aims to demonstrate the applicability of the tools tested thus far, achieved by producing a series of genetic manipulations. These will include an attempt to increase the transformation efficiency of *P. thermoglucosidasius*, to allow more efficient genetic engineering, and ultimately the development of a strain which can produce a high value chemical.

Chapter 2

Materials and Methods

2.1 Bioinformatics

2.1.1 Analysis of DNA Sequences

DNA-based sequence analysis was carried out using basic local alignment search tool (BLAST) through the National Centre for Biotechnology Information (NCBI) (U.S. National Library of Medicine, Rockville, MD, USA) website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify regions of similarity between nucleotide and protein sequences.

2.1.2 Metabolic Pathway Information

Information regarding general and specific metabolic pathways within *P. thermoglucosidasius* was accessed using the Kyoto Encyclopedia of Genes and Genomes (Kegg) (Kyoto, Japan) (http://www.genome.jp/kegg/) website through the Kegg Pathway Database (http://www.genome.jp/kegg/pathway.html).

2.1.3 Management of DNA Sequence Data and Plasmid Map Design

DNA sequences were managed using a variety of programs that included predominately using A plasmid editor (ApE) to design the sequences, then

exporting these sequences into SnapGene Viewer for visualisation. To a lesser extent Benchling (https://benchling.com/) software was also used for management of DNA sequences and plasmid map design. Alignment and reverse complementation of sequences was carried out using ApE.

2.1.4 Oligonucleotide Design, Analysis and Synthesis

Oligonucleotides for synthetic guide RNA (sgRNA), polymerase chain reaction (PCR) and for sequencing of DNA sequences were designed manually. Secondary structure of oligonucleotides was analysed for the production of primers and for calculating the free energy of Rho-indepenent terminators. DNA sequences were analysed using the OligoAnalyzer tool of Integrated DNA Technologies (Coralville, IA, USA) (https://www.idtdna.com/calc/analyzer). Oligonucleotide synthesis was performed by Sigma Aldrich (St. Louis, MO, USA). Spacer acquisition for sgRNA was completed using Benchling software.

2.1.5 DNA Fragment Synthesis

Specific DNA fragments were synthesized by Atum (formally DNA 2.0) (Newark, CA, USA) and Twist Bioscience (San Francisco, CA, USA).

2.1.6 Prediction of Bacterial Promoters

Prediction of bacterial promoters from nucleotide sequences from *P. thermoglucosidasius* was carried out using the prediction software, Softberry BPROM (155) http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb.

2.1.7 Statistical Analysis

Statistical analysis was carried out using Graphpad Prism 7.03. A one-way analysis of variance (ANOVA) was used to determine whether the data presented from a sample was statistically different to the control. P values ≤ 0.05 were considered statistically significant and are represented by asterisks. The following asterisk format is used throughout; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

2.1.8 NMR Data Analysis

NMR data analysis was carried out using Topspin™ software from Bruker (Billerica, MA, USA). The Human Metabolomics Data Base (HMDB) was used to identify specific metabolite peaks (156).

2.2 Media

2.2.1 Luria-Bertani medium (LB)

LB broth contained tryptone (10 g) (Difco Lab, Oxford, UK), yeast extract (5 g) (Oxoid, Basingstoke, UK) and NaCl (5 g) per litre of deionised water. The pH of the media was adjusted to 7.5 using HCl or NaOH prior to autoclaving. LB agar was prepared by the addition of No. 1 Bacteriological agar (10 g) (Oxiod) per litre of deionised water before being autoclaved.

2.2.2 Soy Peptone Yeast Extract No Glycerol (2SPYNG)

2SPYNG contained soy peptone (16 g) (Sigma-Aldrich, St. Louis, MO, USA), yeast extract (10 g) (Oxoid) and NaCl (5 g) per litre of deionised water. The pH was adjusted to 7.0 using 5M KOH prior to autoclaving.

2.2.3 Soy Peptone Yeast Extract (2SPY)

2SPY was made in the same way as 2SPYNG but included glycerol (10 g). The pH was adjusted to 7.0 using 5M KOH prior to autoclaving.

2.2.4 Tryptone Soya Agar (TSA)

TSA agar was prepared using TSA (40 g) (Sigma-Aldrich) per litre of deionised water and autoclaved.

2.2.5 Clostridium Basal Medium (CBM)

CBM consisted of $MgSO_4 \cdot 7H_2O$ (200 mg), $MnSO_4 \cdot H2O$ (7.58 mg), $FeSO_4 \cdot 7H_2O$ (10 mg), p-aminobenzoic acid (1 mg), biotin (2 µg), Thiamine·HCl (1 mg), casein hydrolysate (acid hydrolyzed) (4 g), K_2HPO_4 (0.5 g) and KH_2PO_4 (0.5 g) per litre of deionised water (O'Brien & Morris, 1971). Vitamins and minerals were made as stock solutions (1.0 mg/mL). Stock solutions of K_2HPO_4 and KH_2PO_4 were also made (50 mM). The pH was adjusted to 7.0 using 5 M KOH before autoclaving. The addition of sugars was carried out as indicated after being autoclaved, from sterile 10 % solutions. CBM agar was prepared by the addition of No. 3 Bacteriological agar (15 g) per litre, prior to being autoclaved.

2.2.6 USYE

USYE media (24) comprised of, $NaH_2PO_4 \cdot 2H_2O$ (1.56 g), K_2SO_4 (1.74 g), citric acid (0.38 g), $MgSO_4 \cdot 7H_2O$ (0.31 g), $CaCl_2 \cdot 2H_2O$ (2.94 mg), $Na_2MoO_4 \cdot 2H_2O$ (0.40 g), urea (2.94 g), $ZnSO_4 \cdot 7H_2O$ (7.19 mg), $FeSO_4 \cdot 7H_2O$

(27.80 mg), $MnSO_4 \cdot H_2O$ (8.45 mg), $CuSO_4 \cdot 5H_2O$ (1.25 mg), $CoSO_4 \cdot 7H_2O$ (2.81 mg), $NiSO_4 \cdot 6H_2O$ (2.63 mg), H_3BO_3 (0.62 mg) and biotin (2.44mg) , per litre of deionised water. Mineral and vitamins were prepared as 1 mg/mL stock solution. Yeast extract were added as indicated. The appropriate sugar was added post autoclaving.

2.2.7 Modified ASYE

Modified ASYE media was produced from ASYE media developed by Lin *et al.* (7). This consisted of 0.5 % yeast extract, 2 mM $MgSO_4$ (0.24 g), Thiamine (0.01 g), Citric acid (0.384 g), $FeSO_4 \cdot 7H_2O$ (27.8 mg), $NiCl_3 \cdot 6H_2O$ (4.6 mg), 0.2 M HEPES buffer (47.6 g), biotin (3.05 mg), 1X dilution Trace Metal Mix A5 (Sigma) and 1X M9 Media (Sigma), per litre of deionised water. Filter sterilised 1 % glucose and 0.1 mM $CaCl_2$ (0.011 g) was added to the media post autoclaving.

2.3 Antibiotics and Supplements

See table 2.1, for the final concentration of antibiotics and other supplements to the growth medium. Antibiotics were prepared from a stock solution according the manufactures instructions.

Table 2.1 – Final working concentrations for antibiotics and other supplements used in this study.

	E. coli	*P. thermoglucosidasius*
Antibiotic		
Kanamycin (Km)	50 µg/mL	12.5 µg/mL
Supplements		
5-fluoroorotic acid (5-FOA)	-	TSA 2000 µg/mL, 2SPYNG 500 µg/mL
Uracil	-	20 µg/mL
Lysozyme	-	20 mg/mL

2.4 Solution, Buffers and Gels

2.4.1 Agarose Gel

Agarose gels were made with 1% agarose in 1 x TAE buffer which contained 40 mM 2-Amino-2-(hydroxymethyl)propane-1,3-diol (Tris), 1 mM Ethylenediaminetetraacetic acid (EDTA) and 0.1% (v/v) glacial acetic acid.

2.4.2 Annealing Buffer (1X)

Annealing buffer consisted of 10 mM Tris (7.5 – 8.0 pH) (Sigma), 50 mM NaCl (Sigma) and 1 mM EDTA (Sigma) in nuclease-free water (Promega).

2.4.3 Protein Extraction Buffer (1X)

Extraction of proteins was carried out using BugBuster® 10X protein extraction reagent (Merck, Darmstadt, Germany). The protein extraction buffer (1X) consisted of BugBuster 10X (2 mL), sodium phosphate buffer 50 mM (18 mL) and EDTA-free protease inhibitor tablet (Roche Holding AG, Basel, Switzerland). Lysozyme (20 mg/mL) was added to the extraction buffer for *P. thermoglucosidasius*.

2.4.4 Electroporation buffer (EPB)

For the preparation of electrocompetent *P. thermoglucosidasius* cells, EPB was used. EPB contained (final concentrations), 0.5 M sorbitol, 0.5 M mannitol, 10% (v/v) glycerol.

2.5 Bacterial Growth Conditions

2.5.1 Escherichia coli

E. coli Top10 strain (Invitrogen Ltd, Carlsbad, CA, USA) was grown in LB media (5 mL) containing the appropriate antibiotic selection at 37 ˚C on a rotary shaker (200 RPM) overnight (O/N) or on LB agar plates with appropriate antibiotic selection at 37 ˚C, O/N. Strains were stored at -80 °C.

2.5.2 *Parageobacillus thermoglucosidasius*

P. thermoglucosidasius was grown in 2SPYNG (5 mL) containing the appropriate antibiotic selection at 52 ˚C shaking (250 RPM) O/N or on TSA agar plates with appropriate antibiotic selection at 52 ˚C, O/N. Strains were stored in 2 mL sterilized microtubes as 1 mL glycerol stock. Culture

(20 mL) was concentrated in 2SPYNG (500 μL) and added to 50% glycerol (500 μL) and stored at - 80 ˚C.

2.5.3 Bacterial Cell Density

Growth of bacterial cultures in liquid medium was monitored by measuring optical density at 600 nm (OD_{600}) using a Jenway 6300 (Cole-Parmer, Stone, UK). Samples measured were diluted to 1×10^{-1}.

2.6 Bacterial Strains

Table 2.2 – A list of all the bacterial strains used and produced in this study.

Strain	Description	Reference
Parageobacillus thermoglucosidasius NCIMB11955	Wild Type Isolate	TMO Renewables, London, UK
Escherichia coli Top10	F- *mcr*A Δ (*mrr-hsdRMS-mcrBC*) Φ80*lacZM15*Δ*lacX74 deoR recA1araD139* Δ*(ara-leu)*7697*galU galK rpsL* (Str^{R})*endAl nupG*	Invitrogen Ltd
Streptococcus thermophilus LMD-9 GenBank: CP000419.1	Genomic DNA	ATCC
P. thermoglucosidasius ML001	Δ*ack*	This study
P. thermoglucosidasius ML002	pNCI001^{-}	This study
P. thermoglucosidasius ML003	pNCI002^{-}	This study
P. thermoglucosidasius ML004	pNCI001^{-} pNCI002^{-}	This study

P. thermoglucosidasius ML005	::*eCGP123*	This study
P. thermoglucosidasius ML006	pNCI001::*eCGP123*	This study
P. thermoglucosidasius Δ*pyrE*	Δ*pyrE*	Sheng *et al.* (118)
P. thermoglucosidasius ML007	Δ*pyrE* pNCI001::N-*pyrE*	This study
P. thermoglucosidasius ML008	Δ*pyrE* pNCI001::I-*pyrE*	This study
P. thermoglucosidasius ML009	ΔRM1	This study
P. thermoglucosidasius ML010	ΔRM2	This study
P. thermoglucosidasius ML011	ΔRM3	This study
P. thermoglucosidasius ML012	ΔRM1ΔRM2	This study
P. thermoglucosidasius ML013	ΔRM2ΔRM3	This study
P. thermoglucosidasius ML014	ΔRM1ΔRM3	This study
P. thermoglucosidasius ML015	ΔRM1ΔRM2ΔRM3	This study
P. thermoglucosidasius ML016	Δ*ldh*	This study
P. thermoglucosidasius ML017	Δ*cox*	This study

P. thermoglucosidasius ML018	Δ*adhE*	This study
P. thermoglucosidasius LS005	Δ*pyrE*Δ*ldh*Δ*pfl*	Sheng *et al.* (118)
P. thermoglucosidasius LS006	Δ*pyrE*Δ*ldh*Δ*pfl*Δ*adhE*	L. Sheng
P. thermoglucosidasius LS007	Δ*pyrE*Δ*ldhpdh*^{up}Δ*pfl*Δ*adhE*	L. Sheng
P. thermoglucosidasius ML019	Δ*ldh*Δ*cox*	This study
P. thermoglucosidasius ML020	Δ*pyrE*Δ*ldh*Δ*pfl*Δ*cox*	This study
P. thermoglucosidasius ML021	Δ*pyrE*Δ*ldh*Δ*pfl*Δ*adhE*Δ*cox*	This study
P. thermoglucosidasius ML022	Δ*pyrE*Δ*ldhpdh*^{up}Δ*pfl*Δ*adhE*Δ*cox*	This study
P. thermoglucosidasius ML023	Δ*sdh*	This study
P. thermoglucosidasius ML024	Δ*ldh*Δ*sdh*	This study

2.7 Plasmids

Table 2.3 – A list of all the plasmids used and produced in this study.

Plasmid	Description	Reference
pMTL61110	*P. thermoglucosidasius* modular shuttle vector version 1, ColE1 + tra, pUB110, km^R	Sheng *et al.* (118)
pTML61110_pldhRBS	pMTL61110 with the P_{ldh} from *G. stearothermophilus* and a synthetic RBS	This study
pTMO372	*G. stearothermophilus* DSMZ 6285 *pheB* based reporter vector, *pheB* flanked upstream by the *P. thermoglucosidasius* P_{idh}, pBM1, repBST1, Km^R, Amp^R	TMO Renewables
pMTL_sGFP	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, sGFP gene	This study
pMTL_eCGP123	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, eCGP123 gene	This study
pMTL_GAF	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS,	This study

	GAF gene	
pMTL_YtvALOV	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, YtvA LOV gene	This study
pMTL_Dasher	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, Dasher	This study
pMTL_Paprika	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, Paprika	This study
pMTL_Cayenne	pMTL61110_pldhRBS containing T1T2 terminator, P_{ldh}, synthetic RBS, Cayenne	This study
pMTL_TRsG	pMTL61110_pldhRBS containing T1T2 terminator, promoter-less, synthetic RBS, sGFP gene	This study
pMTL_TRsG_(Promoter)	pMTL61110_RBS, replaced application module. T1T2 terminator, specific promoter, synthetic RBS, sGFP gene and contains specific promoter sequence	This study
pMTL_TRsG_(R1-14)	pMTL61110_pldh, replaced application module. T1T2 terminator, P_{ldh}, sGFP gene, contains specific SD sequence	This study
pMTL_TRsG_(Terminator)	pMTL61110_pldhRBS, replaced application module. T1T2 terminator, P_{ldh}, sGFP gene, contains specific Terminator sequence	This study

pMTL_RbxE_gapd	pMTL61110_pldhRBS containing T1T2 terminator, P_{gapd}, Riboswitch E, sGFP gene	This study
pMTL_RbxF_gapd	pMTL61110_pldhRBS containing T1T2 terminator, P_{gapd}, Riboswitch F, sGFP gene	This study
pMTL_RbxG_gapd	pMTL61110_pldhRBS containing T1T2 terminator, P_{gapd}, Riboswitch G, sGFP gene	This study
pMTL_RbxE_fbpa	pMTL61110_pldhRBS containing T1T2 terminator, P_{fbpa}, Riboswitch E, sGFP gene	This study
pMTL_RbxF_fbpa	pMTL61110_pldhRBS containing T1T2 terminator, P_{fbpa}, Riboswitch F, sGFP gene	This study
pMTL_RbxG_fbpa	pMTL61110_pldhRBS containing T1T2 terminator, P_{fbpa}, Riboswitch G, sGFP gene	This study
pMTL_RbxE_pfk	pMTL61110_pldhRBS containing T1T2 terminator, P_{pfk}, Riboswitch E, sGFP gene	This study
pMTL_RbxF_pfk	pMTL61110_pldhRBS containing T1T2 terminator, P_{pfk}, Riboswitch F, sGFP gene	This study
pMTL_RbxG_pfk	pMTL61110_pldhRBS containing T1T2 terminator, P_{pfk}, Riboswitch G, sGFP gene	This study

pCas91	pMTL61110 containing *stcas9-1* under the control of P_{ldh} and the *ldh* RBS, sgRNA under the control of P_{gapd}	This study
pCas93	pMTL61110 containing *stcas9-3* under the control of P_{ldh} and the *ldh* RBS	This study
pCas91_Ack	pMTL61110 containing *stcas9-1* under the control of P_{ldh} and the *ldh* RBS, *ack* targeting sgRNA under the control of P_{gapd}	This study
pCas93_Ack	pMTL61110 containing *stcas9-3* under the control of P_{ldh} and the *ldh* RBS, *ack* targeting sgRNA under the control of P_{gapd}	This study
pΔCas91	pMTL61110 containing Δ*stcas9-1* under the control of P_{ldh} and the *ldh* RBS, *ack* targeting sgRNA under the control of P_{gapd}	This study
pΔCas93	pMTL61110 containing Δ*stcas9-3* under the control of P_{ldh} and the *ldh* RBS, *ack* targeting sgRNA under the control of P_{gapd}	This study
pCas91only	pMTL61110 containing *stcas9-1* under the control of P_{ldh} and the *ldh* RBS	This study
pCas93only	pMTL61110 containing *stcas9-3* under the control of P_{ldh} and the *ldh* RBS	This study
pCas91_Ack_hr	pCas91_Ack containing homology arms (500 bps) from up and downstream of *ack*	This study
pCas93_Ack_hr	pCas93_Ack containing homology arms (500 bps) from up and downstream of *ack*	This study

pCas91_Ack_hrT	pCas91_Ack containing a T1 terminator and homology arms (500 bps) from up and downstream of *ack*	This study
pCas93_Ack_hrT	pCas93_Ack containing a T1 terminator downstream of the sgRNA and an *ack* homology cassette (HC)	This study
pCas93_repA_hrT	pCas93 containing *repA* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *repA* HC	This study
pCas93_repB_hrT	pCas93 containing *repB* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *repB* HC	This study
pCas93_tox_hrT	pCas93 containing *tox* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *tox* HC	This study
pCas93_eC_KI	pCas93 containing genome targeting (downstream of *pyrE*) sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a knock-in homology cassette containing the eCGP123 reporter gene between two regions of homology downstream of the *pyrE* gene	This study
pCas93_pNCI001eC_KI	pCas93 containing BCV53_19490 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a knock-in homology cassette containing the eCGP123 reporter gene between two regions of homology from up and downstream of BCV53_19490	This study
pCas93_pNCI002eC_KI	pCas93 containing BCV53_19815 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a knock-in homology	This study

	cassette containing the eCGP123 reporter gene between two regions of homology from up and downstream of BCV53_19815	
pCas93_pNCI001N-*pyrE*_KI	pCas93 containing BCV53_19490 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a knock-in homology cassette containing the eCGP123 reporter gene between two regions of homology from up and downstream of BCV53_19490	This study
pCas93_pNCI001I-*pyrE*_KI	pCas93 containing BCV53_19490 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a knock-in homology cassette containing the eCGP123 reporter gene between two regions of homology from up and downstream of BCV53_19490	This study
pCas93_RM1_hrT	pCas93 containing RM1 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a RM1 HC	This study
pCas93_RM2_hrT	pCas93 containing RM2 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a RM2 HC	This study
pCas93_RM3_hrT	pCas93 containing RM3 targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a RM3 HC	This study
pCas93_LDH_hrT	pCas93 containing *ldh* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *ldh* HC	This study
pCas93_COX_hrT	pCas93 containing *cox* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *cox* HC	This study

pCas93_ADHE_hrT	pCas93 containing *adhE* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *adhE* HC	This study
pCas93_SDH_hrT	pCas93 containing *sdh* targeting sgRNA under the control of P_{gapd}, T1 terminator downstream of the sgRNA and a *sdh* HC	This study

2.8 Oligonucleotide Primers

A list of the primers used in this study and their sequences are provided in the supplementary data (Supplementary Data 9.1).

2.9 DNA Manipulation, Extraction and Sequencing

2.9.1 Extraction and purification of *P. thermoglucosidasius* Chromosomal DNA

P. thermoglucosidasius chromosomal DNA was extracted and purified using the DNeasyR Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

For high purity genomic DNA, a phenol:chloroform extraction was performed. The following procedure was performed within a fume hood. Culture (2 mL) was pelleted and the supernatant was discarded. The cell pellet was resuspended in lysozyme containing buffer (180 µL) and incubated at 37 °C with gentle agitation for 30 mins. Proteinase K (25 µL) (Sigma), ddH_2O (85 µL) and a 10 % w/v SDS solution (110 µL) was added to the mixture and inverted to mix before incubation at 65 °C. RNase (20 µL) (Sigma) was added to the solution and incubated for 10 mins at room

temperature. The solution was added to a phase-lock tube. Phenol:chloroform (400 µL) was added to the tube and inverted. RNase (20 µL) (Sigma) was added to the solution and incubated for a further 10 mins at room temperature before centrifuging at 16,000 x g for 5 mins. Phenol:chloroform (400 µL) mix followed by centrifugation at 16,000 x g for 5 mins was repeated a further two more times. The top layer of the of the solution was transferred to an ice cold microtube containing 3 M Sodium acetate (40 µL) and ethanol (800 µL), mixed by inversion and held at -80 °C for 30 mins. Centrifugation of the sample at 4 °C at full speed for 15 mins was carried out. The supernatant was remover and washed with 70 % ethanol (1 mL) and centrifuged for 3 mins. The supernatant was removed and further span before air drying the pellet and resuspending in ddH_2O (50 µL).

2.9.2 Extraction and Purification of Plasmid DNA

Plasmid DNA from *E. coli* was extracted and purified using the Qiagen QIAprep Miniprep Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. For extraction from *P. thermoglucosidasius*, prior to extraction using Qiagen QIAprep Miniprep Kit, cells were incubated at 37 ˚C with P1 buffer (250 µL) and lysozyme (20 mg/mL) for 30 mins.

2.9.3 Amplification of DNA by Polymerase Chain Reaction (PCR)

Amplification of DNA through PCR was performed using either Phusion DNA polymerase kit (New England Biolabs™, Ipswich, MA) or DreamTaq Green PCR Master Mix (2X) (Life Technologies, Carlsbad, CA, USA). All PCR reactions were performed in Axygen 0.5 mL thin-walled, flat cap PCR (Corning Inc., NY, USA) tubes using a TProfessional TRIO thermocycler (Biometra, Göttingen, Germany).

DreamTaq was exclusively used for colony PCR with was used for screening purposes. Single colonies from agar places were selected using sterile wooden toothpicks and inoculated into nuclease free water (company) (20 µL). 2x DreamTaq PCR Master Mix and primers were added to the solution according to the manufacturer's instructions and PCR steps were carried out following the manufacturer's protocol (Table 2.4 and Table 2.5).

Table 2.4 – Components and their subsequent volumes used to perform a colony PCR using DreamTaq Green PCR Master Mix.

Component	Volume
DreamTaq Green PCR Master Mix (2X)	25 µL
Forward Primer (10 µM)	2.5 µL
Reverse Primer (10 µM)	2.5 µL
Template DNA	20 µL
Total reaction volume	**50 µL**

Table 2.5 – Thermal cycling conditions used for colony PCR using DreamTaq Green PCR Master Mix.

Step	Temperature / °C	Duration	Number of Cycles
Initial denaturation	95	60 seconds	1
Denaturation	95	30 seconds	35
Annealing	T_m -5	30 seconds	
Extension	72	60 seconds / kb	
Final extension	72	10 mins	1

For all other PCR, Phusion® High-Fidelity PCR Master Mix with HF Buffer was used and carried out according to the manufactures instructions (Table 2.6 and Table 2.7).

Table 2.6 – Components and their subsequent volumes used to perform a PCR using Phusion.

Component	Volume
Phusion Master Mix (2X)	25 µL
Forward Primer (10 µM)	2.5 µL
Reverse Primer (10 µM)	2.5 µL
Template DNA	0.5 µL
Nuclease-free water	To 50 µL
Total reaction volume	**50 µL**

Table 2.7 – Thermal cycling conditions used for a PCR using Phusion.

Step	Temperature / °C	Duration	Number of Cycles
Initial denaturation	98	60 seconds	1
Denaturation	98	10 seconds	30
Annealing	T_m +3	30 seconds	
Extension	72	30 seconds / kb	
Final extension	72	10 mins	1

2.9.4 Restriction Endonuclease Digestion of DNA

The restriction endonuclease enzymes from Thermo Scientific™ FastDigest™ enzymes (Life Technologies) and NEB restriction enzymes (New England Biolabs™) were used digest specific DNA sequences. Digestion reactions were carried out in accordance with the manufacturer's instructions.

2.9.5 DNA Blunting

T4 DNA Polymerase (New England Biolabs™) was used to remove sticky-ends. The blunting of DNA was achieved through the synthesis of DNA for 5' over hangs and the removal 3' over hangs through exonuclease activity. Reactions were carried out in accordance with the manufacturer's instructions.

2.9.6 Dephosphorylation of DNA

Antarctic Phosphatase (New England Biolabs™) was used to remove the 5' phosphate group from linear DNA following the manufacturer's instructions.

2.9.7 TA Cloning

TA cloning was carried out using pGEM®-T Easy Vector System (Promega) following the manufacturer's instructions. Screening for correct colonies was accomplished through blue/white screening due to insertion of the desired fragment into the *lacZ*α region of the vector.

2.9.8 DNA analysis by Agarose Gel Electrophoresis

1% agarose gels were mixed with SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) evenly. 6X loading buffer (New England Biolabs™) was mixed with DNA reaction mixtures before loading into the agarose gel. Separation of DNA was then carried out using a PowerPac™ Basic Power Supply (BioRad) (100 – 120 V, 20 – 60 mins). A Gel Doc™ XR+ Gel Documentation System (BioRad) was used to visualised the DNA. The size of the DNA was attained using Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb) (New England Biolabs™).

2.9.9 Agarose gel extraction and purification of DNA

DNA fragments were excised from agarose gel under blue light. Gel extraction was carried out using Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA) according to the manufacturer's instructions.

QIAquick PCR Purification Kit (Qiagen) were used routinely to purify DNA from contaminating oligonucleotides, enzymes and buffers from PCR reactions, restriction digests and blunt-ending reactions in accordance with manufacturer's instructions.

2.9.10 Ligation of DNA fragments

Purified DNA fragments, containing compatible ends, were ligated together using LigaFast™ Rapid DNA Ligation System (Promega) following the manufacturer's instructions. For O/N ligation reactions, T4 DNA ligase (Promega) was used following the manufacturer's instructions.

2.9.11 Site-Directed Mutagenesis (SDM)

Site-directed mutagenesis was carried out using Q5® Site-Directed Mutagenesis Kit (New England Biolabs™) following manufacturer's instructions. Primers were designed using the NEBaseChanger™ tool (New England Biolabs™).

2.9.12 Sequencing of DNA

DNA sequencing was carried out by Source Bioscience UK Limited (Source Biosciences, Nottingham, UK). Samples were prepared in accordance with their specifications.

Genomic DNA was prepared using phenol:chloroform extraction. Paired-end libraries were sequenced using an Illumina MiSeq bench-top sequencer (Deepseq, University of Nottingham, Nottingham, UK).

2.10 Transfer of Plasmid DNA into Bacterial Cells

2.10.1 Preparation of Electro-Competent *E. coli* Cells

A conical flask containing LB media (200 mL) was inoculated with an overnight pre-culture of *E. coli* Top10 cells (1 mL). The culture was incubated (37 °C, 200 RPM), to exponential phase (OD_{600} of 0.5 – 1.0). The culture was placed on ice for 10 mins, split evenly into six pre-chilled falcon tubes. The cells were then collected by centrifugation (5000 x g, 15 mins, 4 °C) and washed with ice cold sterile distilled water (30 mL) before being centrifuged. The cells were then combined into two falcon tubes and washed with 10% glycerol (2 mL) and harvested (5000 x g, 15 mins, 4 °C).

The competent cells were resuspended in a total volume of 600 µL and aliquoted into pre-chilled sterile micro-tubes (45 µL) and stored at -80 °C.

2.10.2 Preparation of Chemical Competent *E. coli* Cells

A conical flask containing LB media (100 mL) was inoculated with an overnight pre-culture of *E. coli* Top10 cells (1 mL). The culture was incubated (37 °C, 200 RPM), to exponential phase (OD_{600} of 0.3 – 0.4). The culture was placed on ice for 10 mins, split evenly into three pre-chilled falcon tubes. The cells were then collected by centrifugation (5000 x g, 15 mins, 4 °C) and then resuspended in 100mM $CaCl_2 \cdot 2H_2O$ (15 mL). The cells were then held on ice overnight at 4 °C. The cells were then transferred to pre-chilled micro-tubes (1.5 mL) and centrifuged (5000 x g, 15 mins, 4 °C). The each micro-tube was resuspended in sterile 100 mM $CaCl_2 \cdot 2H_2O$ with 10% glycerol (500 µL) and aliquoted into pre-chilled sterile micro-tubes (100 µL) and stored at -80 °C.

2.10.3 Transformation into *E.coli* using Electroporation

Plasmid DNA (1 – 2 µL) (100 – 500 ng/µL) was added to electrocompetent *E. coli* Top10 cells (45 µL). This mixture was added to a pre-cooled 2 mm gap electroporation cuvette. Electroporation was carried out using a

MicroPulser (BioRad, Hercules, CA), set at 200 Ω, 2.5 kV and 25 μFD. Immediately after the pulse, LB media (800 μL) was added to the cells and left to recover for 2 hours incubated at 37 °C with shaking (200 RPM). After recovery, the cells (100 – 200 μL) were plated onto LB agar plates containing the appropriate antibiotic selection.

2.10.4 Transformation into *E. coli* using Heat Shock

Plasmid DNA (1 – 2 μL) (100 – 500 ng/μL) was added to chemical competent *E. coli* Top10 cells (60 μL). This mixture was left on ice for 30 – 45 mins, then transferred to 42 °C for 30 seconds, then returned back to ice for 3 – 5 mins. LB media (900 μL) was then added to the cells and left to recover for 1.5 hours, incubated at 37 °C and shaking (200 RPM). After recovery, the cells (200 μL) were plated onto LB agar plates containing the appropriate antibiotic selection.

2.10.5 Preparation of Electro-Competent *P. thermoglucosidasius* Cells

An aliquot of frozen cells were evenly plated onto TSA agar and grown overnight (52 °C). A full loop of cells were used to inoculate pre-warmed 2SPYNG (50 mL) in a 250 mL baffled conical flask and incubated (52 °C, 250 RPM) to exponential phase (OD_{600} 2.0 – 2.5). The cells were then

transferred evenly to two pre-chilled falcon tubes and held on ice (10 mins) before collection by centrifugation (5000 x g, 15 mins, 4 °C). The cells were then washed sequentially, by gentile resuspension using electroporation buffer (EPB) (30 mL, 20 mL, 10 mL, 10 mL), and centrifuged (5000 x g, 15 mins, 4 °C). The competent cells were then resuspended in EPB (2 mL) and aliquoted into pre-chilled sterile micro-tubes (60 µL) and stored at -80 °C.

2.10.6 Transformation into *P. thermoglucosidasius* using Electroporation

An electroporation of *P. thermoglucosidasius* protocol was provided by TMO Renewables (TMO Renewables, Guildford, UK). Plasmid DNA (5 µL) (100 – 500 ng/µl) was added to electro-competent *P. thermoglucosidasius* NCIMB 11955 cells (60 µL). The mixture was added to a pre-cooled 1 mm gap electroporation cuvette. Electroporation was carried out using a GenePulser (BioRad), set at 600 Ω, 2.5 kV and 10 µFD. Immediately following the pulse, pre-warmed 2SPY (1 mL) was added to the cells and left to recover for 2.5 – 4 hours incubated at 52 °C shaking (250 RPM). After recover the cells were concentrated by centrifugation (5000 x g, 5 mins) and re-suspended in 2SPY media (200 µL) before being plated onto pre-warmed TSA agar plates containing the appropriate antibiotic selection.

2.11 Protein Extraction and Visualisation

2.11.1 Protein Extraction through Chemical Lysis of Cells

The samples (20 mL) were harvested by centrifugation (10,000 x g, 10 mins). The supernatant was discarded and the pellet resuspended in the protein exaction buffer (1X) (1 mL) and incubated (37 °C, 5000 x g, 20 mins). The lysed cells were then centrifuged (4 °C, 16000 x g, 20 mins) and the supernatant transferred to fresh pre-chilled micro-tubes.

2.11.2 Western Blot of HIS-tagged Protein

2X reducing loading dye (10 µL) (Invitrogen) was added to the protein lysate (10 µL) and heated (100 °C). The samples were loaded into a gel and separated by electrophoresis (BioRad). The initial stacking of samples was run (80V, 20 mins) before the main run (120V, 60 mins). The proteins were then transferred onto a polyvinylidene fluoride (PVDF) membrane (BioRad) and blocked with Tris buffer saline (30 mL) plus Milk powder (0.45 g) (60 mins, RT). The blocking solution was then discarded before additonal blocking solution (30 mL) containing anti-his Antibody (1/10,000) was added to membrane and incubated (O/N, 4 °C). The membrane was washed with Tris-buffered saline (TBS) (25 mL) containing Tween (0.1 %) (6 washes, 10 mins, 50 RPM). The membrane

was imaged using 3,3',5,5' tetramethylbenzidine (TMB) (Sigma Aldrich) for antibody detection.

2.11.3 Protein Concentration Determination by Bicinchoninic Acid (BCA) assay

Protein lysate (5 µL) was added to H2O (95 µL). Protein estimation was carried out using a Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's instructions.

2.11.4 Analysis of Fluorescent Reporter Proteins

For analysis of fluorescent reporter proteins, both a plate reader and fluorescent microscope were used. The OD_{600} of samples were measured using a spectrophotometer and normalised to an OD_{600} of 1. The samples of cells (500 µL) were loaded into an opaque, clear, flat-bottomed 96 well plate. The plate was loaded into a CLARIOstar plate reader (BMG Labtech, Ortenberg, Germany) and the temperature was set to 52 °C. The excitation spectra were set to the required wavelength and emission spectra of samples were detected. For visualisation of samples on a fluorescent microscope, Nikon SMZ1000 (Nikon Corporation, Shinagawa, Tokyo, Japan), a colony was picked and diluted in phosphate buffered saline (PBS).

2.11.5 sGFP Assay

The strengths of various gene expression elements were routinely assessed in this chapter using the following assay to measure sGFP expression:-

P. thermoglucosidasius: the constructed gene expression vectors were transformed into *P. thermoglucosidasius* competent cells. After recovery, the transformed cells were plated onto TSA plates containing kanamycin (12.5 µg/mL) and grown overnight. A single colony was picked and restreaked onto TSA plates containing kanamycin. A single colony was picked and cultured in a baffled conical flask for 8 hours at 250 RPM, in 2SPYNG media containing kanamycin (12.5 µg/mL), before extracting the protein lysate.

E. coli: the constructed gene expression vectors were transformed into *E. coli* competent cells. After recovery, the transformed cells were plated onto LB plates containing kanamycin (50 µg/mL) and grown overnight. A single colony was picked and cultured in a falcon tube for 8 hours at 200 RPM, in LB media containing kanamycin (50 µg/mL), before extracting the protein lysate.

The level of sGFP was measured using a micro-plate reader (CLARIOstar). The relative fluorescence intensities (RFU) were then normalised using

the protein concentration of each sample. Protein concentration was calculated using the BCA assay.

2.12 Enzymatic Assays

2.12.1 PheB assay

Protein lysate (5 µL) containing 2-hydroxymuconic semialdehyde (HMSA) was added to 50 mM sodium phosphate buffer in a cuvette that was preheated (55 °C). 10 mM catechol (33 µL) was added to the lysate and the absorbance read (375 nm, 1 min) on a Specord spectrophotometer (Analytik Jena, Jena, Germany).

2.13 Analysis of Metabolites

2.12.1 HPLC-UV Analysis

Metabolites produced by *P. thermoglucosidasius* were analysed using High-performance liquid chromatography (HPLC-UV), performed using the Dionex UltiMate 3000 System (Thermo Scientific, Life Technologies). 5 mM H_2SO_4 was used as the mobile phase. The sample supernatant was mixed, in a 1:1 ration, with diluent, which consisted of the mobile phase mixed with 50 mM valeric acid. Samples were filtered to remove

particulate matter through 0.2 µm syringe filters directly into HPLC vials containing 300 µl inserts with split caps. The samples were run on the Bio-Rad Aminex HPX-87H 300 mm x 7.8 mm x 9 µm column (BioRad), at a flowrate of 0.5 mL/min, at 35 °C for 55 mins.

Chapter 3

Development of New and Existing Genetic Tools for *P. thermoglucosidasius*

3.1 Introduction

3.1.1 Genetic Tools Available for Engineering in *P. thermoglucosidasius*

Significant advancements in the field of genetic engineering are reliant on the availability of the requisite gene technologies. This is particularly true of new biological systems, where the necessary tools to make deletions and insertions, as well as more subtle genetic changes, need to be identified and their deployment and utility optimised. The first challenge in any such manipulation is to deliver DNA into the organism. This may be achieved through electroporation, heat-shock, conjugation, or transduction. Currently, only electroporation (24) has been explored for the delivery of genetic material into *P. thermoglucosidasius*.

As discussed in section 1.4.1, delivery of DNA is often achieved through the agency of shuttle vectors. These allow simple construction within an alternative host, such as *E. coli*, and then their subsequent transfer into the desired chassis. A shuttle vector with a modular vector system for *P. thermoglucosidasius,* based on the clostridial modular vector system (157), has already been developed within this group (118). This vector, pMTL61110, contains the kanamycin adenyltransferase (Kan) gene (*kan*), for its heat stable properties, and the pUB110 replicon, from *S. aureus*, for replication in *P. thermoglucosidasius.* Other shuttle vectors currently in use for the *Parageobacillus* species include pNW33N, a commercially

available plasmid used for *G. stearothermophilus* expression (Bacillus Genetic Stock Centre), and pTMO31, pTMO19, pUCG18 and pUB190, all produced by Cripps *et al.* (24). Using these developed vectors, Cripps *et al.* (24) successfully achieved both gene interruptions and integrations in *P. thermoglucosidasius*, through homologous recombination, for the production of the ethanol overproducing strain TM242 (24). Gene integrations can also be achieved via ACE, a method which Sheng *et al.* successfully used in *P. thermoglucosidasius* (118). ACE allows the integration of a gene through the activation and inactivation of the natural selection marker *pyrE* (116), described in more detail in section 4.1.2.

Translation efficiency can also be affected by the degeneracy of the genetic code, as detailed in section 1.4.2. The degeneracy of the codons in DNA means that the 20 amino acids found naturally in nature can be coded for by more than one triplet. Multiple nucleotide sequence combinations can, therefore, encode a polypeptide of an identical amino acid sequence. Codon usage refers to the difference in the frequency of usage of a codon between organisms caused by an evolutionary selection bias towards certain nucleotide sequences and tRNA levels (79-81,86). It is therefore advantageous to determine the existence of any codon usage bias exhibited by *P. thermoglucosidasius* in order to optimise the design of synthesised genes for the most effective translational efficiencies. Codon usage bias information is available for multiple *Parageobacillus* species, but not currently *P. thermoglucosidasius* (158).

Through a selection of gene expression elements, such as promoters, ribosomal binding site (RBS) and terminators, both transcription and translation can be successfully controlled (section 1.4.2). However, current available gene expression elements for use in *P. thermoglucosidasius* are limited. The increased temperature at which *P. thermoglucosidasius* grows is a key issue in the development of tools, as many well-studied gene expression tools do not operate at these high temperatures. Cripps *et al.* utilised the P_{ldh} promoter of the lactate dehydrogenase (Ldh) gene (*ldh*) from *G. stearothermophilus* for expression of pyruvate dehydrogenase (Pdh) in *P. thermoglucosidasius* (24). The strength of this promotor has since been quantified by Bartosiak-Jentys *et al.* (159). Additionally, Lin and co-workers produced a promoter library consisting of several promoters from *P. thermoglucosidasius*, including the P_{ldh} promoter and P_{glpD} promoter of the glycerol-3-phosphate dehydrogenase (GphD) gene (*glpD*) from *Geobacillus thermodenitrificans* (7). However, all the promoters tested from *P. thermoglucosidasius* showed low levels of expression, with the exception of the *G. thermodenitrificans* P_{ldh} promoter (7).

Previous work within the group has led to the identification and characterisation of several useful tools, including a synthetic RBS sequence, the P_{idh} promoter of the isocitrate dehydrogenase (Idh) gene (*idh*) and establishment of ACE, for the engineering of *P. thermoglucosidasius* (118). These have allowed the accomplishment of some basic metabolic engineering, including the recreation of an ethanol

overproducing strain originally engineered by Cripps *et al.* (24). However, currently there is only one promoter and RBS candidate in use within the group, the P_{ldh} promoter from *G. stearothermophilus* and a synthetic RBS sequence produced in-house. At the outset of this project there was, therefore, a need to expand the library of available gene expression elements.

Inducible promoters, discussed in section 1.4.4, allow selective induction or repression of gene expression, through the addition of a compound or molecule. Again, increased temperatures can have a negative impact on the application of existing systems. In addition, the functionality of protein-based inducible systems, such as the *lac* operon, can be organism specific. As a consequence, RNA-based inducible expression can represent a better option (160,161). Riboswitches are regulatory segments of mRNA that in certain configurations sequester the RBS to prevent translation and thereby control gene expression. They contain a secondary structure and an aptamer domain that binds a ligand. Binding of the ligand causes a conformational change, releasing the sequestered RBS, allowing protein synthesis to occur. Hence, an mRNA that contains a riboswitch is directly involved in regulating its own activity, in response to the concentration of its effector molecule. Riboswitch E has previously been identified and tested in a Gram-positive organism, *B. subtilis*, showing inducible and over-expression of targets (162). Ongoing work within our group has facilitated the design, and testing in *Clostridium sporogenes,* of riboswitches F and G, through rational modifications to riboswitch E. (I.

Canadas Blasto, unpublished). These riboswitches sense and bind the molecule theophylline (1,3-dimethylxanthine) as the inducer and have shown potential to be attractive targets for other Gram-positive organisms.

To successfully test the relative strengths of gene expression elements, a quantifiable reporter gene is required. At the initiation of this project, two reporter genes were available for use in *P. thermoglucosidasius*. Most recently Lin *et al.* (7) used a *lacZ* reporter gene, isolated from *G. stearothermophilus* encoding a thermostable β-galactosidase, to evaluate the strengths of several promoters. Furthermore, the *pheB* gene, isolated from *G. stearothermophilus*, has been utilised as a reporter gene to examine the expression of the *G. stearothermophilus* P_{ldh} promoter (159). This reporter assay relies on the production of catechol 2,3-dioxygenase, the gene product of *pheB*, which catalyses the formation of 2-hydroxymuconic semialdehyde (HMSA). The accumulation of HMSA can be recorded spectrophotometrically (λ = 375 nm), and can therefore determine the strength of each promoter relative to one another. Although the use of *pheB* as a reporter gene is a viable method, it is both labour and time intensive.

Fluorescent reporter proteins offer a key advantage over non-fluorescent reporters as not only do they require less labour-intensive methods, they also enable visual observation of protein localisation within individual cells, using confocal microscopy techniques. Traditional fluorescent

proteins, such as green fluorescent protein (GFP) (94), are unstable at high temperatures (163). Therefore, the identification and validation of a thermostable fluorescent protein for use in *P. thermoglucosidasius* is a key target. Several thermostable fluorescent reporters have demonstrated expression at high temperatures. One promising candidate is superfolder GFP (sGFP), a more stable variant of the GFP found in *A. victoria* (91,94), which was generated by Pedelacq *et al.* (164). It has been shown to be fluorescent in the thermophiles, *Thermus thermophilus* (165) and more recently in *G. stearothermophilus* (166). The success of sGFP in these two thermophiles and especially in a *Parageobacillus* species indicates that this could be a promising fluorescent reporter in *P. thermoglucosidasius*.

3.1.2 Chapter Aims

It is clear that the number of genetic tools available in *P. thermoglucosidasius* is currently inadequate to fully explore the potential of this thermophile for chemical and biofuel production. This chapter aims to expand the tools available for the early stages of metabolic engineering in *P. thermoglucosidasius*, specifically targeting the bio-parts involved in the control of transcription and translation. Initial identification and validation of thermostable fluorescent reporter genes will be conducted. Once validated, these can then be utilised to identify a variety of promoters, RBS sequences and terminators and quantify their respective strengths. In addition, this chapter aims to identify candidates for an inducible system and elucidate whether *P. thermoglucosidasius* exhibits a bias within codon occurrence. This will determine whether there is an optimal codon usage system that can be employed when synthesising genes for expression in *P. thermoglucosidasius*.

3.2 Results

3.2.1 Identification and Testing of a Fluorescent Reporter Gene

3.2.1.1 Candidate Selection of Fluorescent Reporter Genes

Potential fluorescent reporter gene candidates were selected primarily for their thermostability, either theoretical or experimentally demonstrated in the literature. Three synthetic, "non-*Aequorea*" proteins, DasherGFP, CayenneRFP and PaprikaRFP, were selected by ATUM from their proteinpaintbox® collection (Figure 3.1). Four additional candidates were identified from the literature, based on studies demonstrating their expression in another thermophile host and/or activity at high temperatures. These candidates were superfolder GFP (sGFP) (164,165), cGMP phosphodiesterase/adenyl cyclase/FhlA (GAF) (167), enhanced consensus green protein variant 123 (eCGP123) (168,169) and YtvA LOV (170).

Once identified, the genes were cloned into the pMTL_61110_pldhRBS vector under the control of the P_{ldh} promoter and a synthetic RBS, to produce the shuttle vectors pMTL_86551_pldhRBS_fluorescent marker (e.g., pMTL_Dash or pMTL_sGFP). Successful construction was confirmed by analytical digest and Sanger sequencing (Source Bioscience).

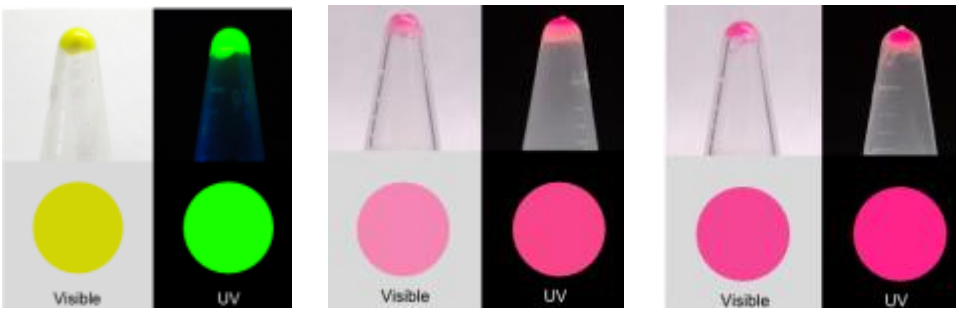


Figure 3.1 – Images illustrating the appearance of the three commercially available fluorescent protein markers in *E. coli*., under both visible and UV light: DasherGFP, PaprikaRFP and CayenneRFP (from left to right). The images are provided by ATUM (www.atum.bio).

3.2.1.2 Expression and Thermostability Testing of Fluorescent Reporter Genes

Shuttle vectors containing the fluorescent markers were transformed into *P. thermoglucosidasius* and were screened to assess expression levels. Screening was conducted using a micro-plate reader (BMG CLARIOstar) to measure fluorescence intensity levels of each candidate. Each sample was cultured and plated (200 µL) onto a black micro-plate, excited at the appropriate wavelength and then emission scans were carried out to measure fluorescence levels, reported as Relative Fluorescent Units (RFU). The relevant excitation and emission wavelengths are summarised in table 3.1. All samples were both grown and tested at 52 °C. Emission was observed only for sGFP and eCGP123, which both emitted similar

fluorescence levels (Figure 3.2). No fluorescence was observed for DasherGFP, CayenneRFP, PaprikaRFP, GAF and YtvA LOV.

Table 3.1 – Excitation and emission wavelengths of tested fluorescent reporter proteins.

Fluorescent Gene	Excitation / nm	Emission / nm
DasherGFP	505	525
CayenneRFP	554	590
PaprikaRFP	554	590
sGFP	485	510
GAF	360	646
eCGP123	493	504
YtvA LOV	450	506

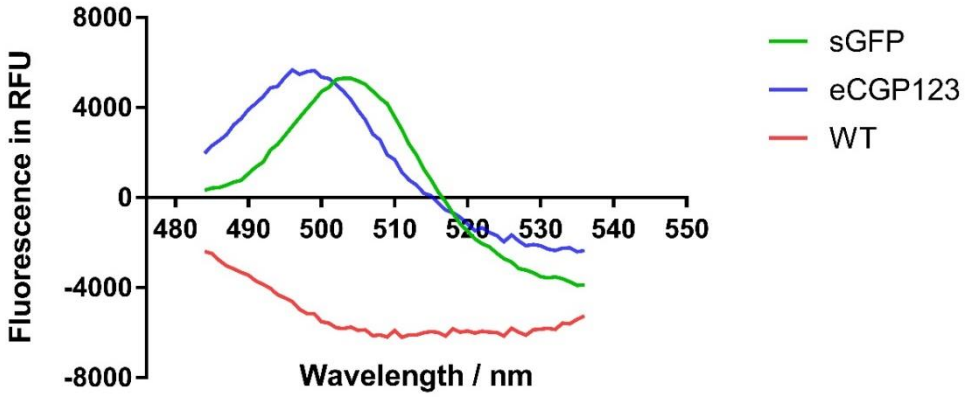


Figure 3.2 – Fluorescence emission analysis using a CLARIOstar microplate reader. The spectrum shows the fluorescence intensity, reported as Relative Fluorescence Units (RFU) produced by *P. thermoglucosidasius* cultures containing sGFP and eCGP123 reporter genes, compared to WT.

It also proved possible to visually observe the fluorescence of sGFP within individual cells using a confocal fluorescence microscope (Nikon SMZ1000) (Figure 3.3). Interestingly, this technique highlighted that fluorescence was not homogenously emitted across the total cell population, with some cells showing no expression. Unfortunately, this could not be compared with the expression for eCGP123, as the excitation wavelength was outside the range of the microscope.

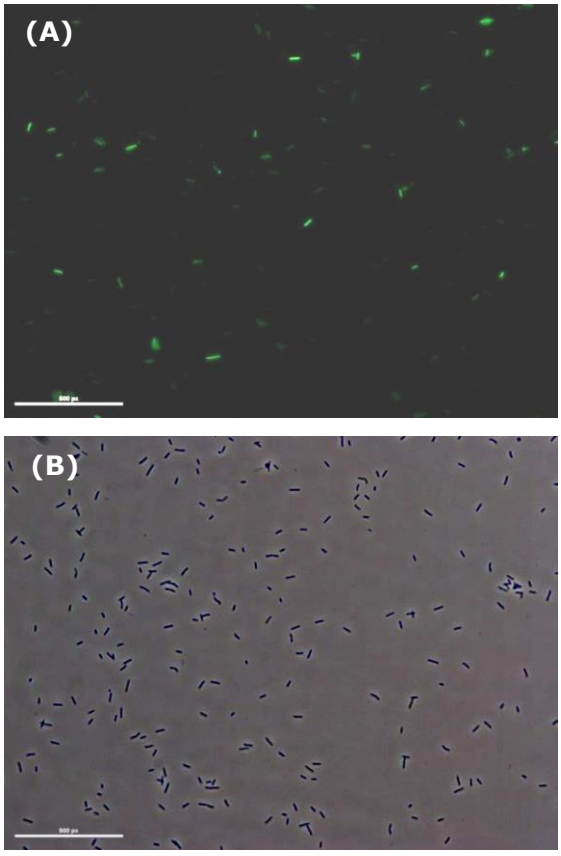


Figure 3.3 – (A) Confocal image of *P. thermoglucosidasius* cultures expressing the fluorescent protein sGFP grown at 52 °C. (B) A bright field illumination of the same frame. Images were taken at x 40 magnification. Scale bars – 500 pixels.

It was decided to further explore the four potential candidates identified from the literature (sGFP, GAF, eCGP123 and YtvA LOV), by confirming protein expression following the addition of a C-terminal histidine-tag (HIS-tag), to determine whether the fluorescent proteins were being translated. A western blot, using antibodies against the HIS-tag, was carried out on the extracted protein lysate of each sample (Figure 3.4). Clear bands were observed for sGFP and eCGP123, corresponding to the correct molecular weight for each protein. However, no protein could be detected for either GAF or YtvA LOV, which probably accounts for the lack of fluorescence observed previously. It was therefore decided to continue with sGFP and eCGP123 only.

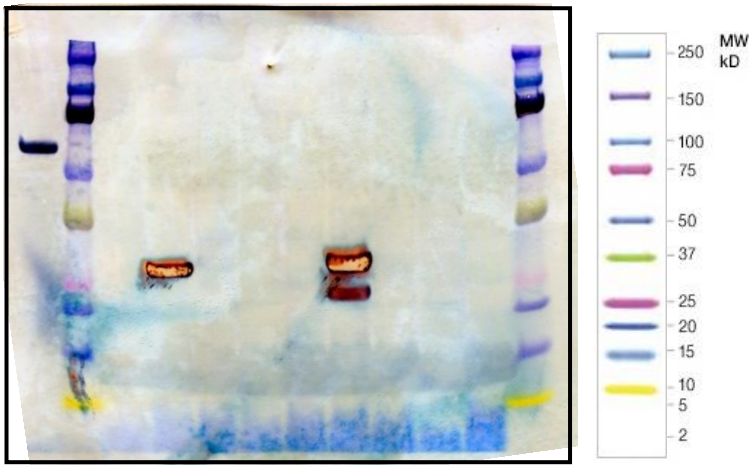


Figure 3.4 – Anti-Histidine Tag Western Blot showing the expression of sGFP and eCGP123. Lane 1 control, Lane 2 ladder, Lane 3 sGFP no promoter, Lane 4 sGFP, Lane 5 GAF no promoter, Lane 6 GAF, Lane 7 eCGP123 no promoter, Lane 8 eCGP123, Lane 9 YtvA LOV no promoter, Lane 10 YtvA LOV, Lane 11 WT, Lane 12 Ladder.

3.2.1.3 Thermostability Testing of Successful Candidates

Following the successful identification of two active reporter genes, it was essential to determine their range of thermostability. After growth at 52 °C, the cell lysate was extracted and exposed to temperatures varying from 25 °C to 75 °C. Following exposure to each specific temperature, the level of fluorescence within the sample was measured and normalised to protein concentration (mg/mL) (Figure 3.5). Both proteins demonstrated activity across the full range of temperatures tested. sGFP reached peak emission at 50 °C and then continued to steadily lose activity; whereas, the emission for eCGP123 remained relatively stable until 60 °C, after which the expression levels slowly declined.

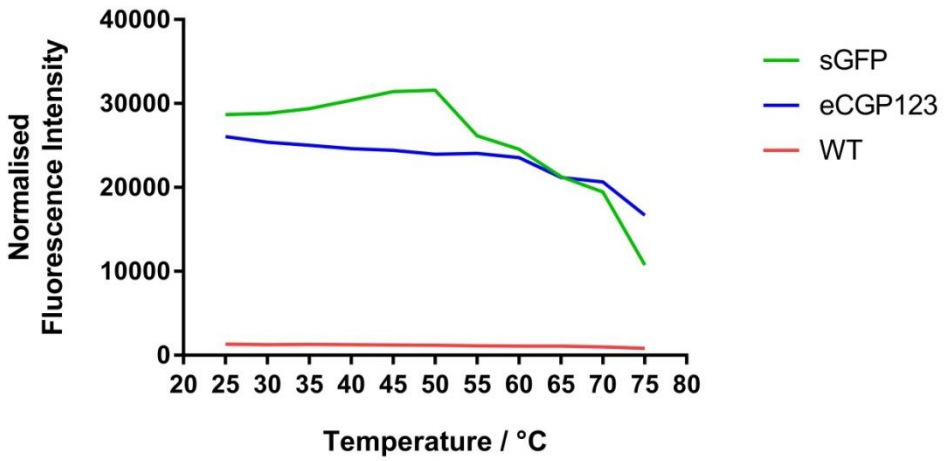


Figure 3.5 – A graph showing the thermostability of the fluorescent protein markers sGFP and eCGP123. The level of fluorescence intensity (RFU) per mg/mL of protein is plotted against temperature for each sample.

3.2.2 Development of a Promoter Library for *P. thermoglucosidasius*

3.2.2.1 Selection of Promoter Candidates and Development of a Modular Expression System

It was decided to select potential promotor candidates from genes within the glycolysis pathway and the TCA cycle, as these are known to be strongly expressed in *P. thermoglucosidasius*. The genome of *P. thermoglucosidasius* was screened and the specific genes selected were those encoding, glyceraldehyde 3-phosphate dehydrogenase (*gapd*) (91 bp), lactate dehydrogenase (*ldh*) (168 bp), citrate synthase (*cts*) (301 bp), fructose-bisphosphate aldolase (*fbpa*) (141 bp), glucose-6-phosphate isomerase 2 (*gpi2*) (135 bp), phosphoglycerate kinase (*pgk*) (101 bp), glycero-3-phosphate dehydrogenase (*gpd*) (94 bp) and 6-phosphofrucokinase (*pfk*) (169 bp). However, the consensus vegetative promoter sequences TATAAT and TTGACA, found at the -10 and -35 positions of *E. coli* and *B. subtilis* promoters, did not appear to be present within the promoters that were selected. In order to identify the promotor sequences, we attempted to use both promoter predicting software and manual alignment of possible promoter sequences, to try and elucidate the promoter consensus sequence within *P. thermoglucosidasius.* However, both methods failed to identify the presence of a specific consensus sequence.

As the extent of the promoter region could not be predicted, the entire region up stream of the targeted gene was selected. The RBS site for each gene had an easily identifiable consensus sequence, which enabled sequences to be taken upstream from the RBS site to the stop codon of the adjacent gene. This method also provided confidence that any additional regulatory sequences associated with the promoter had also been included. The promoters were originally cloned into the pMTL_TRsG vector, upstream of the sGFP gene and using our synthetic Shine-Dalgarno (SD) sequence as the RBS.

To facilitate the future interconversion of bioparts, a standardised vector was created. This was achieved by starting with the empty vector pMTL61110 and editing the application module. The design of the application module allowed for a "cut-and-paste" type approach to bioparts, to allow easier screening. The vector format is illustrated in figure 3.6. It consists of the double transcriptional terminator T1T2 of the *E. coli rrnB* gene flanked upstream by a SbfI site and NotI restriction site, followed by a promoter, an EcoRI site, a synthetic RBS sequence, an XbaI site and sGFP, before ending with an AscI restriction site.

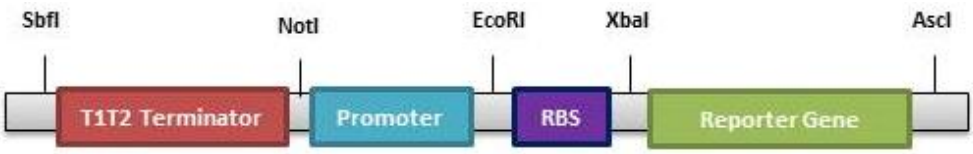


Figure 3.6 – A schematic representation of the application module designed for a quick and easy 'cut and paste' system to test different tools involved in gene expression, in addition to restriction enzyme sites used for cloning.

3.2.2.2 sGFP assay

The strengths of various gene expression elements were routinely assessed in this chapter using the following assay to measure sGFP expression:-

- *P. thermoglucosidasius*: the constructed gene expression vectors were transformed into *P. thermoglucosidasius* competent cells. After recovery, the transformed cells were plated onto TSA plates containing kanamycin (km) (12.5 µg/mL) and grown overnight. A single colony was picked and restreaked onto TSA plates containing km. A single colony was picked and cultured in a baffled conical flask for 8 hours at 250 RPM at 52 °C, in 2SPYNG media containing km (12.5 µg/mL), before extracting the protein lysate.

- *E. coli*: the constructed gene expression vectors were transformed into *E. coli* competent cells. After recovery, the transformed cells were plated onto LB plates containing km (50 µg/mL) and grown overnight.

A single colony was picked and cultured in a falcon tube for 8 hours at 200 RPM, in LB media containing km (50 μg/mL), before extracting the protein lysate.

The level of sGFP was measured using a micro-plate reader (CLARIOstar). The RFU were then normalised using the protein concentration of each sample. Protein concentration was calculated using the BCA assay.

3.2.2.3 Testing the Relative Strengths of Promoter Candidates

Each promoter was cloned into the application module of the sGFP vector. The strength of each promotor candidate was tested by measuring the fluorescence emission, using the standard protocol detailed in section 3.2.2.2, within both *E. coli* and *P. thermoglucosidasius* (Figure 3.7). The screening of promoters for expression in *P. thermoglucosidasius* produced a library with varying strengths. P_{gapd}, P_{ldh}, P_{cts}, P_{fbpa}, P_{gpi2} and P_{pgk} all showed a significant statistical increase compared to the control containing no promoter. In addition, P_{gapd} resulted in a higher expression compared to the previously highest known promoter, P_{ldh}. A 5-fold increase was shown for P_{gapd} compared to P_{pgk}, the weakest viable candidate.

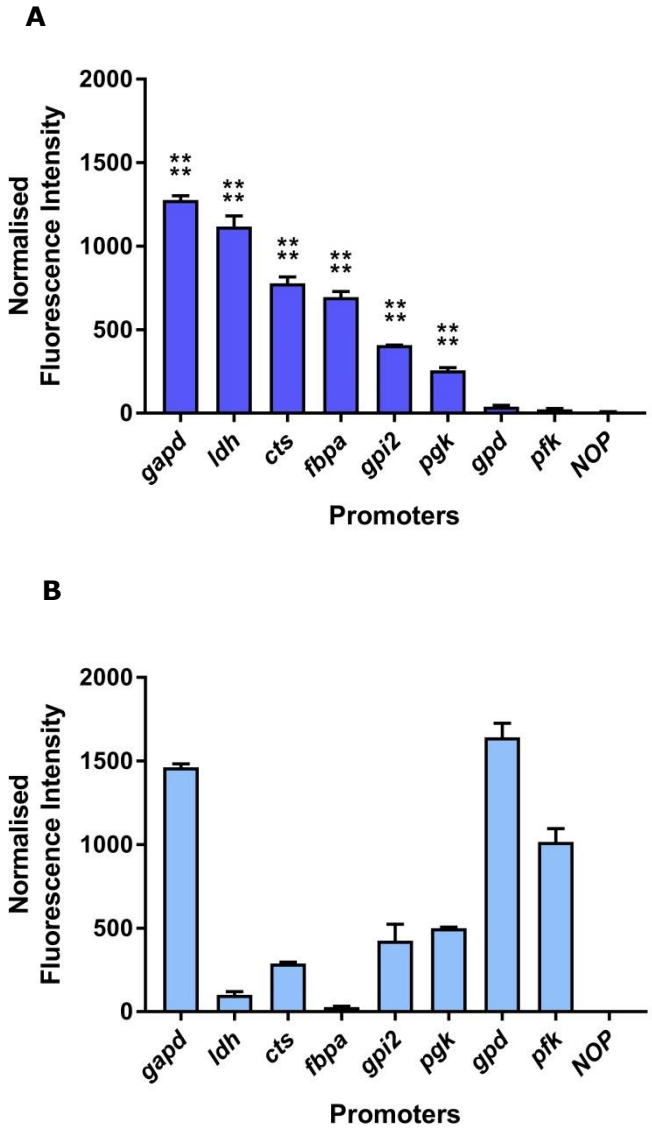


Figure 3.7 – The relative strengths of promoter sequences native to *P. thermoglucosidasius* expressed in both (A) *P. thermoglucosidasius* and (B) *E. coli*. The strength of each promoter was determined by the relative fluorescence units (RFU) of sGFP expression normalised to protein concentration (mg/mL). Promoters from left to right are as follows, those derived from the genes encoding glyceraldehyde 3-phosphate dehydrogenase (*gapd*), lactate dehydrogenase (*ldh*), citrate synthase (*cts*), fructose-bisphosphate aldolase (*fbpa*), glucose-6-phosphate isomerase 2 (*gpi2*) , phosphoglycerate kinase (*pgk*), glycero-3-phosphate dehydrogenase (*gpd*), 6-phosphofrucokinase (*pfk*) and no promoter (NOP). The error bars on the graph denote the standard error of the mean (n = 3). *, **, *** and **** denote levels of statistical significance from the negative control equal to the p values, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$, respectively, as determined by a one-way analysis of variance (ANOVA) assuming unequal variances.

Prior to specific organism testing, plasmids were first constructed in *E. coli*. However, the toxicity of exogenous genes can be problematic and so it is important to understand the activity of genetic expression elements within *E. coli* as well as within the specific organism of interest. Again, a variation in sGFP expression was observed when the promoters were tested in *E. coli*, with P_{gpd} producing the highest expression levels and P_{fbpa} the lowest. Though showing no significant expression in *P. thermoglucosidasius*, P_{pfk} and P_{gpd} showed high expression in *E. coli*. The use of P_{gapd} resulted in high expression in both organisms, whereas, P_{ldh} produced only high expression levels in *P. thermoglucosidasius*.

3.2.3 Validation of a Ribosomal Binding Site Library

An investigation into the relative effectiveness of a RBS library was also conducted. A library of SD sequences had previously been assembled based on sequences present in *Clostridium botulinum* and tested in *C. sporogenes* (Table 3.2). Their effectiveness was tested in *P. thermoglucosidasius*. The P_{gapd} promoter was kept constant and the distance between the SD sequence and the start codon of the gene was varied. The SD sequences were produced through Site-Directed Mutagenesis (SDM) and were tested using the sGFP assay (section 3.2.2.2). A positive control containing the synthetic RBS and a negative control, using no consensus RBS sequence, were also included.

Table 3.2 – Shine-Dalgarno sequences from different RBSs previously identified in *C. botulinum* (Kubiak, Unpublished). **Bold** text denotes the RBS sequence and the underscore indicates the start of the gene.

RBS name	RBS/spacer sequences selected
R1	TGTTTTTA**AAAGGAGGT**ATT_ATG
R2	TATAT**AAAGGAGGT**GCTACT_ATG
R3	TAAAA**AAAGGAGGT**GTTGAT_ATG
R4	ATAAAT**AAAGGAGGT**TATAT_ATG
R5	TTTT**AAAGGAGGT**ATAAGCT_ATG
R6	GAAAT**AAAGGAGGT**GAGTAT_ATG
R7	ACATA**AAAGGAGGT**GGCATT_GTG
R8	CTT**AAAGGAGGT**TTTATATT_ATG
R9	TAGAT**AAAGGAGGT**AGAACT_ATG
R10	AAT**AAAGGAGGT**ATTTAAAT_ATG
R11	GGAAT**AAAGGAGGT**GAGTAT_ATG
R12	AAAAGC**AAAGGAGGT**GTTTT_ATG
R13	GCTG**AAAGGAGGT**GAGAAAT_GTG
R14	AAT**AAAGGAGGT**TTATATAT_ATG
Synthetic RBS	ATT**GAAGGAGGA**TGTCTAGA_ATG

As before, the strength of each individual RBS was tested in both *P. thermoglucosidasius* and *E. coli* and the expression levels of sGFP are illustrated in figure 3.8. A large variation in the level of expression was observed in *P. thermoglucosidasius* dependent on the individual RBS employed. All except R1 showed a significant statistical increase compared to the negative control. Both R14 and R13 resulted in a higher expression than the synthetic RBS, which was previously the only available candidate. Furthermore, a 15-fold increase was observed between the most highly expressed RBS, R14, and the lowest, R1, demonstrating a large variety in strengths.

Unlike in *P. thermoglucosidasius*, the RBS library for *E. coli* did not produce much variation between the sequences. Expression was shown to be in general much lower compared to expression in *P. thermoglucosidasius*, except for the synthetic RBS and R2. The majority of sequences exhibited more than a 3-fold decrease in *E. coli* expression compared to the previously used synthetic RBS, including R14 which gave the highest *P. thermoglucosidasius* expression level.

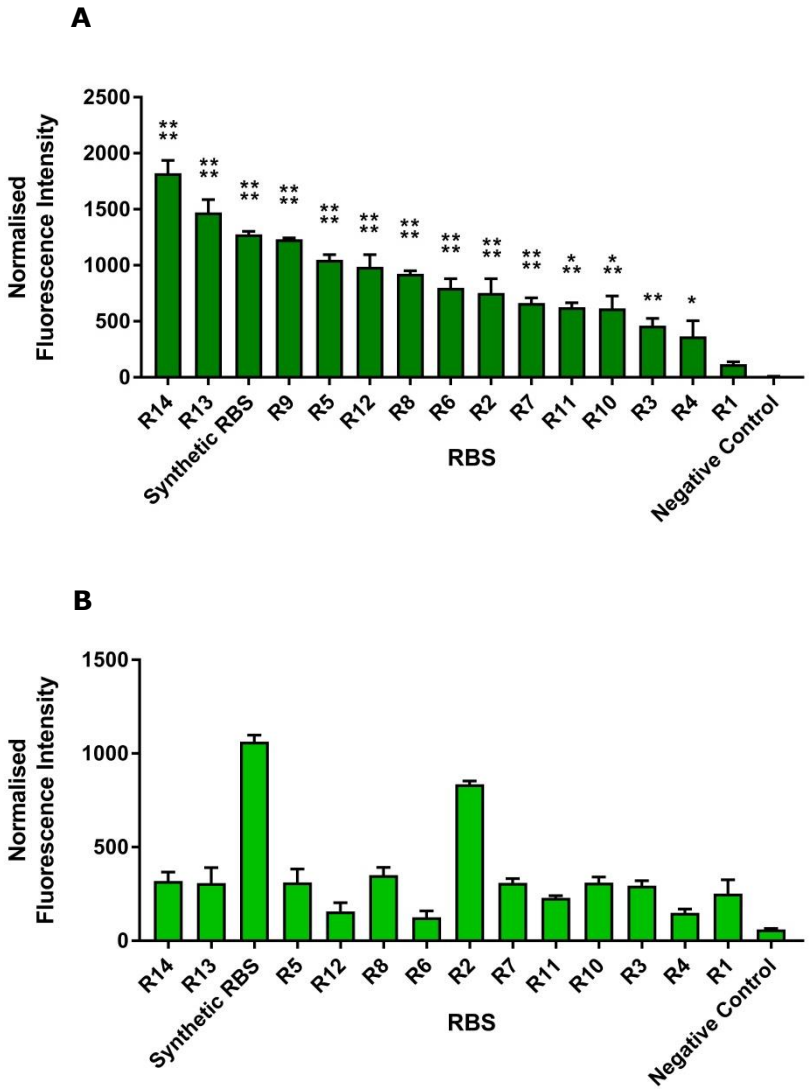


Figure 3.8 – The relative strengths of RBS sequences native to Gram-positive *C. botulinum*, used to express sGFP in both (A) *P. thermoglucosidasius* and (B) *E. coli*. The strength of each RBS sequence was determined by the relative fluorescence units (RFU) of sGFP expression, normalised to protein concentration (mg/mL). The positive control contained the published synthetic RBS sequence used in *P. thermoglucosidasius*, whereas the negative control contained no concensus RBS sequence. The promoter, P_{gapd}, remained constant throughout each of the RBS constructs. The error bars on the graph denote the standard error of the mean (n = 3). *, **, *** and **** denote levels of statistical significance from the negative control equal to the p values, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$, respectively, as determined by a one-way analysis of variance (ANOVA) assuming unequal variances.

3.2.4 Validation of Transcriptional Terminator Strengths

No quantitative data was available regarding the strengths of terminators available for use in *P. thermoglucosidasius*. It was therefore decided to examine the strengths of different Rho-independent terminators that had been previously identified in-house by Willson *et al.* (unpublished) for use in *Clostridium acetobutylicum*. The selected terminator sequences and organism of origin are summarised in table 3.3. Rho-dependent terminators require interaction with external factors, which often entails the use of extra genes and proteins. Often, these external factors do not successfully transfer between organisms. Therefore, using Rho-independent terminators, which only require self-annealing, offers a more reliable solution.

Table 3.3 – Rho-independent terminators selected as potential candidates for use in *P. thermoglucosidasius*. The organism, upstream gene and sequence are provided. Underlined residues contribute to the formation of a hairpin, as predicted by Integrated DNA technologies OligoAnalyzer 3.1 software (see section 2.1.4), with corresponding ΔG values given in the final column.

Terminator	Organism Isolated From	Upstream gene	Sequence	ΔG (predicted)
T*pepN*	*Lactococcus lactis* (171)	*pepN*	taatttataaataaaaatcacctttaggaggtggtttttttattttataaatt	-6.84
T*slpA*_LA	*Lactobacillus acidophilus* (172)	*slpA*	tgaaaaaggcagagcgaaagctctgtctttttt	-11.32

T*slpA*_CD	*C. difficile* ribotype 031/094 (173)	*slpA*	aaatataaaaagactctctcagatgagaagtctttttgtgaaa	-11.89
EcoT1	*E. coli*	*rrnB* T1 terminator	ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctc	-20.52
EcoT2	*E. coli*	*rrnB* T2 terminator	attaagcagaaggccatcctgacggatggcctttttgcgtttc	-10.8
phiTD1	*B. subtilis* Φ29 phage (174)	Late TD1 terminator	aacaatcaaaagaaaagcctatcgtctgaggaacggtaggctctttgtagcatatagttg	-8.6
T*tyrS*	*B. subtilis* (175)	*tyrS* tRNA	ataatcaatcgtccctcgtgtaaacgaaggggcgttttttattt	-12.25
T*gyrA*	*B. subtilis* (176)	*gyrA* gene	aagaagaagtgtgaaaaagcgcagctgaaatagctgcgcttttttgtgtcataa	-15.28
T*fdx*	*Clostridium pasteurianum* (157,177)	*fdx* gene	ataaaaataagaagcctgcatttgcaggcttcttatttttat	-18.34

To test each terminator, the sequences were cloned between the promoter (P_{gapd}) and the RBS sequence (the synthetic RBS) of the standard modular vector produced for sGFP. A strong terminator candidate can be identified two-fold, firstly by a fluorescence reduction compared to a system with no terminator and secondly, by a fluorescence reading similar to a system with no promoter. The sGFP assay was performed, as detailed in section 3.2.2.2, to determine the relative expression levels (Figure 3.9).

It was observed that all the tested Rho-independent terminators produced a statistically significant reduction in sGFP expression compared to the absence of a terminator. EcoT1, phiTD1, T*slpA_LA* and T*fdx* were the strongest candidates, all resulting in sGFP expression levels comparable to (no statistical difference) the system containing no promoter (NOP). The weakest terminator was determined to be T*pepN*, producing the highest fluorescence levels of all the terminators tested.

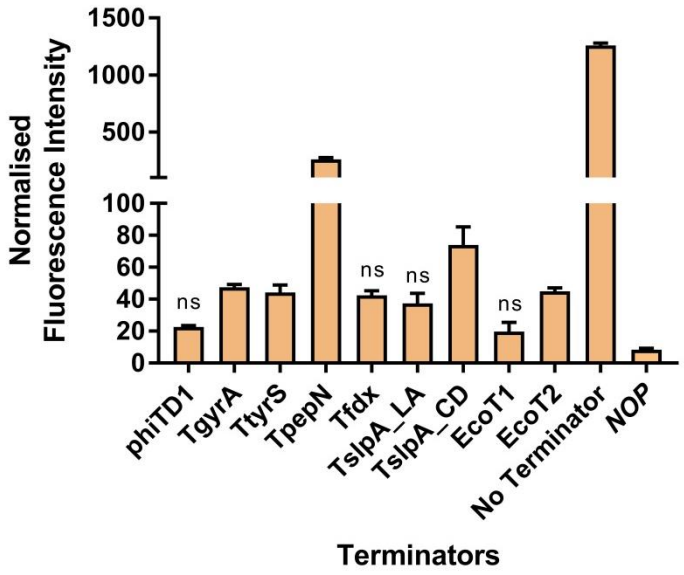


Figure 3.9 – The relative strengths of the Rho-independent transcriptional terminator sequences tested in *P. thermoglucosidasius*. Normalised fluorescence intensity was measured in relative fluorescence units (RFU) per mg/mL of protein. The terminators were expressed downstream of the promoter P_{gapd} used to express sGFP. The error bars on the graph denote the standard error of the mean (n = 3). All of the terminators show a statistically significant decrease compared to the sample containing no terminator. The samples showing no statistical difference compared to no promoter (NOP) are denoted by ns.

3.2.5 Codon Optimisation in *P. thermoglucosidasius*

3.2.5.1 Manual Examination of Codon Usage

Through manual assessment of the genes within an organism, it is possible to observe whether there is a natural tendency for the codon usage to be skewed. In order to assess this in *P. thermoglucosidasius*, four different genes were selected for analysis, those encoding: *pyrE*, *ack*, *gapd* and *fbpa*. For each gene, the amino acids were listed and the corresponding codon was recorded. Figure 3.10 highlights the codon bias for each individual amino acid, displaying the average number of times a specific codon was recorded as a percentage of the total.

From this initial data, it is apparent that a codon bias does exist within *P. thermoglucosidasius* and that this bias presented varying trends across the codons associated with different amino acids. For some amino acids, for example asparagine or serine, there appeared to be no bias, with equal codon frequencies recorded. However, for others there was a significant bias, either in favour or against a particular codon. The amino acid lysine is a clear example of this, demonstrating a striking preference for AAA over AAG. The codon bias presented for arginine was another very interesting result, with AGG showing no observed usage at all. Within arginine there is also bias against the use of AGA and in favour of CGC. These were trends that were replicated across numerous amino acids and demonstrate clearly the presence of a codon bias within *P.*

thermoglucosidasius. Also notable, was the fact that in all four genes the stop codon TAA was used exclusively.

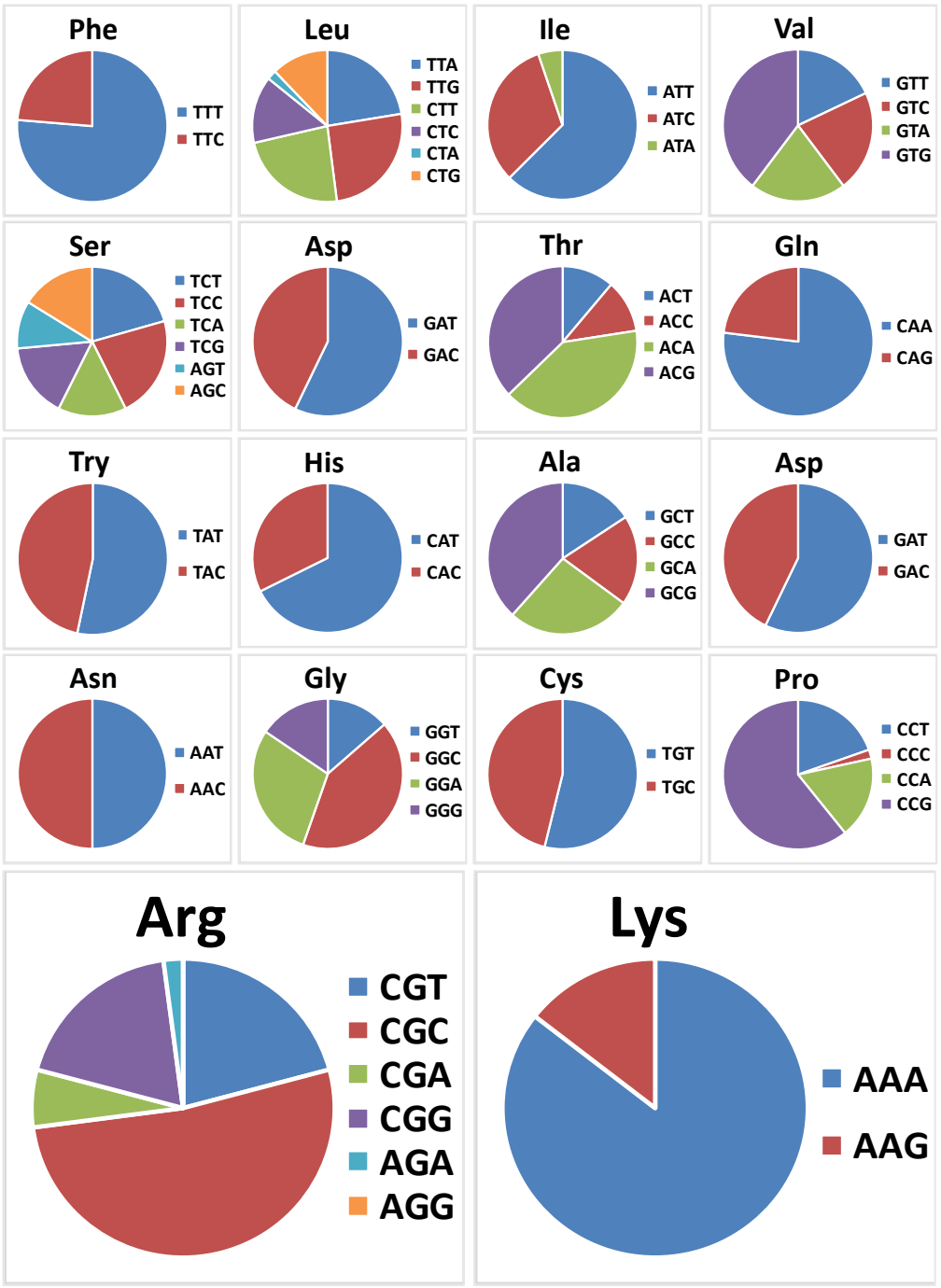


Figure 3.10 – Charts representing the average number of times a specific codon codes for each amino acid within a gene. The codons within the genes coding for orotate phosphoribosyltransferase (*pyrE*), acetate kinase (*ack*), glyceraldehyde 3-phosphate

dehydrogenase (*gapd*) and fructose-bisphosphate aldolase (*fbpa*) were manually counted and recorded. The amino acids methionine (Met) and tryptophan (Trp) are not shown as they are only coded by a single codon.

3.2.5.2 Algorithmic Codon Optimisation using the pheB Gene

The industrial collaborator, ATUM, designed and synthesised 23 variants of the *pheB* gene, from *G. stearothermophilus,* containing a variety of codon variations (178). These variants were transformed into the plasmid pTMO_372, into *E. coli*, before extracting the plasmid and transforming into *P. thermoglucosidasius*. Three colonies from the pates were picked and cultivated to provide three biological replicates for each sample and then were frozen in glycerol stocks at -80 °C. These replicates were revived overnight on TSA km (12.5 µg/mL) plates and used to inoculate 2SPYNG (50 mL) containing km (12.5 µg/mL), to a starting OD_{600} of 0.3. This culture was allowed to grow for 8 hours, before the cells were harvested and the protein lysate extracted. The *pheB* assay (section 2.12.1) was then carried out with the WT *pheB* gene was used as a control. The codon algorithm for each sample was not publicly available; however, the results show clear variation in the activity of catechol 2, 3-dioxygenase, which therefore infers that there is a difference between using specific codons (Figure 3.11). Around two thirds of the variants presented no statistical increase in activity compared to the WT, with several actually exhibiting a decrease. Samples 1, 12, 16, 19, 21, 22 and 23

did show a significant statistical increase, with sample 21 displaying the highest activity, a 2.6-fold increase compared to the WT gene.

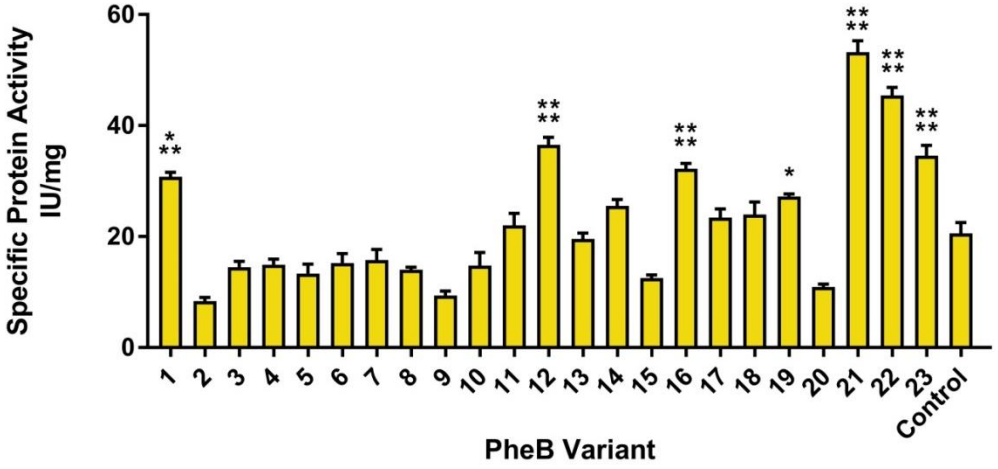


Figure 3.11 – A graph showing the specific protein activity of 23 variants of the *pheB* gene resulting from the optimised translation of codons. The WT *pheB* gene was used as a control. The error bars on the graph denote the standard error of the mean (n = 3). *, **, *** and **** denote levels of statistical significance from the negative control equal to the p values, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$, respectively, as determined by a one-way analysis of variance (ANOVA) assuming unequal variances.

This type of analysis represents only the initial phases of a programme of work designed to derive an improved algorithm for predicting codon usage (178). The data obtained is used to guide the design and synthesis of two further sets of variant genes, which are then tested in terms of expression levels in *Parageobacillus.* However, due to time constraints the study was not completed.

3.2.6 Initial development of an Inducible Expression System for *P. thermoglucosidasius*

3.2.6.1 Riboswitches

It was decided to evaluate whether riboswitches E, F and G, discussed in section 3.1.1, could be used in *P. thermoglucosidasius.* These riboswitches respond to theophylline, so the effect of varying concentrations of theophylline on *P. thermoglucosidasius* viability had to be initially established. *P. thermoglucosidasius* was grown for 8 hours in the presence of varying theophylline concentrations, ranging from 0 – 10 mM, and the cell density was measured to monitor growth. OD_{600} measurements were taken for each sample after 8 hours of growth (Figure 3.12). There was a general linear decrease in cell density with increasing theophylline concentration. However, even at high concentrations, where growth rate appeared to slow, growth was still observed. A decrease in growth of 50 % was observed when the maximum concentration of theophylline was added.

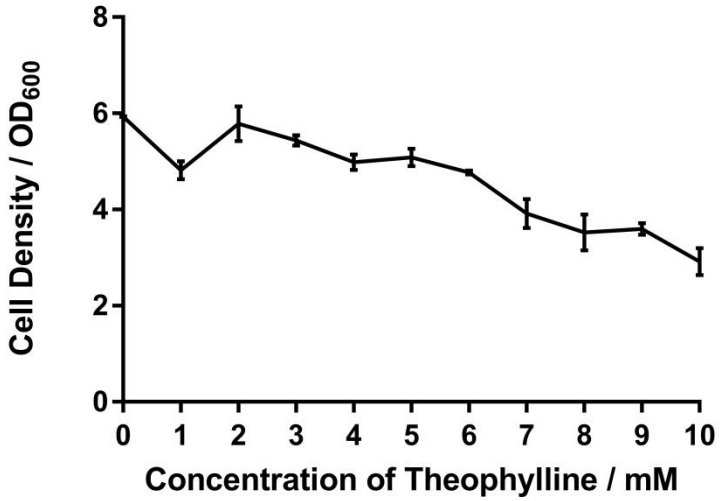


Figure 3.12 – The effect of theophylline concentration on the growth of *P. thermoglucosidasius*. *P. thermoglucosidasius* was grown for 8 hours at 52 °C in a baffled conical flask shaking at 250 RPM in the presence of varying concentrations of theophylline, ranging from 0 – 10 mM. Growth was determined by cell density, calculated by measuring the optical density at a wavelength of 600 nm (OD_{600}). The error bars on the graph denote the standard error of the mean (n = 3).

After establishing the tolerance of *P. thermoglucosidasius* to theophylline, the optimum concentration of theophylline for induction of the riboswitches was established. Riboswitch E, identified from the literature, was cloned downstream of P_{gapd} and upstream of the reporter gene sGFP, replacing the synthetic RBS sequence. The complete plasmid was transformed into *P. thermoglucosidasius* and the sGFP assay was performed as before (section 3.2.2.2), but instead adding theophylline to the cultures, at varying concentrations. Figure 3.13 illustrates the expression levels relating to the varying concentrations and it can be observed that there was generally a linear response, with an increased concentration producing an increased level of sGFP. However, this linear

response was observed only up to a maximum of 8 mM, at which point the response plateaued. The increased expression observed at 8 mM was observed to be higher than the positive control, which contained no Riboswitch E. However, it is also shown that the use of Riboswitch E does not completely negate expression of sGFP as there was still fluorescence observed when uninduced, at 0mM of theophylline, and therefore as an inducible expression system is could be considered as 'leaky' or not optimal.

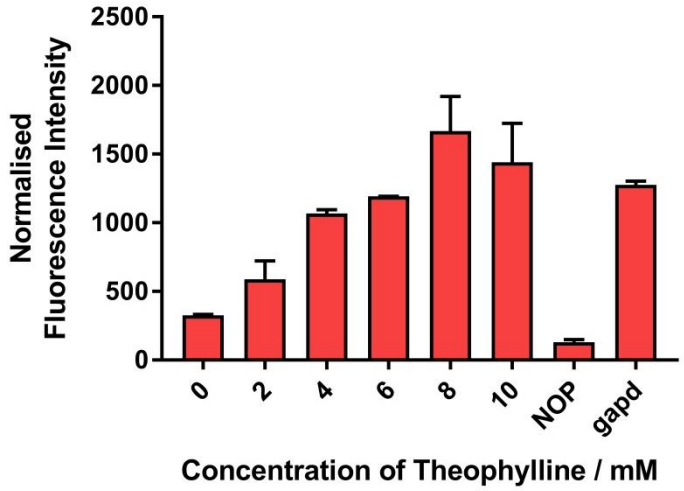


Figure 3.13 – The effect of theophylline concentration used for induction of Riboswitch E on sGFP expression in *P. thermoglucosidasius*. Normalised fluorescence intensity was measured in relative fluorescence units (RFU) per mg/mL of protein. The positive control used the P_{gapd} promoter and synthetic RBS sequence with no riboswitch, whereas the negative control had no promoter sequence (NOP). The error bars on the graph denote the standard error of the mean (n = 3).

After determining an optimal theophylline concentration of 8 mM, testing commenced to evaluate the expression levels for all three riboswitches. Each riboswitch was cloned downstream of three promoters of different

strengths, specifically P_{gapd}, P_{fbpa} and P_{pfk}. The sGFP assay was used as previously described (see section 3.2.2.2), with theophylline (8 mM) added after 2 hours for induction. RFU values per protein concentration were obtained for all samples (Figure 3.14). The expression profiles shown in figure 3.14, illustrated that sGFP expression was increased in theophylline-induced samples. The addition of a riboswitch did not change the relative strength of each promoter, with P_{gapd} remaining the strongest and P_{pfk} the weakest for all riboswitches. Furthermore, no riboswitch provided complete absence of sGFP expression when uninduced. From the data shown in figure 3.14, it can be suggested that the stronger the initial promoter the larger the increase in expression from the uninduced to the induced systems. All three riboswitches, E, F and G, under the control of both promoters, P_{gapd} and P_{pfk}, showed a large increase in expression over the constitutive promoters alone. Overexpression was observed for Riboswitch G compared to the constitutive P_{fbpa}, but this was not the case for Riboswitches E and F. Overall, Riboswitch G appears to have the strongest induction, followed by F and then E.

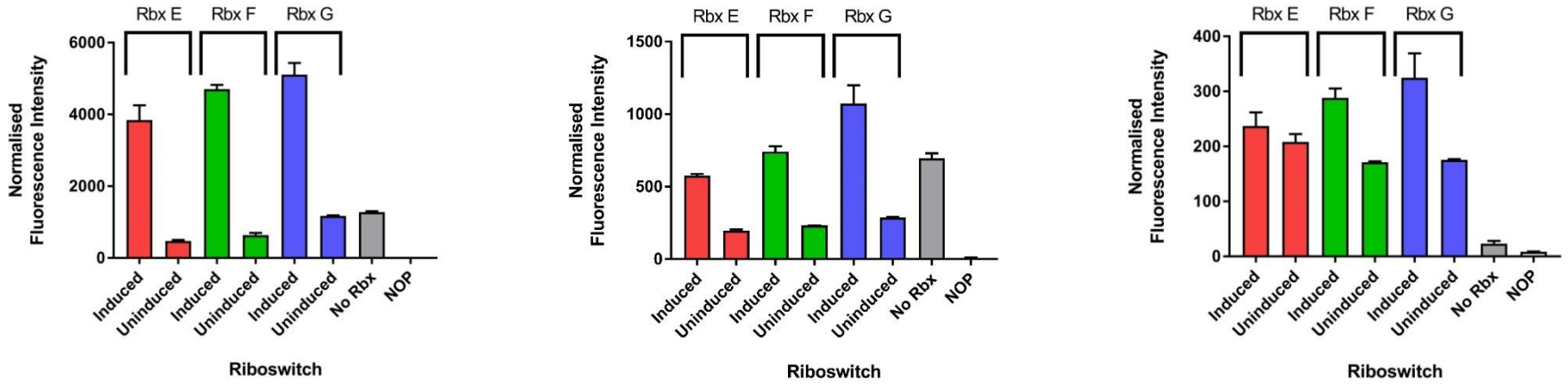


Figure 3.14 – The expression profiles for three different riboswitches, Riboswitch E (Rbx E), Riboswitch F (Rbx F) and Riboswitch G (Rbx G), under the control of three different promoters of varying strength, P_{gapd}, P_{fbpa} and P_{pfk} , respectively. The strength of each promoter-controlled riboswitch was determined by assessing relative normalised fluorescence intensity, measured in relative fluorescence units (RFU) per mg/mL of protein. Each riboswitch was measured in an induced and uninduced state, along with controls that contained the synthetic RBS with no riboswitch (No Rbx) and a construct containing no promoter (NOP). The error bars on the graph denote the standard error of the mean (n = 3).

3.3 Discussion

An increasing demand to produce alternative energy sources that are both renewable and carbon-neutral has led to an investigation into thermophiles, such as *P. thermoglucosidasius,* as potential industrial organisms to produce desirable products. Prior to wide scale industrial use, the organism must first be optimised, to improve performance and productivity, and it is therefore essential to have genetic tools and bioparts capable of achieving these goals. However, due to the novelty of the field, the multitude of tools that exist for well-studied organisms, such as *E. coli* and *S. cerevisiae,* do not yet exist for *P. thermoglucosidasius* (52). This work has therefore taken steps towards increasing the availability of thermostable tools that can be utilised to explore the potential of *P. thermoglucosidasius.*

Validation of Two Fluorescent Reporters

Fluorescent reporter genes are a common tool used in microbiology for the screening of gene expression and protein localisation. However, at the start of this project there was no fluorescent reporter protein available for use in *P. thermoglucosidasius*. Through testing of both commercially available proteins and candidates selected specifically from the literature, this work describes the successful identification and subsequent validation of two different thermostable reporter proteins, sGFP and

eCGP123, active in *P. thermoglucosidasius* at its optimum growth temperature of 60 °C. By observing protein expression, through both the emission of fluorescence and a western-blot, the results determined that the fluorescent reporter genes were successfully expressed within *P. thermoglucosidasius*. Furthermore, thermostability testing showed that both proteins maintained their ability to fluoresce up to and including 75 °C. The validation of these two reporters enable simple testing of gene expression libraries to be conducted, in addition to an ability to perform protein localisation studies within *P. thermoglucosidasius*.

These results corroborate previous findings that show both sGFP and eCGP123 expression at high temperatures (165,169). *In vivo* expression of sGFP has previously been shown in both *T. thermophilus* at 70 °C (165) and *G. stearothermophilus* at 60 °C (166); whilst eCGP123 fluorescence has been successfully demonstrated *in vitro* at 80 °C (169) and *in vivo* in the thermoacidophilic archaeon *Sulfolobus acidocaldarius* (179). However, importantly, this study details the first *in vivo* expression of eCGP123 in a bacteria species. Following the completion of this work, expression of sGFP in *P. thermoglucosidasius* was also confirmed by Reeve *et al.* (180), supporting the findings of this study. One area that could be further explored is the use of these fluorescent proteins as *in vivo* localisation-tags, in order to show protein distribution. This would present a significant advantage for exploring the biology of *P. thermoglucosidasius* and sGFP has already shown success in this area, used in *T. thermophilus*

to visualise the distribution of three proteins, GroES, NarC and PhoA (165).

One interesting, yet currently unexplained observation relating to sGFP, is that there was a non-homogenous expression across the population, with some cells failing to fluoresce. This was observed through analysis of these cultures using a confocal microscope, to isolate fluorescence from individual cells. There are several possible reasons for this observation. Firstly, fluorescence can be quenched within a sample in many ways, such as collisional quenching, excited state reactions and other energy transfers, e.g., Förster resonance energy transfer (FRET) (181). It has been shown that collisional quenching can be caused by molecular oxygen (182). Hence, an increased length of time exposed to oxygen on a slide could result in a lowered fluorescence intensity. Another possible reason for the low percentage of fluorescence across the population may be cell death and/or a slowing of metabolic activity, due to the reduction in temperature from 52 °C to room temperature, at which the analysis took place. However, the precise reason for this is still unclear and needs further investigation. This non-homogenous expression across the population has not been discussed by any paper currently describing the use of sGFP in thermophiles. Conversely, the lower levels of fluorescence may be a consequence of reduced levels of transcription. It has been shown that the P_{ldh} exhibits low expression during aerobic growth (159). However, one possible way to explore this may be to immobilise the cells and immediately treat with formaldehyde to 'capture' the active state of

the bacteria which was performed by Cava *et al.* (165) when determining expression in *T. thermophilus*.

In addition to testing sGFP and eCGP123, two other thermostable fluorescent reporters identified from the literature were tested, YtvA LOV and GAF. However, this study was unable to demonstrate fluorescence of these reporters at 52 °C in *P. thermoglucosidasius*, suggesting that these candidates are unsuitable for use in the *Parageobacillus* chassis. A HIS-tag western blot, using the protein lysate from *P. thermoglucosidasius*, determined that the protein expression of the two genes was unsuccessful. The thermostable variant of YtvA LOV, tested in this study, was created from the native protein, found in *B. subtilis,* through rational design and tested *in vitro* following production in *E. coli* (170). Similarly, the GAF domain, isolated from two thermophilic cyanobacteria *Synechococcus* species, was produced through truncation of the full length *SyA*-Cph1 and *SyB*-Cph1 gene and expressed in *E. coli* before *in vitro* testing of thermostability (167). However, currently, no *in vivo* expression has been shown in thermophiles for either of the suggested candidates.

The failure to express YtvA LOV and GAF *in vivo* in *P. thermoglucosidasius,* suggests that translation of these genes is affected. As discussed in sections 1.4.2 and 3.1.1, the degeneracy of the genetic code can cause an evolutionary selection bias towards certain nucleotide sequences and tRNA levels (79,81). It has been shown that this selection bias leads to codon bias that differs between species (77,78). Therefore, it can be

suggested that the lack of protein expression in *P. thermoglucosidasius* could be explained by differences in codon usage of the synthesised genes. This could also explain the lack of fluorescence observed for the commercial reporter proteins tested, DasherGFP, CayenneRFP and PaprikaRFP, of which vivid colour observed in *E. coli* was eliminated in *P. thermoglucosidasius*. The gene expression elements selected to drive transcription and translation of the reporter genes, P_{ldh} and the synthetic RBS sequence, were later demonstrated to be highly proficient. This suggests that these elements are not the cause of the lack of protein expression. Despite the lack of protein expression, the true cause is still unknown. However, due to time pressures and the success in identifying two usable reporter genes, no further investigation was conducted.

Development of a Promoter Library

Promoters are a key regulator of gene expression and it is essential to have a characterised library available that can exploit a range of strengths, to allow fine-tuning of synthetic systems. At the beginning of this project, the number of promoters available for selection in *P. thermoglucosidasius* was inadequate, consisting of a single, strongly expressed promoter, P_{ldh} from *G. stearothermophilus*, and a limited number of weakly expressed promoters (7,159). Therefore, constructing a promoter library of varying strengths would end the reliance on a single promoter and improve the potential for successful genetic engineering of *P. thermoglucosidasius*.

This work describes the production of a promoter library containing a range of strengths, in addition to a promoter (P_{gapd}) that shows higher expression than P_{ldh}, the strongest promoter previously (7). Candidates were selected from the glycolysis pathway and the TCA cycle, highly expressed metabolic pathways, and used to drive the expression of the newly identified sGFP reporter gene. As mentioned, the only significant candidate previously available for selection for gene expression in *P. thermoglucosidasius* was P_{ldh}, identified from *G. stearothermophilus* (24). Cripps *et al.* (24) characterised the strength of this promoter alongside P_{glpD} from *G. thermodenitrificans*, the P43 promoter from *B. subtilis* and seven native glycolytic promoters. The seven native *P. thermoglucosidasius* promoters all exhibited very low expression, showing a limited increase compared to the empty control plasmid. The results discussed in this study demonstrate a clear improvement on this, validating several significantly expressed promoters and offering a much wider range of strengths. In a similar method to this study, Lin *et al.* (7) took 200 – 300 bps upstream from the start codon of the gene to acquire promoters. However, this selection process includes the native RBS sequence of the gene and therefore expression of the reporter is determined by two factors. This restricts the ability to accurately quantify the strengths of different promoters relative to one another. The promoters characterised in the present study are all under the control of an identical RBS, allowing confident comparisons. The current strongest and most widely used promoter, P_{ldh}, has been shown to have an expression profile dependent on redox conditions (159), exhibiting strong expression under aerobic

conditions yet low expression under anaerobic conditions, leading to concerns about its wide spread use. Therefore, the identification of a stronger promoter (P_{gapd}), validated by this study, eliminates this dependence on a single, variable promoter.

Although the promoter library produced in this study has demonstrated a significant advancement for promoter use in *P. thermoglucosidasius*, since completion, two further promoter libraries have been successfully constructed (180,183). Reeve *et al.* (180) produced a promoter library through mutagenic PCR of the promoter from the ribosomal protein RplS, which has also shown strong expression in other *Bacilli*. The library exhibited a 100-fold range and produced one promoter stronger than the WT. Pogrebnyakov *et al.* (183) instead reproduced a method described by Jensen and Hammer (184), detailing the production of a promoter library in *L. lactis,* through the randomisation of the spacer sequences between the -10 and -35 consensus sequence. Using the native P_{groESL} promoter, Pogrebnyakov *et al.* (183) were able to produce a library with a 76-fold range in expression with sequences that vary more significantly from the native sequence than that of the library produced by Reeve *et al.* (180). In comparison to the library produced within this study, these two libraries offer significantly more promoters, with a considerable broader range of expression. It could be argued that the possibility of homologous recombination could occur within the library produced within this study, due to the use of native promoters. This has been suggested by Reeve *et al.* (180) and Pogrebnyakov *et al.* (183), both of whom produced a

synthetic library generated from a native promoter. However, one advantage of using native promoters is that it ensures any transcription factor binding sites or other interactions within the 5' untranslated regions (UTRs) are present, which could contribute to gene expression variability. Furthermore, it has previously been suggested that the context of the promoter may be as important as the -35 and -10 regions (185,186).

The promoter candidates were tested in parallel in both *P. thermoglucosidasius*, a Gram-positive bacterium, and *E. coli*, a Gram-negative bacterium. As plasmids are first constructed in *E. coli*, it is important to also understand the activity of genetic expression elements within this species, in case toxicity problems arise for the exogenous genes of interest. Therefore, knowledge of promoter strength within *E. coli* enables a more informed choice when initially building operons. Promoter strengths within the library were shown to vary between the two species, a result also observed in the synthetic library constructed by Reeve *et al.* (180). The wide range observed in both species, in addition to reverse expression profiles exhibited by some of the promoters, is of clear benefit to genetic engineering in *P. thermoglucosidasius*.

Regarding future work, there are several advancements that could be made. Firstly, temperature, media, and growth conditions have all been shown to have an effect on the strength of promoters (187). Therefore, by varying these conditions, a more intricate picture of the expression profile of these promoters can be obtained. Furthermore, as these promoters are

native, this would provide further insight into the growth of *P. thermoglucosidasius* in general.

Construction of a Ribosomal Binding Site Library

SD sequences, the RBS on mRNA, represent another essential tool to enable gene expression. Heterologous expression often utilises the native SD sequence; however, by using the native sequence, translation efficiency is unknown. Therefore, to achieve confident expression levels, a RBS library should also be produced. Currently within *P. thermoglucosidasius*, heterologous expression has occurred using the native SD sequences and no RBS library exists.

This work demonstrates the successful production of a RBS library containing a comprehensive range of strengths. In addition, the library produced two SD sequences that showed higher expression than the current in-house synthetic RBS. The RBS sequences were identified from *C. botulinum*, a pathogenic anaerobe, and using the sGFP reporter, their relative strengths were established. On completion, this represented the only RBS library tested in *P. thermoglucosidasius*. The binding of the ribosome to the SD sequence occurs through the anti-SD sequence present in the 16S rRNA. The production of the RBS library is based upon this consensus sequence, identified from several Gram-positive bacteria. These included several clostridial species and *B. subtilis*, the sequences of which were all observed to be identical (Kubiak, Unpublished) (Table 3.4). The

16S rRNA SD consensus sequence identified from *P. thermoglucosidasius* showed variation from this sequence and may lead to a weaker response than a true consensus match. Therefore, more information could be gained by changing the consensus sequence of the library to match the consensus sequence observed from the *P. thermoglucosidasius* 16S rRNA. The SD sequence greatly influences expression, through the efficiency of the ribosome to bind to the SD sequence. Weak base pairing with the SD sequence can lead to interference of initiation by mRNA secondary structure; whereas, a sequence with a strong affinity can be deleterious through stalling of initial elongation of translation (190-193). The context of the DNA surrounding the SD sequence also requires consideration, as mRNA structure can affect expression. Occlusion of the RBS and/or the start codon, by the secondary structure of the mRNA, has been demonstrated to impair protein expression (86,194-196), and therefore should also be analysed when testing and producing synthetic RBS.

Table 3.4 – A table detailing a segment of 16S rRNA sequences, containing the SD consensus sequence, from several species, showing the anti-SD of the ribosome. This was used for construction of the library consensus sequence for production of the RBS library for use in *C. sporogenes* (Kubiak, Unpublished). The difference between the library consensus sequence and the anti-SD of *P. thermoglucosidasius* is displayed.

Species	16S rRNA sequence
P. thermoglucosidasius	3'-GAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGC-5'
B. subtilis	3'-ATCTTTCCTCCACTAGGTCGGCGTCCAAGAGGATG-5'
C. acetobutylicum	3'-ATCTTTCCTCCACTAGGTCGGCGTCCAAGAGGATG-5'
C. difficile	3'-ATCTTTCCTCCACTAGGTCGGCGTCCAAGAGGATG-5'
C. sporogenes	3'-ATCTTTCCTCCACTAGGTCGGCGTCCAAGAGGATG-5'
C. botulinum	3'-ATCTTTCCTCCACTAGGTCGGCGTCCAAGAGGATG-5'
Library consensus sequence	5'-AAAGGAGGTGT-3'

Following this, studies achieved by Reeve *et al.* (180) and Pogrebnyakov *et al.* (183) also produced a simple synthetic library containing four and six RBS sequences, respectively. The RBS library produced in this work is therefore still the largest in existence. The use of both heterologous and synthetic RBS sequences, as detailed in this study and the following studies by Reeve *et al.* (180) and Pogrebnyakov *et al.* (183), offer an advantage over a purely native RBS library, by limiting the potential for homologous recombination. However, in the future, a more effective library could be produced using the method employed by Guiziou *et al.* (76). This identified eight highly expressed constitutive genes from within *B. subtilis* and produced a synthetic library by maintaining the SD consensus sequence, GGAGG, and varying the surrounding sequences.

By using a constant SD consensus sequence throughout the SD sequences tested, the relationships between the length of the spacer sequence and the strength of the RBS can be compared. The results appeared to show a positive correlation between the two; however, as this correlation was weak it was apparent that additional parameters were also affecting the RBS strength. It has previously been recognised that the spacer sequence length does affect the strength of the RBS (197). However, ultimately the overall sample size of the library needs to be increased in order to further investigate this phenomenon. Furthermore, it may be that the sequence upstream of the consensus sequence is having an effect on the relative strengths, due to the production of secondary structures.

As before, the SD sequences were also tested in *E. coli*, demonstrating a relatively low overall expression compared to the previously used synthetic RBS. This low expression is beneficial as expression in *E. coli* is often unnecessary; however, it is also unusual as the SD sequence, AGGAGG, has been shown to be the consensus sequence of *E. coli* (198). Furthermore, a range of spacer sequences, from 5 – 13 bps, have shown success in *E. coli*, with the optimum spacer sequence being 5 bps (197). As discussed, further exploration of this RBS library can be achieved by utilising the method employed by Guiziou *et al.* (76), to produce a larger and potentially more efficient library.

Testing of Transcriptional Terminators

Transcriptional terminators are necessary to stop transcription and cause the RNA polymerase to release mRNA. Often, strong promoters can result in transcriptional read-through past weak terminators, which affects gene expression. Transcriptional termination can be achieved through Rho-dependent and -independent terminators. Rho-independent terminators are an ideal candidate for use in *P. thermoglucosidasius* as they produce a self-annealing, secondary structure, such as a hairpin loop, to cause termination, as opposed to the external factors used by Rho-dependent terminators. There are many available and characterised Rho-independent terminators that can be utilised. However, these terminators are affected by high temperatures, which cause melting of the RNA structure (199,200). For this reason, in addition to the absences of any terminator testing in *P. thermoglucosidasius*, construction of a terminator library has been achieved to expand the application of genetic engineering in this organism.

Through the development of the modular system, to 'cut-and-paste' gene expression elements into the sGFP reporter vector, the successful testing and production of a transcriptional terminator library is presented. This study aimed to assay the strength of the terminators in stopping transcription of the sGFP reporter gene, by placing them between the promoter and the gene. Terminator strength was determined by the lack of sGFP expression observed, demonstrating that the RNA structure was

maintained and transcription could not occur. All the terminator candidates significantly inhibited the transcription of sGFP compared to the control, at the temperature assayed. Importantly, the study validated the use of four terminator constructs that were able to maintain a fluorescence emission level equivalent to a construct containing no promoter. T*pepN* showed the weakest level of termination. This agrees with the predicted stability suggested by the free energy of its secondary structure, which was the lowest of all the terminator candidates. Current identification of terminators from *Geobacilli* species is limited. The use of EcoT1 and EcoT2 terminators has been observed in plasmids constructed for *G. stearothermophilus* (201). However, this library represents the only transcriptional terminator library currently available for *P. thermoglucosidasius*. In future work, it would be interesting to explore the identification and testing of native terminators, to determine whether comparatively their efficiencies would be higher. Identification of terminators in *G. stearothermophilus* has been previously shown (202), suggesting that it may also be possible for *P. thermoglucosidasius*.

Codon Optimisation in *P. thermoglucosidasius*

As previously discussed in sections 1.4.2 and 3.1.1, it is well established that the degeneracy of the genetic code has led to codon bias. This bias is often observed differently between different organisms (203), which can cause problems during heterologous expression if the codon usage between the organisms is exceedingly different. Often, it can be corrected

for by synthesising the gene using codons that are optimised for the specific organism. However, this cannot occur if the codon bias of the organism is not first established. Currently, there is no established codon usage system available for *P. thermoglucosidasius*.

This research has taken steps towards identifying codon bias within *P. thermoglucosidasius* and producing an optimised codon usage system. In an initial approach similar to Sharp and Li (82), manual analysis of codon bias was conducted on four key genes, *gapd*, *fbpa*, *pyrE* and *ack*. These genes were selected due to their roles in the glycolysis pathway, uracil biogenesis and mixed acid fermentation respectively. Sharp and Li (82) developed the Codon Adaptive Index (CAI), which is a measure of codon bias for highly expressed genes. Similar to the technique described here, the CAI measures the frequency of codons in certain genes and determines the 'favoured' codons, allowing synthesis of genes adapted to the codon bias of the organism. This initial screening displayed evidence that a codon bias does occur in *P. thermoglucosidasius*, with certain codons observing more frequent inclusion, or discrimination. For instance, the codon AGG showed no selection across all four genes, suggesting evolutionary selection against this codon. It could be suggested that this is due to a similarity to the SD consensus sequence. However, reasons for codon selection bias are difficult to elucidate and this initial study is largely limited by the limited number of genes selected for been assessment. One further consideration is that manual analysis only provides information on the frequency of codons and not on the

placement of the codon within the gene. It may be, for example, that one codon may feature more heavily at the beginning or the end of the gene. Therefore, codon placement is a factor that should be a key consideration for future algorithms designed for codon optimisation. A similar codon usage system, tRNA Adaptive Index (tAI), has also been proposed, whereby the cognate tRNA concentrations are considered as opposed to the codon frequency (204). However, the tAI makes the assumption that the tRNA concentration is static and that high concentration of tRNA correlates with high expression.

Although no codon usage system is available for *P. thermoglucosidasius*, codon optimisation data is available for several other *Geobacilli* species (158) and it has also been successfully implemented to produce a codon optimised sGFP for expression in *G. stearothermophilus* (166). A process call 'codon harmonisation' has however been investigated within *P. thermoglucosidasius*, in order to express a pyruvate decarboxylase gene from *Gluconobacter oxydans* (205). The frequency, or infrequency, of codons affects the speed of translations, which can often affect the folding of the protein and specific domains. Therefore, codon harmonisation attempts to match the codon frequencies for specific regions within the native gene with native host codons that match these frequencies (206-208). The codon harmonisation of the pyruvate decarboxylase gene in *P. thermoglucosidasius* enabled successful expression of the gene, demonstrating high activity despite previously being shown to not express at temperatures over 45 °C (205).

These methods discussed, have shown success in several organisms; however, more recently, studies have demonstrated additional significant factors affecting protein expression (86,178,209,210). Further evidence has also demonstrated that utilising the most common codon can cause detrimental effects (178,211). These studies have demonstrated that local and global variability in the mRNA sequence can have a drastic effect on expression. Through selecting the highly expressed codon or tRNA, factors such as mRNA structure, cryptic sites e.g. terminators and RNase sites, GC content and repetitive elements are not considered (212). Variation in local sequence motifs can have a deleterious effect on heterologous protein expression. Sites such as RNase recognition, DNA recombination and transcriptional and translational sites can reduce expression, for which elimination is constrained by the selection of 'optimised' codon selection. Furthermore, sequences such as terminators and repetitive elements can also compromise stability of translation and cause a decrease in expression, in addition to misincorporation of amino acids (213). In addition to specific site incorporation, the mRNA structure can also influence the efficiency of protein expression. Both Kudla *et al.* (86) and Allert *et al.* (209) demonstrated the influence of 5' mRNA structure on expression. Secondary structure arising from the nucleotide sequence can occlude the RBS and translational start codon, which was demonstrated to be a defining factor correlating to expression (86,209). Furthermore, both studies showed no significant correlation with CAI. Nevertheless, codon usage has shown significance, corresponding to both global and local variations. It is clear that clusters of rare codons can effect translation

(214) through stalling (215), frameshifting (216), decreased translational velocity (78) and amino acid misincorporation (217). Furthermore, it has been demonstrated that inclusion of rare codons early in the mRNA sequence, suggested to be between 15 – 25 codons, can also have deleterious effects on expression (86,209,218,219). This role is still currently not fully understood, as specific codons have been demonstrated to influence the folding rate of the protein by modulating the translation rate at critical points (207,220-222). Allert *et al.* (209) observed, in addition to the mRNA 5' structure, a correlation between expression and GC content, a result that was not shared by Kudla *et al.* (86). These studies demonstrated the success of algorithms to predict and increase protein expression. Welch *et al.* (178) also constructed an algorithm which successfully increased expression by two orders of magnitude, and like Kudla, Allert and co-workers observed no correlation with CAI. They did not, however, observe a correlation with mRNA 5'structure, yet it was postulated that this was due to the lack of tertiary structure exhibited by the proteins tested (178). In contrast to previous work, Welch *et al.* (178) demonstrated that protein expression was increased through the selection of codons exhibiting tRNAs that remained highly charged during starvation. It was hypothesised that this was due to the physiological shock inflicted upon the host cells due to the synthesis of large amounts of a single protein.

This specific algorithm produced by Welch *et al.* (178) is the basis of the codon optimisation strategy employed in this study. This initial

experiment-based algorithm, tested in *E. coli*, has since been employed to improve gene design algorithms for several species including *S. cerevisiae*, *Kluyveromyces lactis*, *Agaricus bisporus*, and several *Clostridium* species (212). Multiple combinations of the *pheB* gene, from *G. stearothermophilus*, using different 'synonymous' codons were synthesised and tested. The results identified seven variants with a significant statistically increased activity compared to the WT gene, with the best variant exhibiting a 2.6-fold increase, demonstrating the achievement of a more optimised codon usage. These results therefore confirm what was identified previously from the manual analysis, that there is indeed a codon bias within *P. thermoglucosidasius*. Although this initial data is promising, these highly expressed *pheB* gene variants should now undergo further optimisation in an iterative process to fine tune the algorithm, in order to confidently produce a codon usage system that is fully optimised for *P. thermoglucosidasius*.

Development of an Inducible Expression System

Inducible expression systems are important to genetic engineering, as they provide controllable timing of expression. Furthermore, inducible expression systems allow regulated gene expression as induction is typically controlled by the concentration of the inducer. This study details the first successful dosage-dependent inducible system that is regulated at the mRNA-level in *P. thermoglucosidasius*, validating three riboswitches, E, F and G. These riboswitches have previously been shown to sequester

protein expression and allow induction upon binding of the ligand theophylline to its aptamer domain (162) (I. Canadas Blasto, unpublished). Significantly, the three riboswitches tested in this study demonstrated overexpression compared to the same constitutive promoters. This overexpression is consistent with what has been previously observed, with Riboswitch E showing overexpression in *B. subtilis* (162).

Thus far, only ligand-based inducible promoters that act at the transcriptional level have been identified and successfully utilised within *P. thermoglucosidasius*. The β-glucanase promoter was identified from within *P. thermoglucosidasius* and showed induction on addition of cellobiose, in addition to a weaker induction using glucose and xylose respectively (56). A xylose-inducible promoter from the xylose isomerase gene (*xylA*), native to *P. thermoglucosidasius*, has also been demonstrated as a successful inducible promoter. However, a second carbon source of either glucose or glycerol was required in addition to the inducer xylose (183). In addition, a sucrose inducible promoter and a maltose inducible promoter have been expressed in *G. stearothermophilus* and *G. kaustophilus,* respectively (166,223). The riboswitches validated from this study offer a key advantage over the currently available inducible promoters for use in *P. thermoglucosidasius*, which are both native promoters involved in carbon catalysis. Unlike these riboswitches, which utilise theophylline as the inducer, both previous systems are limited by

carbon catabolite repression and other regulatory mechanisms, e.g., repressors like XylR.

The data obtained through this study does appear to show that the riboswitch systems are leaky, with expression observed in the uninduced samples indicating that the RBS is not completely sequestered. This result is corroborated by similar observations in both *B. subtilis* and *Streptomyces coelicolor* when under the control of riboswitch E (162,224). This is where the previous systems may be at an advantage, as Pogrebnyakov *et al.* (183) were able to successfully repurpose the XylR repressor, in order to both reduce the leakiness of the system and produce overexpression during times of high inducer concentration. Furthermore, the $P_{\beta glu}$ inducible promoter, validated by Bartosiak-Jentys *et al.* (56), showed almost no activity when uninduced.

Prior to demonstrating inducible expression via the riboswitches, it was first important to determine the tolerance of *P. thermoglucosidasius* to theophylline. It is known that theophylline can affect the concentration of both cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), as it is a weak non-selective inhibitor of phosphodiesterase (Pde) enzymes (PDE) that catalyse their breakdown. This molecule therefore has the potential to effect viability of the cells (225). Raj and Dhala (226) demonstrated that at high concentrations xanthines are bactericidal and bacteriostatic. This was shown for multiple prokaryotes, with the most susceptible organism being *B. subtilis.*

However, the study by Raj and Dhala (226) also elucidated that susceptibility to xanthines varied greatly, even in species that were related. Therefore, it was crucial to first identify whether *P. thermoglucosidasius* was susceptible to theophylline over a range of concentrations necessary for induction. The data attained illustrated that *P. thermoglucosidasius* is indeed susceptible to theophylline, as illustrated by a clear reduction in cell density, following a linear trend, observed as the concentration is increased. However, although reduction in cell viability was observed, even at the highest theophylline concentration a cell density of 50 % was still achieved after growth for 8 hours compared to the WT. This demonstrated that *P. thermoglucosidasius* was able to tolerate theophylline at the relevant concentration needed to successfully use the chosen riboswitches for the creation of an inducible system.

Following determination that *P. thermoglucosidasius* was able to withstand the relevant theophylline concentrations, an investigation was conducted to determine whether the riboswitches were dose dependent. A range of theophylline concentrations, 0 – 10 mM, was tested to observe the effect on induction of Riboswitch E, again by monitoring the production of the sGFP reporter protein. A dose-dependent response was observed, showing an increase in fluorescence up to a maximum of 8 mM. This dose-dependent response was also observed for riboswitch E in *B. subtilis* (162). This increase in expression was not observed at 10 mM, suggesting that at this concentration the toxicity effect of theophylline is

starting to become more prevalent. This dosage dependent response is beneficial as it will allow a tailored response to gene expression.

One limitation of this study is that the RBS sequences are different between the riboswitch induced sample and the constitutive promoter control. This means that the intensity levels produced from each are not completely comparable. However, it does indicate that the SD sequence within the riboswitches is stronger than the synthetic RBS sequence used in the riboswitch-less expression plasmids. It would therefore be interesting to observe the effect of the riboswitch SD sequence on a constitutive promoter, in order to determine whether this is stronger than any of the currently validated RBS sequences. Another consideration for future exploration is the possibility that *P. thermoglucosidasius* may be able to metabolise theophylline. It has been shown, for example, that theophylline is metabolised by *E. coli* to produce 1- and 3-methylxanthine (227). If this also occurred within *P. thermoglucosidasius*, the expression levels would therefore decrease over time as the inducer is metabolised. Previous observations in-house have shown that often desirable metabolites, such as 3-Hydroxypropionic acid (3-HP), have been metabolised by *P. thermoglucosidasius* (Bashir, unpublished) so it is possible that this could also be the case for theophylline. Therefore, in the future, the ability of *P. thermoglucosidasius* to utilise theophylline as a carbon source should also be investigated.

3.4 Conclusions and Future Work

At the onset of this project, a very limited selection of genetic tools was available to facilitate the genetic engineering of the thermophile *P. thermoglucosidasius*, identified as a potential source for the production of renewable biofuels. The aim of this work was to identify and validate new genetic tools and to ultimately develop a whole range of tools with a variety of uses and strengths, which the data described in this chapter successfully demonstrates. Specifically, this work details the identification and validation of two stable fluorescent proteins, sGFP and eCGP123, for use in *P. thermoglucosidasius* and the utilisation of these proteins to validate and quantify the relative strengths of various promoters, RBS, terminators and an inducible expression system. Furthermore, initial steps have been taken towards identifying and developing a codon optimised usage system in *P. thermoglucosidasius*, achieved through both manual analysis of codons and using an algorithm produced by ATUM.

Future work should focus on the continued optimisation of these libraries, through an exploration of factors such as the effect of growth conditions. It has been shown that gene expression is intrinsically linked to cellular growth and so the strength of promoters expressed at different times during the grow cycle should also be considered (228,229). Successful identification of the promoter consensus sequence would be another useful target for future work and could possibly be achieved using a method called Rapid Amplification of cDNA (complimentary DNA) ends

(RACE). This technique enables the acquisition of a fully sequenced specific mRNA transcript, allowing the identification of the 5' UTR. Further expansion of the RBS library could be achieved by testing RBS sequences identified from within *P. thermoglucosidasius*. This would allow the spacer sequences around the consensus sequence to be varied, in order to produce an optimised-semi synthetic library with a range of strengths. Furthermore, the use of an N-terminal tag for protein extraction of the fluorescent proteins YtvA LOV and GAF could allow further indication as to the expression, as a C-terminal tag may be subjected to any cleavage events. Further optimisation of the *pheB* gene variants is necessary to produce a complete usage system optimised for protein expression. The data obtained here will be used to guide the design and synthesis of two further sets of variant genes, which can again be tested in terms of expression levels in *P. thermoglucosidasius* (178). Finally, further testing of the riboswitches is necessary to compare the overexpression profile against a constitutive promoter expressing the same SD sequence, in addition to identifying whether theophylline is utilised as a carbon source.

3.5 Key Outcomes

- Identification, testing and expression of two thermostable fluorescent reporter genes in *P. thermoglucosidasius*.

- Development of a promoter library, with relative strengths for each promoter identified in both *E. coli* and *P. thermoglucosidasius*.

- Development of a RBS library, with relative strengths for each RBS identified in both *E. coli* and *P. thermoglucosidasius*.

- Development of the first terminator library produced for *P. thermoglucosidasius*, with the relative strengths of each terminator identified.

- Development of a riboswitch mediated inducible expression system, with relative strengths of each riboswitch identified for three different promoters in *P. thermoglucosidasius* and elucidation of the tolerance of *P. thermoglucosidasius* to theophylline.

- Identification of a codon bias present in *P. thermoglucosidasius* and initial steps taken towards producing a codon optimised usage system

Chapter 4

Development of CRISPR/Cas-based Knock-In and Knock-Out Technology

4.1 Introduction

4.1.1 Manipulation of Genes and Genomes

The ability to manipulate an organism, by KO and KI of genes of interest, is one of the key foundations of genetic engineering. As discussed in sections 1.4.5 and 1.4.6 this is generally achieved through replacement of a wildtype allele with the requisite a mutant allele through allelic exchange in a process reliant on homologous recombination (HR) between homology arms that flank the alleles being exchanged. In standard allelic exchange (AE) approaches the flanking homology arms are of an equal size (generally between 500 to 1000 bp). ACE, in contrast, uses asymmetric homology arms comprising a long (ca. 1200 bp) and short (ca. 300 bp) homology arm which dictates the order in which homologous recombination events are selected. Single crossover integration of the ACE plasmid preferentially occurs as a result of homologous recombination between the longer, 1200 bp homology arms. Whilst the subsequent excision of the plasmid via the same longer homology arms is preferred, excision via the shorter 300 bp homology arms can be selected due to the creation of a new allele with a selectable phenotype. Thus, in the one configuration the allele formed is a defective *pyrE* allele, which leads to resistance to fluoroorotic acid (FOA), a phenotype that is easily selected through incorporation of FOA into the agar medium. In the opposite configuration, the mutant *pyrE* allele is converted back to wildtype, a phenotype that is easily selected through selection for

prototrophy, ie., plating on agar medium lacking uracil where the parent is unable to grow due to the absence of uracil, but the desired double crossover excision variant can. Both standard AE and ACE have been successfully implemented in *P. thermoglucosidasius* (24,118). However, both techniques have limitations with regard to efficiency.

4.1.2 CRISPR/Cas9 System

The discovery of CRISPR systems and their subsequent exploitation in genome editing represent a 21st century breakthrough. Clustered regularly interspaced short palindromic repeats, more commonly referred to as CRISPR, and CRISPR-associated proteins (Cas) have been reviewed in detail in section 1.5. First observed by Ishino *et al.* (120) in the genome of *E. coli* and identified as a bacterial adaptive immune system, this process has been harnessed to selectively target locations in the genome for editing purposes. By transforming the DNA encoding the Cas9 nuclease complex into the bacteria, in addition to a synthetic guide RNA (sgRNA), the genome of the organism is cleaved at a specific, defined location. This ability to accurately target precise locations within the genome allows the efficient addition or removal of specific genes through homologous recombination, thereby overcoming the limitations of the currently available techniques.

The CRISPR/Cas9 system most commonly used in bacteria was initially identified in the mesophile *S. pyogenes.* It is a Type II *cas9* gene (spCas9), shown to cause cell death of *Streptococcus pneumoniae* by chromosomal targeting (230). Since then, this spCas9 has been successfully applied to facilitate genome editing in numerous different bacteria including *S. pneumoniae*, *E. coli* (231), *Clostridium cellulolyticum* (232), *B. subtilis* (233), *C. acetobutylicum* (234,235) and *Clostridium ljungdahlii* (236). It is clear that this technology has the potential to provide huge advancements in the field of genetic engineering and therefore should be exploited for use in *P. thermoglucosidasius*. Successful utilisation of spCas9 for genome editing has been shown in the moderate thermophile *Bacillus smithii* (237), which grows between 25 – 65 °C. However, the activity of spCas9, to enable lethal double-stranded DNA breaks, was only possible following growth at 37 °C. In confirmation of this, work by Mougiakos *et al.* (237) demonstrated that spCas9 is inactive at temperatures above 42 °C. This therefore suggests that spCas9 would be unsuitable for use in *P. thermoglucosidasius*.

4.1.3 Streptococcus thermophilus Cas9

The organism *Streptococcus thermophilus* is a non-pathogenic, Gram-positive thermophile, with an optimum growth temperature of 45 °C (238,239). Its widespread use in food manufacturing means that it is widely studied, with an extensively analysed genome. CRISPR/Cas

systems are known to be prevalent in streptococci species (*S. agalactiae*, *S. mutans*) (240,241), with the first programmable use of CRISPR/Cas9 for genome editing originating from *S. pyogenes* (231).

S. thermophilus encodes up to four CRISPR/Cas systems, two of which are categorised as Type II CRISPR/Cas9 systems, specifically stCas9-1 and stCas9-3 (123,242,243). These CRISPR/Cas9 systems have been shown to provide immunity against viruses and plasmids in *E. coli*, through the nuclease activity of Cas9. In addition, the PAMs for stCas9-1 and stCas9-3 have also been investigated and determined to be 5'-NNAGAA(W)-3' and 5'-NGGNG-3', respectively (where N stands for any nucleotide and W stands for A or T) (244-247). Similar to *cas9* from *S. pyogenes*, genome editing has also been successfully pursued and demonstrated with *S. thermophilus cas9* (246-248).

4.1.4 Plasmids

Plasmids are independently replicating, autonomous, extrachromosomal DNA that are widespread in prokaryotes and archaea, but have also been identified in eukaryotic organisms (249). Aside from their practical applications in genetic engineering (Section 1.4.1), plasmids play a central role in horizontal gene transfer, together with integrative conjugative elements and phages. They vary in size, from < 2 kb to as large as > 400 kb (250). At the upper end of the scale, the distinction between a

megaplasmid and a small chromosome becomes borderline and in some cases plasmids can constitute more than 25 % of the total genetic material within the cell (251,252). These megaplasmids often contain genes involved in antibiotic resistance, detoxification of toxic organic compounds, specific ecological interactions and virulence (250). Plasmids have a high rate of interchange between prokaryotes, flowing back and forwards with ease, and can be thought of as a shared genetic pool of extrachromosomal DNA. This allows rapid evolution and adaptation, via homologous recombination and integration, respectively, which facilitates a fast response to stressful conditions and new environments (250,253). It is also believed that plasmids drive microbial cooperation through sharing of genetic material (254).

Maintaining such a large extrachromosomal element is not without its costs, and may impose a measurable burden on the cell (250,255). For example, in multiple *Pseudomonas* species, the acquisition of the vector pMPPla107 imposed a fitness cost resulting in slower growth, both in isolation and competitive growth assays (256). Furthermore, the curing of the megaplasmid pTT27 from *T. thermophilus* resulted in an enhanced transformation efficiency (257). However, this effect appears to be species specific as conversely, the curing of the megaplasmid from *Sinorhizobium meliloti* showed no effect on growth compared to WT (258).

The deliberate removal of a megaplasmid from a cell population can be accomplished in a number of different ways. One approach is to integrate

into the plasmid a negative selection marker, such as *sacB* (258). This marker confers sensitivity to sucrose (259). Growth of the modified cell line in the presence of this sugar therefore leads to the selection of cells that have lost the megaplasmid. Integration of the negative selection marker into the plasmid is accomplished by allelic exchange using a suicide plasmid that additionally carries a positive selection marker, such as a tetracycline resistance gene (256). Alternatively, transposons have also been used to integrate the *sacB* gene, in order to cure a megaplasmid (260). Incompatibility curing can also be employed to remove a plasmid. Here, a second plasmid which is reliant on the same replication region as the plasmid to be cured. Plasmids with an identical replicon cannot exist stably within the cell and by imposing selection for the newly introduced plasmid alone; the original plasmid is cured (257,261). In addition, plasmids can also be cured through curing agents, elevated growth temperature or thymine starvation (262). Megaplasmids occur in many thermophilic bacteria; however, curing of these plasmids is relatively unexplored. Currently, two megaplasmids have been identified in *P. thermoglucosidasius* (263). These are pNCI001 (CP016623), and pNCI002 (CP016624) which are 83.9 and 47.9 kb in size, respectively). However, the role of these megaplasmids is still largely unknown, with many of the genes they carry encoding hypothetical proteins of unknown function.

4.1.5 Chapter Aims

The ability to selectively target areas of a genome with relative ease is what makes the CRISPR/Cas9 technology so attractive as a genetic engineering tool. At the start of this project, no *cas9* had been shown to be active at the high temperatures required for *P. thermoglucosidasius* growth. This work described in this chapter aimed to identify and validate a thermostable Cas9 protein for use in *P. thermoglucosidasius*. Following identification of a suitable candidate, the ability of this protein to cause a double stranded cleavage was explored, before repurposing the system to facilitate KO and KI of specific genes by allelic exchange. If successful, it was envisaged that the system could be used elucidate the function of the two megaplasmids pNCI001 and pNCI002 and to explore their potential exploitation as a stable location for deliberately introduced traits that would exists at an elevated copy number compared to the chromosome.

4.2 Results

4.2.1 Development of a Thermostable CRISPR/Cas9 System in *P. thermoglucosidasius*

4.2.1.1 Construction of a Single Plasmid CRISPR/Cas9 System

It was decided to explore whether the two Type II CRISPR/Cas9 systems found within *S. thermophilus* could be used for genetic engineering of *P. thermoglucosidasius*, based on their previous success as genetic engineering tools in a multitude of species. To achieve this, a series of vectors were designed and constructed to initially demonstrate that the *cas9* genes were active in *P. thermoglucosidasius* at its optimum growth temperature.

The shuttle vector pMTL61110 was used as the backbone, which contained the ColE1 origin of replication from *E. coli* and the temperature sensitive RepB pUB110 replicon from, isolated from *S. aureus* and used for replication in *P. thermoglucosidasius*. All shuttle vectors used a kanamycin resistance (*km*^{R}) gene for selection. The application specific module, which was located between the SbfI and AscI restriction sites, was modified to accept the introduction of the *cas9* gene and the sgRNA. Primers were designed to amplify the *cas9* genes (*stcas9-1* and *stcas9-3*) from the genome of *S. thermophilus* LMD-9 (GenBank: CP000419.1). The two

strongest promoters from the previously developed promoter library (section 3.2.3.2), P_{ldh} and P_{gapd}, were used to drive the *cas9* and sgRNA expression, respectively. Splicing by overlap extension (SOE) PCR was used to fuse P_{ldh} and *cas9* together and the construct was then cloned between SbfI and XbaI. Ack was selected as the initial KO target. As two different Cas systems were to be employed, separate sgRNAs specific for the PAM site of each Cas9 were necessary.

The Guide Design software of Benchling (www.benchling.com) was employed to search for 5'-N_{20}NNAGAAW-3' and 5'-N_{20}NGGNG-3' sequences, for *stcas9-1* and *stcas9-3*, respectively, within the non-essential acetate kinase (*ack*) gene. This was needed in order to target the corresponding PAM recognition sequence, which is immediately downstream of the 20 nt sgRNA sequence. The 20 nt sgRNA sequence was then synthesised attached to the sgRNA handle, a specific sequence corresponding to each *cas9* to allow association. The number of viable sequences available for each of *stcas9-1* and *stcas9-3* varied widely with only 4 PAM sequences for *stcas9-1* and 35 for *stcas9-3* (Table 4.1). To test the expression and targeting of each *cas9*, the top sgRNA sequence was selected, with regard to the On-target score of the sequence. SOE PCR was used to attach P_{gapd} to the sgRNA and this was then cloned between the restriction sites XbaI and SalI, to complete the vectors pCas91_Ack and pCas93_Ack (Figure 4.1).

Table 4.1 – Spacer and PAM analysis from within the *P. thermoglucosidasius ack* gene, for use with both *stcas9-1* and *stcas9-3*. All of the PAM sequences are displayed for *stcas9-1 ack*, whereas only the 10 highest scoring PAMs are displayed for *stcas9-3 ack,* based on the On-target score. On- and Off-target scores refer to guidelines set by Doench *et al.* (2016). The two sgRNA sequences used in the vectors pCas91_Ack and pCas93_Ack are highlighted.

Position	Strand	Sequence	PAM	On-target Score	Off-target Score
			stCas9-1		
1333	1	TTTACGATCACCGTCAATGG	CGAGAAA	42.80	100
1810	-1	TGGCGCCATTGCCAAGGTGG	CAAGAAA	19.05	100
1498	-1	ATACTTCATCCGTAATGAGT	ACAGAAT	17.47	100
1990	1	GAAGTAGTGGAAGTATTAAA	TAAGAAA	17.03	100

Position	Strand	Sequence	PAM	On-target Score	Off-target Score
			stCas9-3		
1462	-1	TTTCGCCACCATGCACAACG	CGGTG	78.17	100
1764	-1	CGGACGGCCGAGCAATTCAG	CGGCG	73.51	99.65
1841	1	CCAGCATCGCTGCAGTAGAA	GGGGG	53.24	100
1816	1	ATTTCTTGCCACCTTGGCAA	TGGCG	45.95	100
2052	1	CTTGCGCGATTTGGAAAAAG	CGGCG	45.56	100
1317	-1	ATTGACGGTGATCGTAAAAA	TGGCG	31.59	99.48
1614	1	GCTGCCGAATGTGCCTGCCG	TGGCG	30.81	100
1377	1	CATTCCGAACCATGCAGTAG	CGGTG	28.49	100
1801	-1	TGCCAAGGTGGCAAGAAATG	AGGCG	22.41	100
1553	-1	ATGTTTGCTGGATTATGAAG	GGGAG	19.68	100

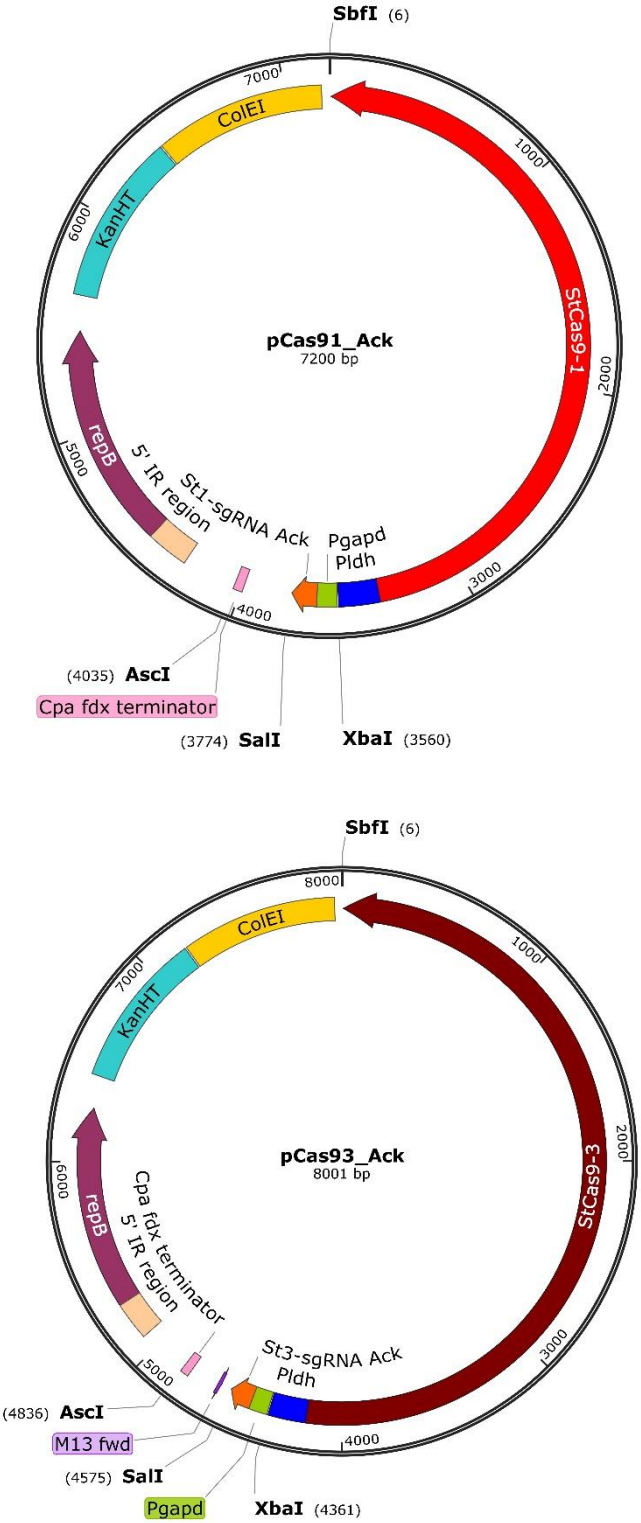


Figure 4.1 – A schematic representation of the vectors (A) pCas91_Ack and (B) pCas93_Ack, illustrating the gene locations and the restriction enzyme sites used in the cloning of each vector.

Two further variants of the initial vectors, pΔCas9 and pCas9only, were created for each of *stcas9-1* and *stcas9-3*. The two pΔCas9 plasmids contained a frame shifted variant of each *cas9* gene. They were produced by cutting the *stcas9-1* and *stcas9-3* genes with PacI and Eco81I, respectively, blunt-ending the sticky-ends produced through treatment with T4 DNA polymerase, and then subjecting the two linear fragments to self-ligation. As a further control, plasmids carrying functional copies of both Cas9 genes were constructed, but which lacked a sgRNA.

4.2.1.2 In vivo Expression of stcas9-1 and stcas9-3

To evaluate whether *stcas9-1* and *stcas9-3* were active, pCas91/3_Ack, pΔCas9, pCas9only and pMTL61110, an empty plasmid control, were transformed separately into *P. thermoglucosidasius* competent cells. The transformations were plated onto TSA plates containing kanamycin (km) and incubated O/N at 52 °C, after which the plates were visually examined for growth. On plates containing pCas91/3_Ack, there was no observed growth of any colonies. Conversely, successful growth of cells was observed on all plates containing pΔCas9, pCas9only and pMTL61110 (Figure 4.2).

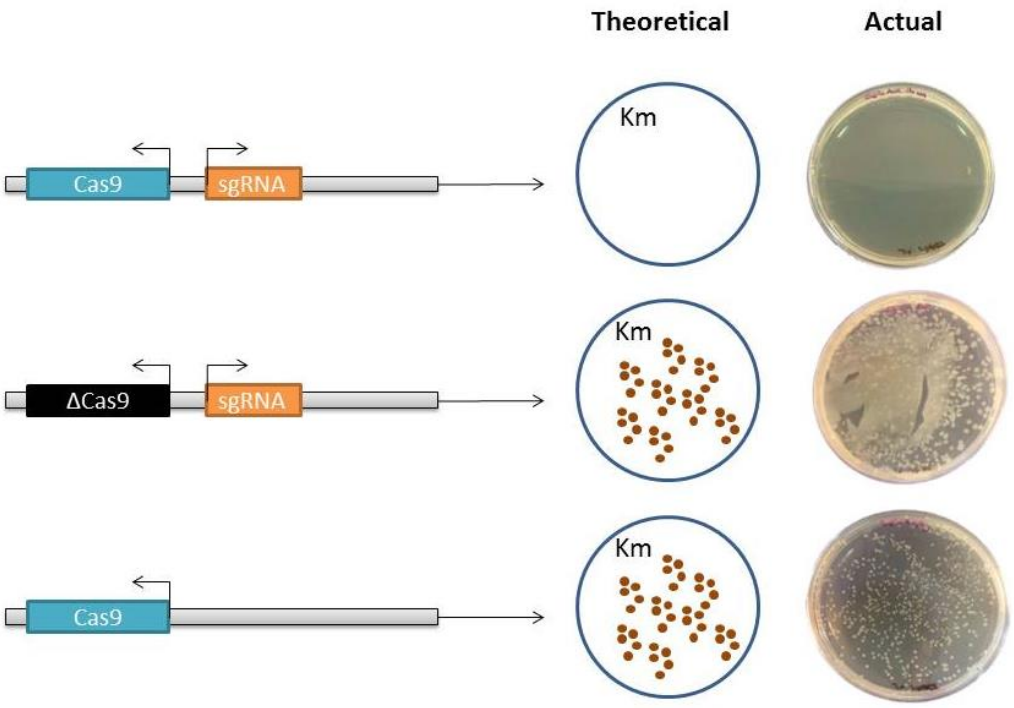


Figure 4.2 – A schematic representation of a successful WT *cas9* compared to a *cas9* without functional activity (Δ*cas9*) and without a sgRNA. The CAS9 is guided using synthetic guide RNA (sgRNA) and cultured on kanamycin containing plates. Theoretical images illustrate how the expected result should appear. Actual images show the results observed after transformation.

4.2.1.3 Cas9-Mediated KO of Acetate Kinase through Homologous Recombination

After successfully establishing that both *stcas9* genes were active in *P. thermoglucosidasius*, the next step was to develop an editing system using these genes, through AE. The plasmids pCas91_Ack_hr and pCas93_Ack_hr were produced (Figure 4.3A), comprising the relevant *stcas9* gene, *ack* targeting sgRNA (relative to the *cas9* gene) and a 'editing template' comprising fused left and right homology arms (HAs) of 500 bp in size that

flanked the *ack* gene. The editing template was produced using SOE PCR. The new plasmids were transformed into *P. thermoglucosidasius* competent cells alongside the previously created plasmids, pΔCas9, pCas9only and pMTL61110, which acted as controls. After an O/N incubation, the number of colonies obtained was similar to that of the controls. This was unexpected, as WT cells expressing both *cas9* and gRNA targeting *ack* should be 'killed', due to cleavage of their chromosomes. The only cells that should survive are those in which the *ack* gene carrying the sgRNA target sequence has been replaced by AE with the mutant allele carried by the editing template. The number of survivors should be low. Indeed, screening of a random selection of the colonies by PCR revealed they were all WT.

One possible explanation for failure of the systems was that the strong promoter responsible for production of the sgRNA, P_{gapd}, was reading through the sgRNA terminator. To explore this possibility, an additional terminator was sought which could be positioned downstream of the sgRNA. Originally, when cloning the editing template into the shuttle vector, the downstream terminator between the application module and the Gram-positive replicon had been removed. Therefore, a terminator from the previously established terminator library (section 3.2.5), EcoT1 was included to eliminate any potential read-through, producing the vectors pCas91_Ack_hrT and pCas93_Ack_hrT (Figure 4.3B).

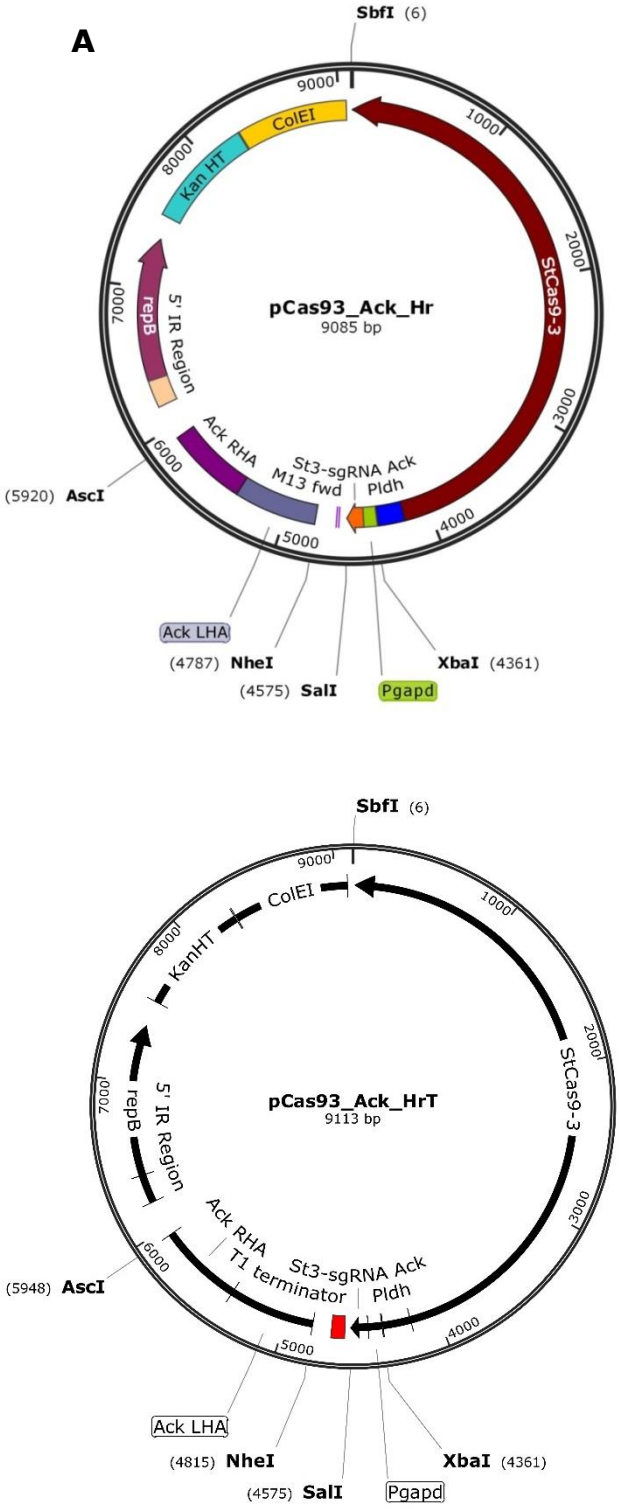


Figure 4.3 – A schematic representation of the vectors (A) pCas93_Ack_hr, illustrating the inclusion of the homology arms homologous to acetate kinase (*ack*) into the vector pCas93_Ack and (B) pCas93_Ack_hrT, illustrating the addition of the T1 terminator to pCas93_Ack_hr.

The vectors pCas91_Ack_hrT and pCas93_Ack_hrT were then transformed into *P. thermoglucosidasius* competent cells. After O/N incubation colonies were evident but at a significantly reduced level compared to the cells that had been transformed with the control plasmids, pΔCas9 and pCas9only. As before, colonies were picked, cultured in 2SPYNG km and incubated O/N at 52 °C, before a colony PCR was conducted to determine whether the KO of the *ack* gene had been successful. It was observed that the KO had successfully been created; however, the colonies were not pure, being composed of a mixture of mutant and WT. To rectify this the colonies were further cultured in 2SPYNG km O/N at 52 °C for three passages, sequentially transferring the cells to fresh media each time to achieve clean Δ*ack* KO mutants (Figure 4.4).

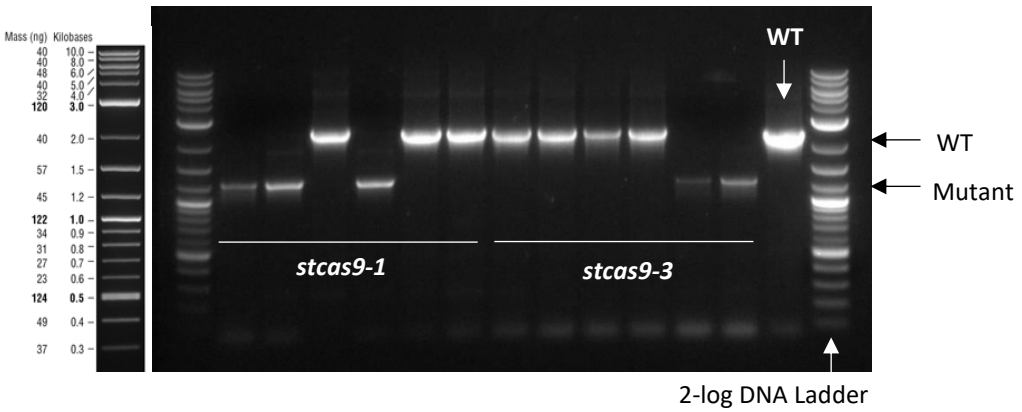


Figure 4.4 – A colony PCR, utilising Ack_F and Ack_R primers, was performed to assess whether the acetate kinase (*ack*) gene had been successfully knocked out of the *P. thermoglucosidasius* genome using *stcas9-1* and *stcas9-3*. A PCR of WT *ack* was performed as a control (~2.4 kB). Colonies that have the Δ*ack* would yield a band at ~1.2 kB. A 2-log DNA ladder was used to determine the size of each band.

4.2.2 Elimination of Megaplasmids and Expansion of Tools

Having established functional CRISPR/Cas9 systems, the possibility of using them to manipulate the two *P. thermoglucosidasius* megaplasmids pNCI001 and pNCI002 was explored. Rather than use both systems, only that based on *stcas9-3* gene was used. This was because its PAM sequence is more ubiquitous (NGGNG) and therefore offers more potential targets, as demonstrated by the significantly larger number of available hits produced when designing the sgRNAs for the *ack* KO.

4.2.2.1 Knock-out of Megaplasmids

To eliminate pNCI001 and pNCI002 from the cell, the strategy initially adopted was to target the gene encoding the replication protein, RepB and RepA, respectively. The requisite KO plasmids were designed as previously described (Section 4.2.1.3), using 500 bp HA from both upstream and downstream of the rep genes. The editing template was produced using SOE PCR and was cloned between the restriction enzyme sites NotI and AscI. Following transformation of the complete plasmids, pCas93_SMPR_hrT and pCas93_BMPR_hrT, the desired mutant was only obtained in the case of the pNCI002, with a complete loss of the plasmid observed by PCR (Figure 4.5).

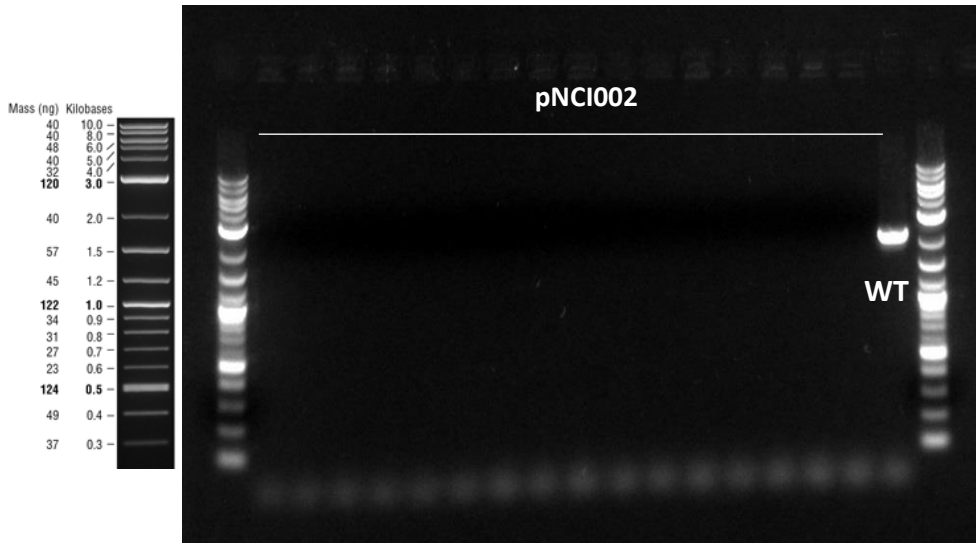


Figure 4.5 – A colony PCR, utilising SMPR_CP_F and SMPR_CP_R primers, was performed to assess whether the pNCI002 replication gene (*repA*) had been knocked out. A PCR of the WT *repA* was performed as a control (~2.2 kB). Colonies containing Δ*repA* would yield a band at ~1.2 kB. However, no PCR product was observed when analysed on a 1% agarose gel, indicating the complete loss of the small megaplasmid (pNCI002). A 2-log DNA ladder was used to determine the size of each band.

In the experiments targeting pNCI001, not a single colony was obtained using all three sgRNA regions designed to target the replication gene. On further inspection of the pNCI001 sequence, a putative toxin-antitoxin system was identified. As antitoxins have a significantly shorter half-life than toxins, removal of the plasmid would result in cell death. Accordingly, a new CRISPR/Cas9 KO vector, pCas93_Tox_hrT (see section 4.2.1.3) was produced that would target the toxin gene. Following its transformation into *P. thermoglucosidasius*, PCR screening of the colonies obtained revealed that rather than generating a pNCI001 derivative lacking the toxin gene, the entire plasmid had been cured, as evidenced by the absence of the replication gene *repB* (Figure 4.6A). As this result was

unexpected, a secondary confirmation PCR was performed to ensure that the isolate was *P. thermoglucosidasius* and not a contaminant. This was achieved by performing a colony PCR that targeted the native *pyrE* gene (Figure 4.6B). Following this, the vector, pCas93_Tox_hrT, was cured from the strain, by raising the temperature to 60 °C.

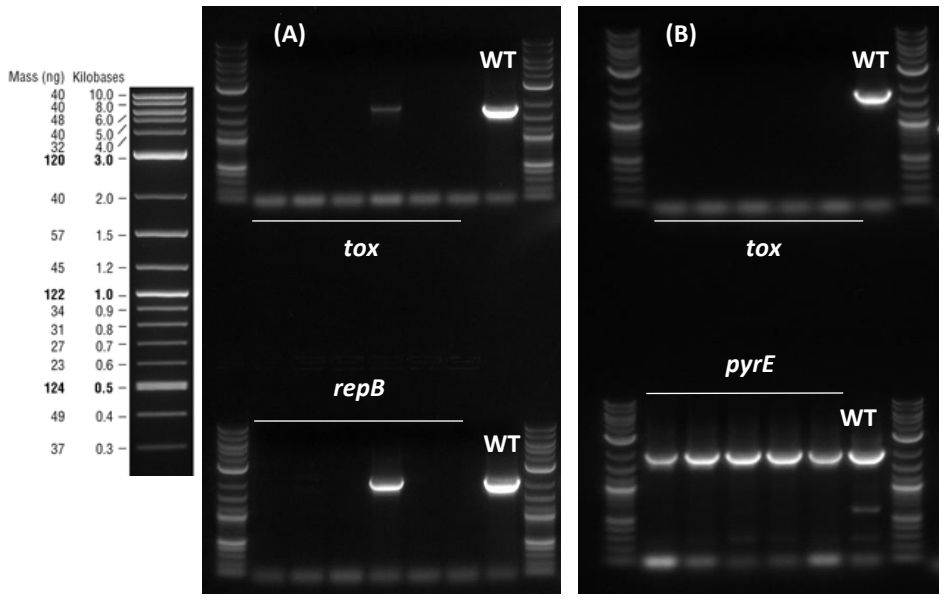


Figure 4.6 – Colony PCRs were performed to show the loss of pNCI001 from *P. thermoglucosidasius*. (A) Two separate PCR reactions of the same colony. PCR of the *tox* gene utilised the primers Tox_CP_F and Tox_CP_R and that of the *repB* replication gene the primers BMPR_CP_F and BMPR_CP_R. A PCR of the WT *tox* gene and the *repB* was performed as a control (~1.8 kb and ~2.3 kb, respectively). Colonies containing Δ*tox* would yield a band at ~1.3 kb. However, no PCR product was observed when analysed on a 1% agarose gel indicating the complete loss of the pNCI001. The unsuccessful PCR of the *repB* further confirmed this finding. (B) A PCR of the Δ*repB* mutants to confirm the strain as *P. thermoglucosidasius*. PCR of *pyrE* was performed, utilising the primers PyrE_F and PyrE_R, alongside a PCR of *tox*, as performed previously. PCR of WT *pyrE* was implemented as a control (~1.7 kb). A 2-log DNA ladder was used to determine the size of each band.

Following the production of both single plasmid deficient strains, it was decided to also construct a double deficient strain. The plasmid pCas93_SMPR_hrT was transformed into pNCI001-free *P. thermoglucosidasius* competent cells. Using the method described previously in this section, the pNCI002 was deleted, demonstrating a successful double deletion of both mega-plasmids in *P. thermoglucosidasius* (Figure 4.7).

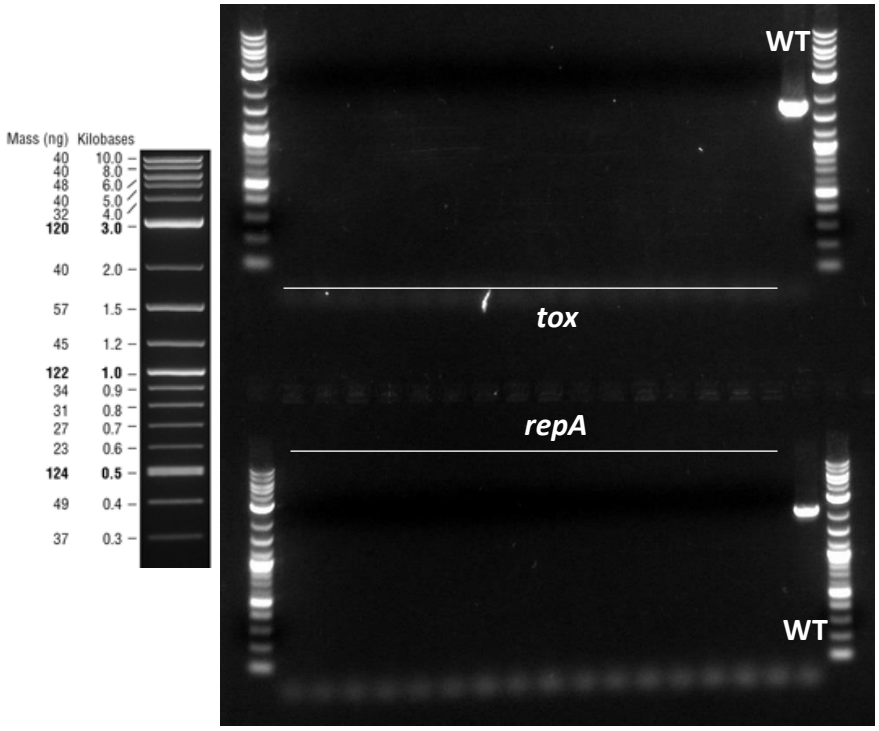


Figure 4.7 – A colony PCR showing the generation of a double megaplasmid deficient strain. The PCR controls and the primers for each strain were as described previously in figures 4.7 and 4.8 respectively.

4.2.2.2 Characterisation of Megaplasmid Deficient Strains

After producing both single (pNCI001-free and pNCI002-free) and double (pNCI001/2-free) megaplasmid deficient strains in *P. thermoglucosidasius*, it was of interest to assess any differences to the WT in terms of growth and metabolism. All three strains were grown aerobically in baffled conical flasks in the defined media, modified ASYE (7) supplemented with 1% glycerol (50 mL) at 60 °C over 36 hours (Figure 4.8). It was observed that both the single (Figure 4.8A) and double (Figure 4.8B) megaplasmid deficient strains showed no signs of a reduction in fitness in comparison to the WT, over a growth period of 36 hours.

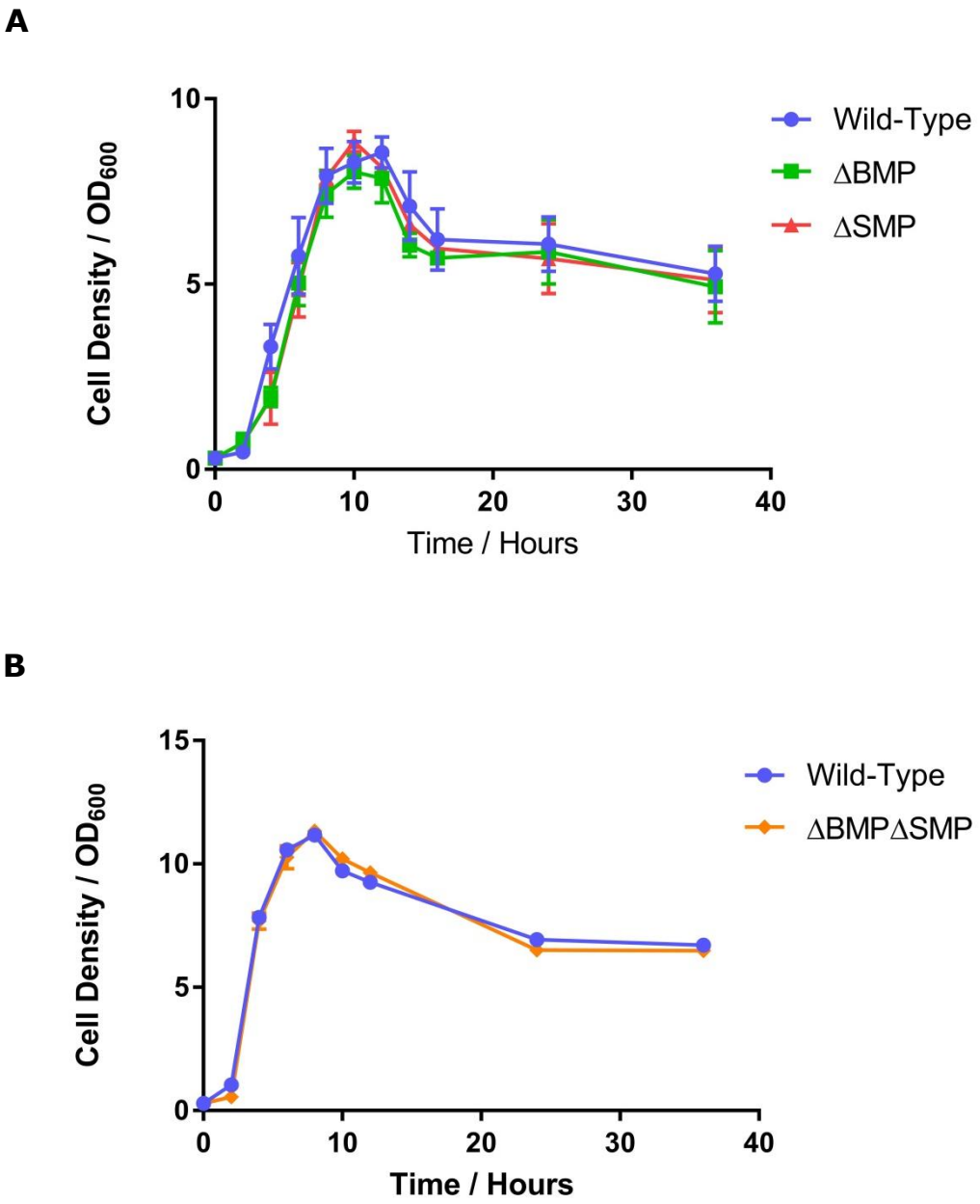


Figure 4.8 – Growth curves of the megaplasmid deficient strains of *P. thermoglucosidasius*, monitoring the effect on growth over a 36-hour time period resulting from the (A) single and (B) double plasmid loss, compared to wild-type. The error bars on the graph denote the standard error of the mean (n = 3).

In order to assess whether the loss of these plasmids affected the production of metabolites, each single plasmid deficient strain was grown and samples of the supernatant were collected over a period of 36 hours. Several metabolites of interest were selected and their levels measured using high-performance liquid chromatography ultraviolet (HPLC-UV). These strains were grown using the defined media, modified ASYE (7) supplemented with 1% glycerol under aerobic conditions in baffled conical flasks. The supernatant was collected and mixed with HPLC diluent in a 1:1 ratio, the standard in-house method for HPLC analysis. The concentration of each metabolite was calculated using calibration curves produced from analytical standards. The concentration of each metabolite produced from both strains is displayed as a function of time in figure 4.9. The rate of glycerol usage by both strains is also assessed.

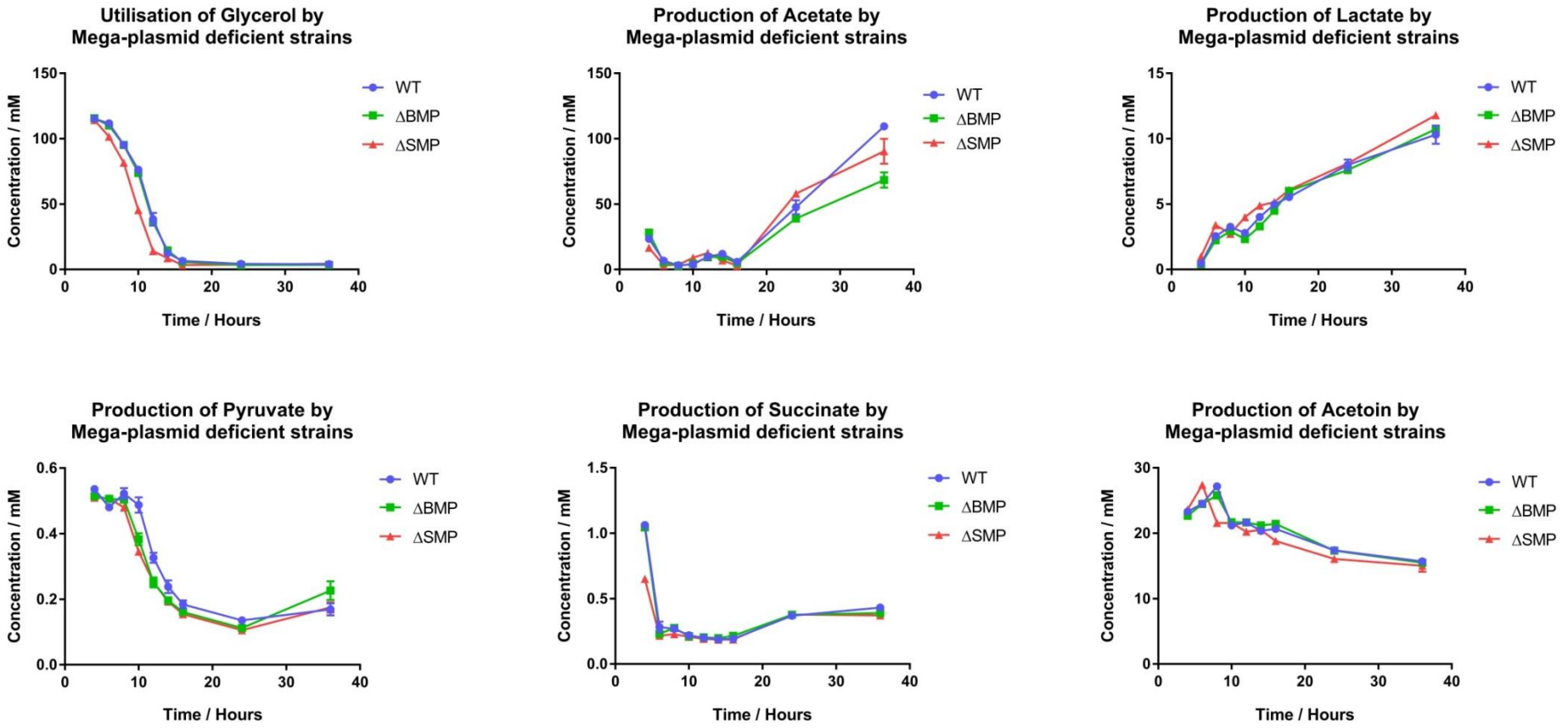


Figure 4.9 – The concentration of metabolites produced and glycerol utilised, by the pNCI001-free and pNCI002-free strains compared to wild-type (WT). The samples were grown using aerobic conditions in baffled conical flasks at 60 °C. The error bars on the graph denote the standard error of the mean (n = 3).

In general, across all strains over the 36-hour growth period, no significant differences were observed in the production of the targeted metabolites or in the utilisation of glycerol, between the mutants and the WT sample. The rate of acetate production did appear to decrease after around 24 hours of growth for both megaplasmid deficient strains, more significantly for the pNCI001-free strain compared to pNCI002-free isolate. One limitation of HPLC analysis is that it requires the successful separation of metabolites on the column and, to fully identify and quantify a metabolite, a calibration step must be undertaken using a reference sample. If a sample presents an unknown composition, it can therefore be laborious to ensure all compounds within a sample have been correctly separated and accounted for. Nuclear magnetic resonance (NMR) instead offers the ability to analyse the sample composition as a whole, which may help to identify any compounds present which were not detected using HPLC.

Prior to NMR analysis, the samples were grown in a minimal ASYE media, supplemented with 0.5 % glucose. The samples were then centrifuged and the supernatant analysed. The spectra, displayed in figure 4.10, were all base line corrected using the internal standard peak, deuterium D_20, identified in the spectra by the red arrow. A large number of the peaks in the region 3.2 – 3.9 ppm were confirmed to correspond to the glucose present in the samples, highlighted in figure 4.10A. However, the NMR spectra for the megaplasmid deficient strains did exhibit some clear differences compared to the WT spectrum. Key peaks present in the

spectra for the mutants, but absent in the WT spectrum, include two doublets at 7.18 and 7.17 ppm (Figure 4.10B), a singlet at 2.24 ppm (Figure 4.10C) and a doublet at 0.96 ppm (Figure 4.10D). It was observed that the internal standard peak was shifted by 0.1 ppm within the obtained spectra. A singlet at 2.24 ppm is within the region expected for an alkyl group attached to a benzene ring (expected around 2.34 ppm). There is no splitting of the peak, which confirms that the hydrogens from this group are adjacent to a carbon atom void of any hydrogens. The pair of doublets at 7.18 and 7.17 ppm are indicative of hydrogens present on a benzene ring. The splitting of the peaks to form doublets denotes the presence of carbons attached to single hydrogen atoms, which supports this assignment. The doublet at 0.96 ppm is the hardest to identify as it falls within a region generally assigned to alkyl hydrogens. A significant peak reduction in the mutant spectra compared to the WT is also observed, for a doublet at ~ 1.21 ppm. This peak is interesting as it is completely lost in the spectrum for the BMP strain, yet present but at a reduced level in the SMP strain compared to the WT.

There are other, smaller differences observed within the spectra produced for the plasmid-free strains, both increases and decreases in peak intensities compared to the WT. In addition, another potential difference between the strains may be elucidated by a shift in the position of the peak indicative of the acetate group, from ~ 1.81 ppm in the WT spectrum to ~1.84 ppm in both the mutant spectra.

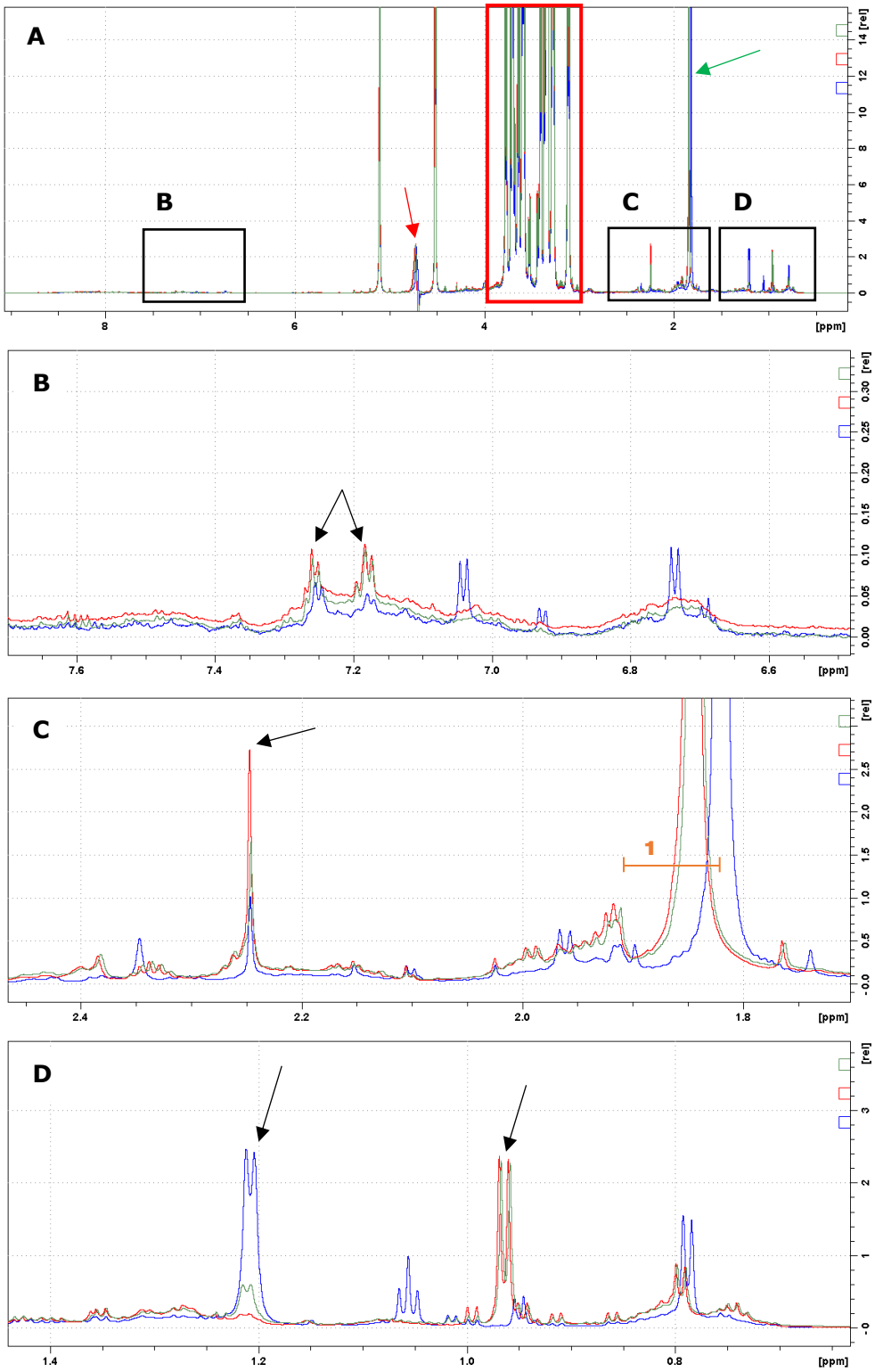


Figure 4.10 – Nuclear magnetic resonance (NMR) spectra elucidating the supernatant composition of pNCI001-free *P. thermoglucosidasius* (Red) and pNCI002-free *P. thermoglucosidasius* (Green) compared to WT *P. thermoglucosidasius* (Blue). (A) The entire spectra for all the samples overlaid. The red arrow indicates the D_2O peak, the green arrow denotes the acetate peak and the red box highlights the region populated

with glucose-related peaks. (B) A zoom in showing the region 7.6 – 6.6 ppm, with the black arrow indicating the pair of doublets at 7.18 and 7.17 ppm (C) A zoom in of the region from 2.4 – 1.8 ppm, with the black arrow indicating a singlet peak at 2.24 ppm (D) A zoom in showing the region from 1.4 – 0.7 ppm, with the black arrow indicating the doublet peak at 0.96 ppm. All spectra denote the chemical shift (ppm) on the X-axis against the normalised intensity on the Y axis.

Prior to the curing of the two megaplasmids, *E. coli*/*Parageobacillus* modular shuttle vectors utilising the pNCI001 or pNCI002 replicons could not be used due to plasmid incompatibility. However, once the megaplasmids had been cured, use of such plasmids becomes feasible. To demonstrate that this was indeed the case, vectors were produced containing the replication proteins RepA and RepB, in addition to the reporter gene eCGP123, and transformed into the megaplasmid-free strains. The plasmids were successfully transformed into the strains and expression of eCGP123 was observed (data not shown).

4.2.2.3 Knock-ins using CRISPR/Cas9 Technology

Currently, ACE is the preferred method used for the KI of genes and operons within *P. thermoglucosidasius* (Sheng, 2017). Therefore, the potential offered by CRISPR/Cas9 to selectively KI at almost any location offers greater flexibility. To validate that this was possible, an initial vector was designed to KI the eCGP123 reporter gene (Figure 4.11). A similar method was used to the one employed in section 4.2.2. However, instead

of producing a KO cassette consisting of HA flanking the gene targeted for removal, a HA containing exogenous DNA is produced, flanking the region that removes the PAM sequence targeted by Cas9. The chosen reporter gene, eCGP123, was expressed under the control of P_{gapd} and was targeted for KI downstream of the orotate phosphoribosyltransferase (*pyrE*) gene. After construction, the vector was transformed into *P. thermoglucosidasius* and passaged as previously described (Section 4.2.2) to produce the mutant. The plasmid was cured and the KI was successfully produced. The mutant was confirmed by both a PCR of the KI site and through the fluorescence intensity detected from the reporter protein, illustrated in figures 4.12A and B respectively.

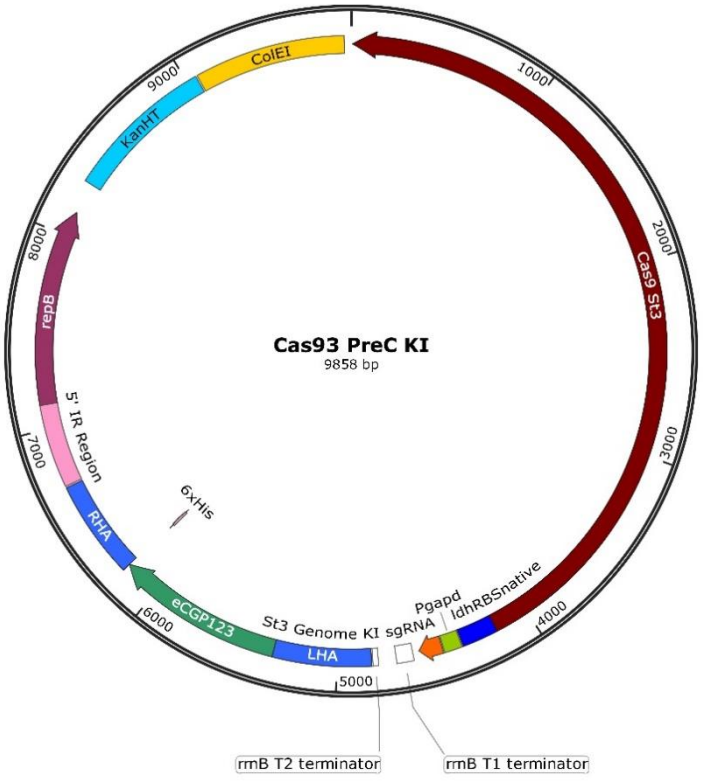
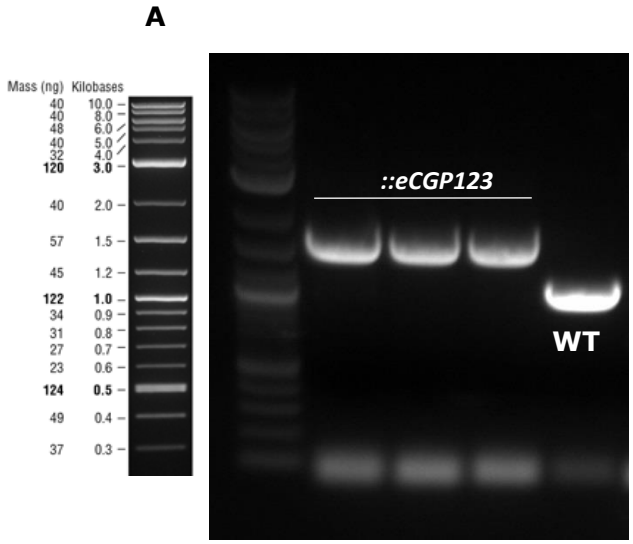


Figure 4.11 – A schematic representation of the vector pCas93_eC_KI containing the homology cassette for the knock-in of the fluorescent gene eCGP123.



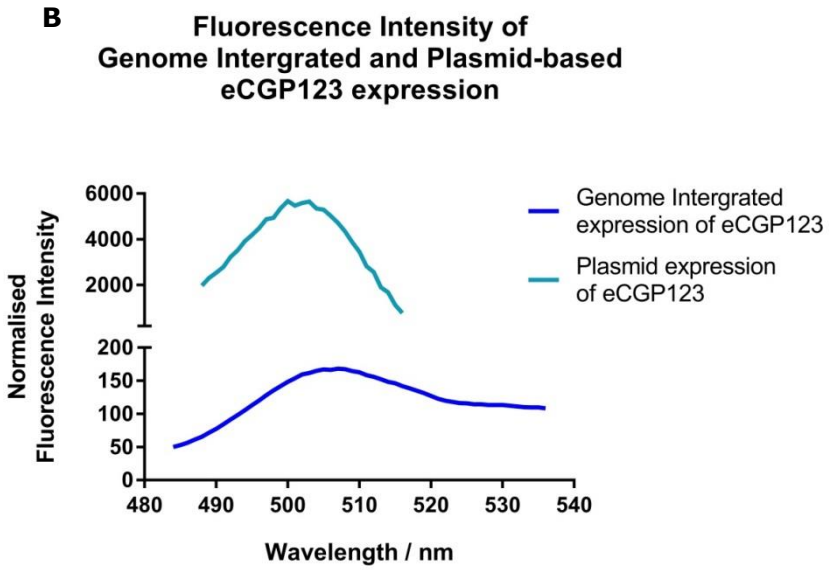


Figure 4.12 – (A) A colony PCR, utilising CP_Genome_eC_F and CP_Genome_eC_R primers, was performed to assess whether the eCGP123 reporter gene had been knocked-in to the genome of *P. thermoglucosidasius*, downstream of *pyrE*. A PCR of WT *P. thermoglucosidasius* was performed as a control (~1.1 kb). Colonies that have the ::eCGP123 would yield a band at ~1.6 kb. A 2-log DNA ladder was used to determine the size of each band. (B) An emission scan (484 – 536 nm) of the protein lysate extracted from the strain ::eCGP123, demonstrating a clear increase in relative fluorescent units (RFU), indicating the presence of the eCGP123 gene. The normalised fluorescence intensity of plasmid-based expression of eCGP123 is also displayed to demonstrate the difference in expression level.

Since the loss of the megaplasmids conferred no advantage, or indeed disadvantage, with respect to fitness, utilising the plasmids as recipients for exogenous DNA was instead suggested. Through knocking-in and expressing genes on the megaplasmids, as opposed to the genome, higher expression should theoretically be observed due to their higher copy number. To explore this, it was decided to KI and express the reporter gene, eCGP123 on both the megaplasmids. This could then be compared to the expression levels shown by the KI of eCGP123 in the genome. Using the methods previously described in this section, reporter gene KI vectors were designed for both strain pNCI001 and pNCI002. The location for the KI of eCGP123 was selected to be the two pseudo genes, BCV53_19490 and BCV53_19815, in the pNCI001 and pNCI002, respectively. KI was successful for the pNCI001 (Figure 4.13); however, this could not be replicated for the pNCI002.

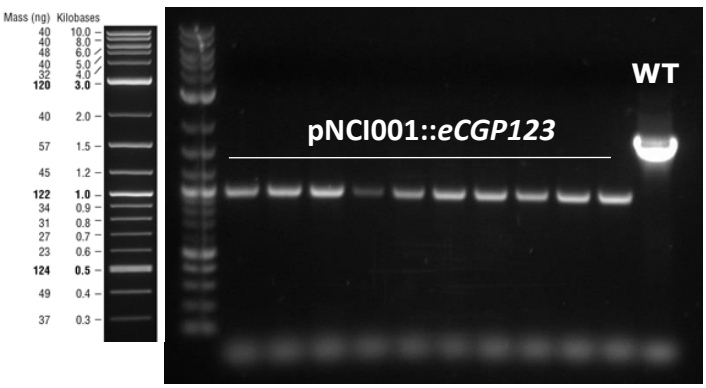


Figure 4.13 – A colony PCR, utilising CP_BMP_eC_F and CP_BMP_eC_R primers, was performed to assess whether the eCGP123 reporter gene had been knocked-in to the pNCI001 of *P. thermoglucosidasius*. The replacement of the pseudo gene, BCV53_19490, produced a band the same size as the WT. Therefore, the primers were designed to bind outside the editing template (CP_BMP_eC_F) and inside eCGP123 (CP_BMP_eC_R) to confirm KI. A PCR of WT *P. thermoglucosidasius* was performed as a control (~2.0 kb),

utilising the primers CP_BMP_eC_F and CP_BMP_eC_R2. Colonies that have the BMP::eCGP123 would yield a band at ~1.0 kb. A 2-log DNA ladder was used to determine the size of each band.

Furthermore, an attempt was made to KI a functional, heterologous (from *G. kaustophilus*) *pyrE* gene into pNCI001 in a Δ*pyrE P. thermoglucosidasius* strain background. The site of insertion was the pNCI001 pseudo gene previously identified. The vectors were constructed as before, but instead using the *pyrE* gene between the HAs for KI. Two sets of editing templates were produced containing *pyrE*, one utilising the native *pyrE* promoter (P_{pyrE}) and the other using the isocitrate dehydrogenase (*idh*) promoter (P_{idh}) from *G. stearothermophilus*. After successful construction and confirmation by sequencing, the vector was transformed into Δ*pyrE P. thermoglucosidasius* competent cells. Following transformation, clean mutants for both *pyrE* constructs were produced using the CRISPR method employed in section 4.2.2. Both mutants were tested on minimal media agar plates, clostridium basal media (CBM), both with and without uracil and also with 5-FOA. Visual observation of the growth is illustrated in figure 4.14, and the results summarised in table 4.2.

It was observed that *pyrE* under the control of P_{idh} was successfully expressed, demonstrating growth both with and without uracil, but not in the presence of 5-FOA. However, *pyrE* under the control of the native promoter was unable to grow without uracil supplementation. The progenitor Δ*pyrE* strain, from which the KIs were produced, was grown in

parallel to confirm the deletion of the native Δ*pyrE*. This deletion was confirmed as the strain did not grow in the absence of uracil and consequently survived in the presence of 5-FOA, unlike the WT control, which was also tested in parallel.

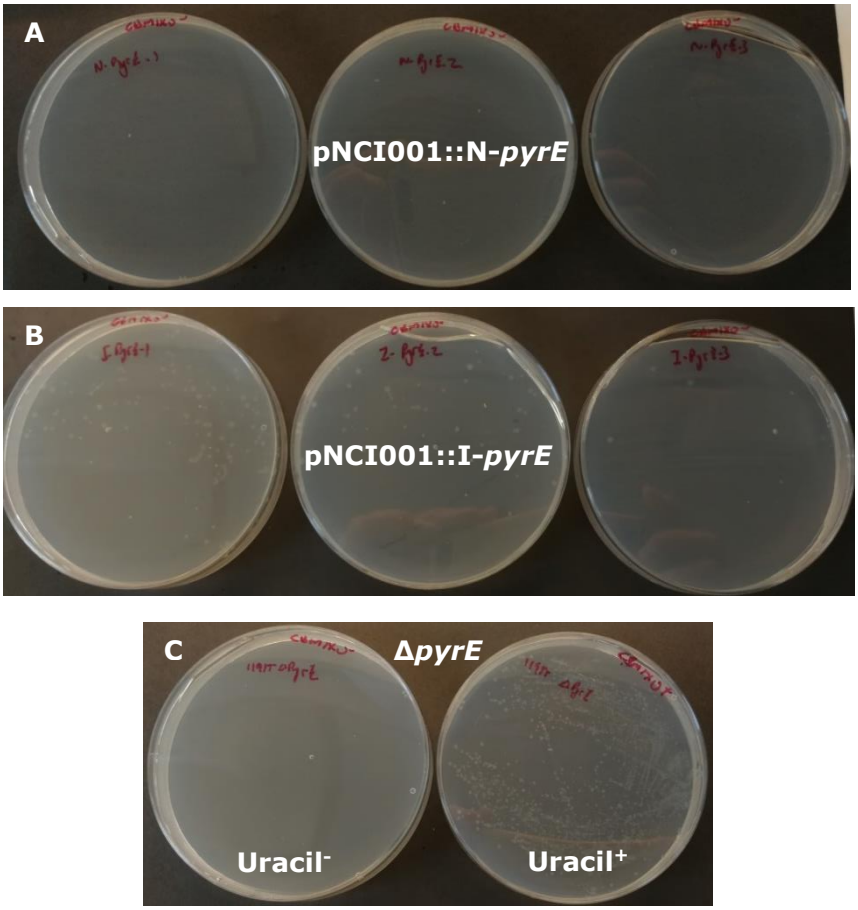


Figure 4.14 – *P. thermoglucosidasius* strains grown on CBM plates. (A) The KI of *pyrE* containing the native *G. kaustophilus* promoter (pNCI001::N-*pyrE*) and (B) The KI of *pyrE* containing the P_{idh} from *G. stearothermophilus* (pNCI001::I-*pyrE*), both without uracil supplementation. (C) Growth of Δ*pyrE* on CBM plates with and without uracil supplementation.

Table 4.2 – A table summarising the growth of *P. thermoglucosidasius* strains on Uracil$^-$, Uracil$^+$ and 5-FOA$^+$, minimal media plates. The KI *pyrE* strains contain either the native promoter (pNCI001::N-*pyrE*) or the P_{idh} isolated from *G. stearothermophilus* (pNCI001::I-*pyrE*), and are compared to Δ*pyrE* and WT *P. thermoglucosidasius* strains. Positive – growth observed, negative – no growth observed.

	Strains			
	pNCI001::N-*pyrE*	**pNCI001::I-*pyrE***	**Δ*pyrE***	**WT**
Uracil$^-$	Negative	Positive	Negative	Positive
Uracil$^+$	Positive	Positive	Positive	Positive
5-FOA$^+$	-	Negative	Positive	Negative

4.3 Discussion

Testing of CRISPR/Cas9

Genetic manipulation of genomes, by knocking-out native genes and knocking-in exogenous DNA, is an essential application for the production of industrially useful strains. Currently, KOs and KIs of genes in *P. thermoglucosidasius* is achieved using AE and ACE, respectively. However, both of these techniques are limited as AE is inefficient and ACE is a location specific method. Therefore, implementation of the CRISPR/Cas9 system, which is both efficient and selective, would simplify and enhance genetic engineering within *P. thermoglucosidasius*. Currently, CRISPR/Cas9 editing occurs in many mesophilic bacteria; however, at the beginning of this study no *cas9* had been demonstrated at temperatures optimum for thermophiles.

This chapter describes the successful implementation of two thermostable *cas9* genes, at 52 °C, to facilitate efficient and selective KI and KO in *P. thermoglucosidasius*. It was demonstrated that *cas9* genes, isolated from *S. thermophilus*, were able to select for successful recombination events that removed the *ack* gene to produce Δ*ack* deficient mutants. Furthermore, successful KI of the eCGP123 reporter gene at two specific locations demonstrated the selective nature *cas9* provides for KI.

Genome editing of *P. thermoglucosidasius* was achieved through the utilisation of two Type II CRISPR/Cas9 systems, *stcas9-1* and *stcas9-3*. These *cas9* genes have previously been demonstrated to produce selective double-stranded (ds) breaks in DNA (125,264), which would be fatal to *P. thermoglucosidasius* due to the lack of non-homologous end-joining (NHEJ) as a mechanism of repair. This was successfully demonstrated in *P. thermoglucosidasius* by the transformation of the *cas9* genes, along with a sgRNA, which resulted in an absence of survivors when plated on agar medium. To confirm that the failure to obtain transformants was due to lethal cleavage of the host chromosome, the two Cas9 genes (*stcas9-1* and *stcas9-3*) were deliberately inactivated through the introduction of a frame-shift mutation. Plasmids carrying these mutated genes transformed *P. thermoglucosidasius* at equivalent frequencies to those observed from plasmids not containing a *cas9* gene, a result also observed by Sapranauskas *et al.* (264) in *E. coli*.

The selectivity of the ds breaks is determined by the ability of *cas9* to identify a specific recognition sequence, unique to the Cas9, called a PAM (130). The PAMs of these *cas9* genes have previously been investigated and determined to be 5'-NNAGAAW-3' and 5'-NGGNG-3' for *stcas9-1* and *stcas9-3* respectively (244-247) and have now been successfully utilised for the first time in a thermophile, demonstrated by the elimination of growth. Furthermore, this study observed the negative effect of transcriptional read through on sgRNA, restricting its ability to form a complex with *cas9*, further highlighting the necessity for a validated

terminator library. Despite the presence of the native hair-pin loop within the sgRNA, a strong promoter can cause transcriptional read through (265), indicating that a stronger terminator is necessary. The strong promoter, P_{gapd}, inhibited the nuclease activity of the *cas9*, which was subsequently returned with the addition of the EcoT1 terminator, validated in section 3.2.5. This result also further confirmed the essential nature of the sgRNA for successful utilisation of the CRISPR/Cas9 system.

Subsequent to demonstrating the activity of both Cas9 proteins at the optimum temperature of *P. thermoglucosidasius*, genome editing was established through the KO of *ack*. Elimination of *ack* was selected due to successful deletion in many different species, including thermophiles, such as *E. coli*, *C. acetobutylicum* and *Thermoanaerobacterium saccharolyticum* (266-268). Due to the function of *ack* in the mixed acid fermentation pathway, it is often deleted to improve expression of valuable products, such as ethanol or succinic acid. This is discussed in more detail in section 6.1.1. Transformation of the constructed *cas9* vector produced successful KO transformants, which were confirmed by colony PCR. Successful genome editing using the *S. thermophilus cas9* has previously been described in human, plants and microorganisms, demonstrating similar activity to that of the more commonly used *cas9* from *S. pyogenes* (246-248). The KO of *ack* in *P. thermoglucosidasius* detailed in this chapter, demonstrates the first use of *stcas9-1* and *stcas9-3* in a thermophile. Initially, this work demonstrated the only example of Cas9-editing at temperatures exceeding 42 °C; however, recently a thermophilic *cas9*,

ThermoCas9, has been identified and characterised from *G. thermodenitrificans*. ThermoCas9 has shown to be active between 20 – 70 °C and has been successfully used for gene deletion in both *B. smithii* and *Pseudomonas putida* at 55 °C and 37 °C, respectively (269).

For future investigations, it would be beneficial to further elucidate the efficiency of each *cas9*. Initial experimentation indicates that *stcas9-3* exhibits better efficiency than *stcas9-1*, a result also demonstrated by Muller *et al.* (247) in human cells. Furthermore, additional KOs at various locations should be attained in order to fully validate CRISPR/Cas9 as a successful genome editing tool in *P. thermoglucosidasius*.

Knock-out of the Megaplasmids

In addition to offering a selective advantage, megaplasmids can impose a metabolic burden on the host (250,255). Unknown functions of megaplasmids can often be elucidated by the curing of the plasmids (256-258). Like many species of geobacilli, *P. thermoglucosidasius* contains extra DNA that is distinct from the chromosome, in the form of two highly stable megaplasmids, a large, ~85 kb megaplasmid pNCI001 and smaller, ~49 kb plasmid pNCI002. However, the genome sequencing carried out by Sheng *et al.* (263) indicated the presence of mostly hypothetical proteins, leaving the function of the megaplasmids unclear. Currently, the function of these two megaplasmids is unexplored. This work detailed in this chapter describes the first successful removal of both megaplasmids from

P. thermoglucosidasius, successfully achieved by utilising the newly implemented CRISPR/Cas9 technology. The curing of the megaplasmids demonstrated no disadvantage to the growth of the cell and further indicated a function devoted to non-laboratory WT bacteria.

Having shown the success of the CRISPR/Cas9 technology in *P. thermoglucosidasius*, it was decided to utilise the KO capabilities to remove the megaplasmids present within *P. thermoglucosidasius*. Initial attempts were made to KO both pNCI001 and pNCI002, through targeting the plasmids replicons. The open reading frames responsible for their replication were identified as RepB (BCV53_19210, BCV53_19380) and RepA (BCV53_19650) respectively (263). As previously discussed, *stcas9-3* appears the more advantageous *cas9* identified from *S. thermophilus* and was therefore selected for use in this study. This attractiveness is due to the more prevalent PAM sequence (NGGNG), shown by the higher number of targets identified within the *ack* gene (Section 4.2.1.1), in addition to the better efficiency exhibited by *stcas9-3*, as demonstrated by Muller *et al.* (247). Curing of megaplasmids is commonly achieved through the addition of negatively selectable markers (256,258,260), such as *sacB*, which are often integrated into the megaplasmid through suicide plasmids or transposons. In addition, megaplasmids have been removed through incompatibility curing, the use of curing agents and by elevating growth temperature (257,261,262). By targeting the SMP replicon using *stcas9-3*, it was observed that the megaplasmid was completely lost, representing a new, more efficient and less laborious method for the removal of plasmids.

In contrast, no viable cells were observed after transformation of *P. thermoglucosidasius* using the plasmid containing the pNCI001 KO cassette. Further inspection identified that this was due to a toxin-antitoxin system, which ensures the survival of transferable genetic elements, such as plasmids, through post-segregational killing (PSK) (270). Hence, by targeting the replicon of the pNCI001, cell death was observed through the action of the toxin, shown by the lack of viable cells post transformation. To circumvent this, the toxin gene was instead targeted to eliminate the PSK, to allow removal of the pNCI001 replicon and therefore subsequently the plasmid. Yet, through targeting the toxin the plasmid itself was lost from the cell, without the need to target the replicon. This loss suggests that pNCI001 is inherently unstable and that its maintenance is dependent on the toxin-antitoxin system (271).

Plasmid maintenance by the mechanism of a toxin-antitoxin system is a commonly reported phenomenon (270,272), as well as the ability for toxin-antitoxins to stabilise specific areas of DNA (273,274). For instance, it has been shown that a large area within the *E. coli* genome was lost when the toxin-antitoxin gene was removed (275). Other reported functions of toxin-antitoxin systems include regulating gene expression, bacterial population control and programmed cell death, which occurs under environmental stress such as DNA damage. Activation of toxin-antitoxin systems can also occur during a phage infection, which also demonstrates the symbiotic nature of bacterial communities (276,277). In

addition to the creation of strains deficient in one or other megaplasmid, a single strain deficient in both megaplasmids was also created.

All three megaplasmid deficient strains showed no growth difference compared to WT, indicating that the megaplasmids neither gave a growth advantage nor conveyed a metabolic burden on the cell, similar to the result observed by Oresnik *et al.* (258) in *S. meliloti*. In addition, no clear differences in the production of common metabolites or glycerol utilisation were observed between the mutants and WT, as determined by HPLC-UV. Glycerol was utilised for this study as it has been demonstrated to produce the same overall levels of cell density as glucose, but does not cause the same significant drop in the pH of the culture (278). This indicates that the plasmids have a non-essential role in *P. thermoglucosidasius* under the conditions tested. This has previously been observed in other bacteria, such as *Bacillus anthracis* (279) and *S. meliloti*, which were also shown to grow as efficiently as WT in both complex and defined media (258). However, Oresnik *et al.* (258) did show growth aberration for some sources of carbon when provided independently in defined media. Similarly, Ohtani *et al.* (257) also demonstrated the return of normal growth on nutrient-rich medium, suggesting that genes present on the megaplasmid are non-essential during optimal growth conditions. Furthermore, it has been noted that gene transfer, from a plasmid to the genome, can occur for essential genes (280). Growth in this study was performed on the defined media, modified ASYE (7), containing glycerol as a carbon source; therefore, further growth conditions should be

considered for future exploration, including growth on a minimal media to fully assess the effect of curing the plasmids. Moreover, it should be noted that curing of four toxin-antitoxin modules of a megaplasmid in a different *S. meliloti* strain caused serious growth defects (281), demonstrating that the effect of losing a megaplasmid is species dependent.

Testing of Megaplasmid KO strains using NMR

In order to assess metabolite production using HPLC-UV, analytical standards for common metabolites are analysed in parallel, in order to correctly identify and also quantify eluted peaks. However, assessing compounds present within a sample with an unknown composition requires a significant amount of method development and can be laborious and time consuming. As discussed, the HPLC-UV analysis did not elucidate an altered production, related to the loss of the megaplasmids, for any of the common metabolites tested. However, as only a small number of metabolites were tested using this method, it could not be guaranteed that the megaplasmid loss was having no effect at all. It was therefore decided to investigate the use of NMR as an alternative analysis technique, as this allows simultaneous analysis of the whole sample composition, which means that is more adaptable to detecting the presence of unexpected compounds.

Prior to NMR analysis, growth was carried out in minimal media in order to reduce the number of peaks produced from complex media

components. As demonstrated in section 4.2.2.2, significant differences between the mutant and the WT spectra were observed, both increases and reductions in numerous different peaks. Interestingly, some of these peaks related to hydrogen groups connected to a benzene structure. These peaks could be interpreted to originate from aromatic hydrocarbon compounds such as toluene Ethyltoluene or possibly aromatic alcohols, such as phenol. This suggestion is supported by the presence of numerous genes on the pNCI001 that are involved in the catabolism of aromatic hydrocarbons. One of these is the enzyme 4-oxalocrotonate tautomerase, which catalyses the isomerization of unsaturated ketones and is also part of the oxidative metabolic pathway that catabolises these aromatic hydrocarbons into intermediates of the TCA cycle (282-284). Several other genes present on pNCI001, namely catechol 2,3-dioxygenase, 4-oxalocrotonate decarboxylase, 4-hydroxy-2-oxovalerate aldolase, 2-keto-4-pentenoate hydratase and an acetaldehyde dehydrogenase (285-288), have also been found to be present in the TOL degrading plasmid pWW0, which is involved in the degradation of toluene (287). These genes enable strains of soil dwelling bacteria to utilise simple aromatic hydrocarbons, such toluene, as their sole carbon source (289). The presence of these same genes on pNCI001may therefore indicate that this plasmid is essential in countering environmental stress caused by limited carbon sources, a theory that is supported by the fact that *P. thermoglucosidasius* was isolated from soil in Japan (53), providing a reason for the acquisition of this mobile element. Furthermore, these genes, in collaboration with flavin-dependent monooxygenase and flavin reductase (290) have been

demonstrated by Omokoko *et al.* (288) in *G. stearothermophilus* to participate in an aromatic degradative meta-pathway, which facilitates the degradation of phenol. It is therefore suggested that by removing pNCI001, the metabolic pathways that catabolise these aromatic compounds are negatively affected, resulting in an increase in their concentration, which was detected within the mutant NMR spectra. This suggests that the megaplasmids are involved in the degradation of aromatic hydrocarbons as a carbon source to facilitate the production of pyruvate. Future work should therefore be focused on further elucidating this suggested function, which could be assessed by growing the strains in the presence of aromatic hydrocarbons to observe whether they can utilise these compounds as a sole carbon source.

Another interesting observation from the NMR was a shift in the position of the acetate peak for the mutant strains compared to the WT. This spectral shift denotes that the acetate group produced by the mutants is within a different chemical environment to that produced by the WT. However, without further investigation this is purely speculative. This is the case for the NMR analysis as a whole; the full characterisation of these spectra, including precise peak assignments, requires a further, more in-depth exploration. However, the NMR analysis has highlighted the limitations surrounding the traditional HPLC method for exploring an unpredictable sample composition and importantly, has elucidated that the removal of the megaplasmids has led to some observable differences in metabolite production.

One limitation to this study is the conditions in which the bacteria have been grown. Often, the role of plasmids is to allow the host to survive unfavourable environmental conditions and hence an *in vitro* expression of the bacteria may not be resulting in utilisation of the plasmid. In addition, large numbers of hypothetical proteins have also been identified on the pNCI002, with their functions yet to be explored. Therefore, the lack of metabolic effect caused by removing these plasmids may indicate a nonessential role under the conditions tested, however, further investigation is necessary. Under the conditions used for testing, the lack of metabolic differences demonstrates that the replicons of each the megaplasmids can be utilised as additional replicons, in order to produce further shuttle vectors for use in *P. thermoglucosidasius*, as is frequently applied in other bacteria (257).

Lastly, it should be considered that pNCI001 may be present purely due to the toxin-antitoxin module, as shown by the loss of the plasmid when the toxin is removed. One hypothesis is that the toxin-antitoxin module can be viewed as an addiction system (292). The PSK effect that toxin-antitoxin systems exhibit, ensures the maintenance of these DNA modules. It has been suggested that this perceived stability conferred to the plasmid is inconsistent and more likely a consequence of plasmid competition, as opposed to conserving a necessary function, leading to the theory of 'selfish DNA' (293). This can also be seen in chromosomally encoded toxin-antitoxin modules which are capable of moving from genome to

genome, through horizontal gene transfer. These modules appear to be moving continuously and can be viewed as recent events, as shown by different toxin-antitoxin distribution patterns between different *E. coli* isolates. This suggests that these toxin-antitoxin modules act as selfish elements (271). Nevertheless, it appears most likely that these megaplasmids are genetic relics from growth in a WT environment and are no longer necessary to *P. thermoglucosidasius* grown under laboratory conditions, with the toxin-antitoxin system maintaining the prevalence of this advantageous plasmid. However, with respect to producing biofuels, these aromatic hydrocarbon degradation pathways may present a key role in the breakdown of polyphenols present in lignocellulosic biomass, offering a significant advantage over similar microorganisms.

Utilisation of CRISPR/Cas9 for Knock-ins

In addition to removing specific genes from an organism's genome, KI of exogenous DNA can further drive production of desired metabolites. KI can be achieved by AE, but this is often inefficient. Therefore, ACE is used within *P. thermoglucosidasius* to increase efficiency and allow the selection of recombinant cells (118). However, gene integration using ACE represents a location specific method and therefore does not allow flexibility. The ability to utilise the specificity of the CRISPR/Cas9 technology would offer a key advantage to genetic engineering of *P. thermoglucosidasius*.

This study demonstrates the first use of *stcas9-3* to KI exogenous DNA into *P. thermoglucosidasius*. Using similar methods developed in this study, *stcas9-3* was employed to KI the reporter gene, eCGP123, into both the chromosome and the pNCI001. The chromosomal location to KI the reporter gene was selected downstream of the *pyrE* gene. This locus is theoretically a safe place for KIs, as it has previously been shown to be successful using the ACE method (118). The reversion of the Δ*ack* gene back to WT would also have been a viable choice, recently demonstrated by Mougiakos *et al.* (237) using spCas9 to reconstitute lactate production in the *B. smithii* strain ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR*. However, it was decided that successful integration of exogenous DNA would be a more beneficial target. In addition, eCGP123 is a convenient marker as it produces a clear phenotype which can be easily validated. After successful transformation, KI transformants were confirmed by PCR of the chromosome locus and through the detection of fluorescence using the eCGP123 assay. Expression was observed to be lower than plasmid expression of eCGP123; however, this is as expected as copy number of the genome is lower than that of the plasmid. However, chromosomal KIs are far more stable than genes expressed on a plasmid. Efficiency of *stcas9-3* showed 100 % success of KI of eCGP123, a much higher efficiency than spCas9, which showed 20 % and 5 % for constructs containing HAs of 1 kb and 0.75 kb respectively (237). Both spCas9 constructs contained longer HAs compared to the ones used in this study, which were 0.5 kb. Mougiakos *et al.* (237) demonstrated that efficiency increased with the length of the HAs, which suggests that any decrease in efficiency observed for future

constructs can be compensated by increasing the length of the HAs. Although 100 % efficiency was achieved for *stcas9-3*, this was only demonstrated for a screen of three colonies. Therefore, it is imperative to screen a much larger sample size to fully assess efficiency.

After successfully showing the capabilities of *stcas9-3* to be used for KI in *P. thermoglucosidasius*, it was decided to attempt a KI into the megaplasmids. Both the pNCI001 and pNCI002 have demonstrated stability during replication and are present at a higher copy number compared to the genome (3X and 5X, respectively) (263). Furthermore, after showing no observable *in vitro* advantage or disadvantage they appear attractive targets to exploit as potential extra-chromosomal vehicles in *P. thermoglucosidasius*. Therefore, they may offer a desirable place for the expression of operons for the production of desired metabolites. The pseudo genes BCV53_19490 and BCV53_19815 were selected as the locations to KI eCGP123. The pseudo gene BCV53_19490 has both a missing start and stop codon and is also frameshifted, which produces an internal stop codon. The gene BCV53_19815 is also frameshifted to produce an internal stop codon. Successful KI of the eCGP123 reporter gene was shown in the pNCI001; however, no KI was attainted for pNCI002. The KI of eCGP123 demonstrated the success of a second KI using *stcas9-3* at a second location. This establishes *stcas9-3* as a viable tool for efficient and specific KI within *P. thermoglucosidasius*, overcoming the previous limitations of KI methods. Conversely, KI to pNCI002 was unsuccessful, showing that the plasmid was instead cured

from the cell. This indicates that there is no selection pressure conserving pNCI002and the drive for HR is weaker than the efficiency of *stcas9-3* to perform a lethal ds break. Unlike pNCI001, pNCI002 contains no observed toxin-antitoxin module to ensure the survival of the plasmid. This further corroborates with the early observation that pNCI002 is nonessential to *P. thermoglucosidasius*, as instead of undergoing homologous recombination to remove the theoretically lethal Cas9 incision, the plasmid is cured. However, as previously discussed, this non-essential nature may be a result of *in vitro* growth. Again, further exploration is necessary to confidently confirm or refute this theory. The successful production of the desired strains was made a priority and therefore once achieved, progression to testing and creation of further strains was conducted. To fully determine efficiency of KI using *stcas9-3,* a more stringent method of screening must be used. Furthermore, altering the growth conditions may allow subsequent KI into the plasmid. Overall, successful KI of eCGP123 was observed for both the chromosome and pNCI001, showing the successful application of the CRISPR/Cas9 system for gene integrations.

Knock-in of *pyrE* as a selectable site for exogenous DNA on pNCI001

The production of valuable chemicals or biofuels often requires the repurposing of the metabolism of industrially useful microorganisms. This often occurs through both KOs and KIs. When integrating exogenous DNA, it is often desirable to overexpress these genes, which can be achieved through strong gene expression elements. However, the use of native,

stable plasmids offers the opportunity for increased expression through an increased copy number. As discussed in section 1.4.6, a positively and negatively selectable marker, such as *pyrE*, is advantageous when knocking-in exogenous DNA. The ability to screen for successful double cross-overs using *pyrE* is one advantage the ACE method has over the CRISPR/Cas9 system. However, as already stated, the limitation of location dependence is the reason for pursuing a method such as CRISPR/Cas9, which is essentially location independent.

This study therefore details the exploitation of CRISPR/Cas9, in combination with the ACE method, to utilise the megaplasmids as vehicles for increased expression of exogenous DNA. Using the previously identified pseudo gene, BCV53_19490, the *pyrE* gene was knocked-in to a Δ*pyrE* deficient strain on pNCI001. Positive selection of these CRISPR mutants was successfully achieved on uracil deficient media, of which the progenitor Δ*pyrE* deficient strain could not grow on. This pNCI001::*pyrE*$^{+}$ can then be truncated and used for KI at this *pyrE* locus on the megaplasmids. This facilitates both positive and negative selection pressure and allows the use of the ACE method to KI genes onto the megaplasmids, which is a proven efficient method for gene integration (118,157). The generation of this mutant can now be used for integration of exogenous genes in order to express a desired product.

4.4 Conclusions and Future Work

This work addresses the development of CRISPR/Cas9 technology for KOs and KIs in *P. thermoglucosidasius*, as well as initial investigation into the function of two megaplasmids that reside within the cell. Specifically, this work describes the identification of two active *cas9* genes from *S. thermophilus*, which are able to successfully target and cut chromosomal DNA. Furthermore, it was shown that the CRISPR/Cas9 technology could be successfully utilised to KO *ack* in *P. thermoglucosidasius*. It was then employed to also target and remove two megaplasmids from within *P. thermoglucosidasius*, through targeting the replicon of pNCI002 and the toxin of a toxin-antitoxin module on pNCI001. Through the curing of the two megaplasmids it was demonstrated that they were non-essential for growth, carbon utilisation or specific metabolite production, under the conditions tested. However, it was speculated that the TOL degradation pathway and/or the aromatic degradative meta-pathway, which breaks down simple aromatic hydrocarbons, may have been affected by the loss of the megaplasmids. Finally, it was shown that *cas9* can also be utilised for KIs, shown through the KI of both the reporter gene eCGP123 and *pyrE*.

Despite this work demonstrating the success of CRISPR/Cas9 in *P. thermoglucosidasius* and an initial investigation into the megaplasmids, there are clear areas within the project that need to be further explored.

This includes testing the efficiency of each *cas9* through screening an increased number of colonies (237). The efficiency of *cas9* could then be optimised through gene expression using different strength promoters or an inducible promoter (294), both of which have been identified in chapter 3. For instance, it has been shown that using a weaker promoter for sgRNA expression results in a decrease in off-targeting, without affecting the on-target score (295). Therefore, it would be prudent to optimise the gene expression system of the CRISPR/Cas9 system. It has also been shown that transformation efficiency declines with increased vector size (296). Therefore, testing a variety of differing HA lengths may also lead to better *cas9* efficiency by increasing transformation efficiency. Another area of CRISPR/Cas9 which is of interest is CRISPR interference (CRISPRi), which utilises a catalytically dead *cas9* (*dcas9*) through the inactivation of the nuclease domains RuvC and HNH, to affect gene transcription (297). CRISPRi can both repress and activate transcription through the sequence specific binding of a promoter. Repression of gene expression is achieved through the blocking of either transcription initiation or elongation, whereas activation can be achieved in the same manner but through the function of activators to the *dcas9* (298-300). The addition of these tools further enhances the stringent control over transcription of an organism, therefore increasing the precision of genetic engineering.

Through this work, it was observed that the megaplasmids may house the gene required for the degradation of aromatic hydrocarbons. To further

identify if this is indeed the case, the megaplasmid deficient strains, along with WT, should be grown using simple aromatic hydrocarbons, such as toluene or phenol, as a sole source of carbon. Furthermore, transformation efficiency has been shown to increase in megaplasmid deficient strains (257) and therefore this should also be tested for the *P. thermoglucosidasius* megaplasmid deficient strains.

As observed through this study, one of the problems of using CRISPR/Cas9 to make KOs or KI in non-essential plasmids is the absence of selection for the desired allelic exchange mutant. When targeting the chromosome, Cas9-mediated cleavage of the chromosome leads to cell death. Accordingly, cells that survive are immune to Cas9 due to removal of the Cas9 target as a consequence of the successful AE-mediated generation of a KO or KI mutant. In contrast, when targeting a plasmid, its cleavage by Cas9 is of no consequence to the cell and so there is no means of counter selection of the WT plasmid, as observed with pNCI002. To circumvent this problem, an attempt can be made to add a gene to the megaplasmid, by AE, whose presence would lead to a specific sensitivity under a defined condition, such as *pyrE*, which was added to pNCI001. Inactivation of this gene would allow which would allow counter-selection. However, to achieve this, the editing cassette would have to be present on a suicide plasmid to ensure AE.

Finally, pNCI002 displays a conjugational DNA transfer system (TrsK). Exploration of the activity of this system would be of interest and could be

demonstrated by conjugation into a pNCI002-free strain. However, this would require successful KI of a selection marker into pNCI002, which has not currently been achieved.

4.5 Key Outcomes

- Identification, construction and testing of two thermostable CRISPR/Cas9 genes, *stcas9-1* and *stcas9-3*, in *P. thermoglucosidasius*.

- Exploitation of the CRISPR/Cas9 system to knock-out *ack* in *P. thermoglucosidasius*.

- Creation of megaplasmid deficient strains through the knock-out of the small megaplasmid replicon and the toxin gene of the big megaplasmid in *P. thermoglucosidasius*.

- Recognition of the non-essentiality of the small and big megaplasmid to the growth, carbon utilisation and metabolite production in *P. thermoglucosidasius*, under the conditions tested.

- Utilisation of *cas9* to allow knock-in of exogenous DNA into both the chromosome and the big megaplasmid.

Chapter 5

Increasing Transformation Efficiency through the Knock-out of R-M Systems

5.1 Introduction

5.1.1 Restriction Modification Systems in *P. thermoglucosidasius*

Bacterial restriction modification (R-M) systems (section 1.4.1.1) are a primitive form of immunity, which prevent the entry of foreign DNA into the cell. This potential barrier to plasmid DNA can present an insurmountable barrier to the effective genetic modification of a particular chassis. Roberts and co-workers have developed a database for DNA restriction and modification (REBASE), which has identified three Type II R-M systems and one Type III system within *P. thermoglucosidasius* NCIMB 11955 (301) (Figure 5.1). This database indicates that one of the Type II systems does not include a restriction endonuclease (REase) domain and is unlikely, therefore, to play a role in restricting foreign DNA. As such, it is most probably does not play a role in the efficiency of plasmid transformation. The recognition sequences of the three other R-M systems are currently unknown.

The ability of R-M systems to inhibit the entry of exogenous DNA often varies between species. For example, within the genus *Clostridium*, representative strains of the species *C. beijerinckii*, *C. perfringens* and *C. difficile* (302-306) have been transformed without hindrance. Whereas, for others, such as *C. acetobutylicum*, *C. cellulolyticum* and *C. cellulovorans*

(307-311) the R-M systems had to be circumvented before successful transformation could be achieved.

Often, R-M systems can be circumvented through the appropriate methylation of the recognition sequences present on the plasmid. This can be achieved *in vivo* by using an *E. coli* donor strain, which contains specific methylase enzymes to methylate the DNA in a specific pattern (312). It is also possible to methylate the DNA *in vitro*, using commercially purchased methylase enzymes (313). However, as previously discussed, the recognition sequences of the R-M systems in *P. thermoglucosidasius* are unknown; therefore, specific methylation of the DNA to circumvent the native restriction system is not currently possible. A third method requires the identification and inactivation of the requisite component of the R-M system, specifically the gene encoding the REase (314).

Although R-M systems are present within *P. thermoglucosidasius*, successful transformation is achievable without the need to evade these systems. However, the transformation efficiency within this species is considered to be low compared to other bacteria (315,316), with efficiencies previously observed in the range of $1 \times 10^3 - 1 \times 10^4$ colony forming units (CFUs) per µg of DNA (51,118,180). Moreover, a reduction in transformation efficiency has also been observed when using the CRISPR/Cas9 system. The removal of these native R-M systems represents a strategy for increasing transformation efficiency within *P. thermoglucosidasius*. As AE is directly linked to transformation efficiency,

this would lead to higher efficiency achieved for methods such as CRISPR/Cas9 and ACE, which both rely on AE.

As previously alluded to, not all R-M systems appear to play a role in immunity. Other roles have been suggested for R-M systems, including transcription regulators (317,318) and the stabilisation of genomic islands (319,320). Many of the features associated with R-M systems have also been attributed to toxin-antitoxin systems. Therefore, R-M systems can also be viewed as addiction systems and selfish DNA. Currently, the function of the R-M systems present in *P. thermoglucosidasius* is unknown and therefore by knocking-out, their role can be elucidated.

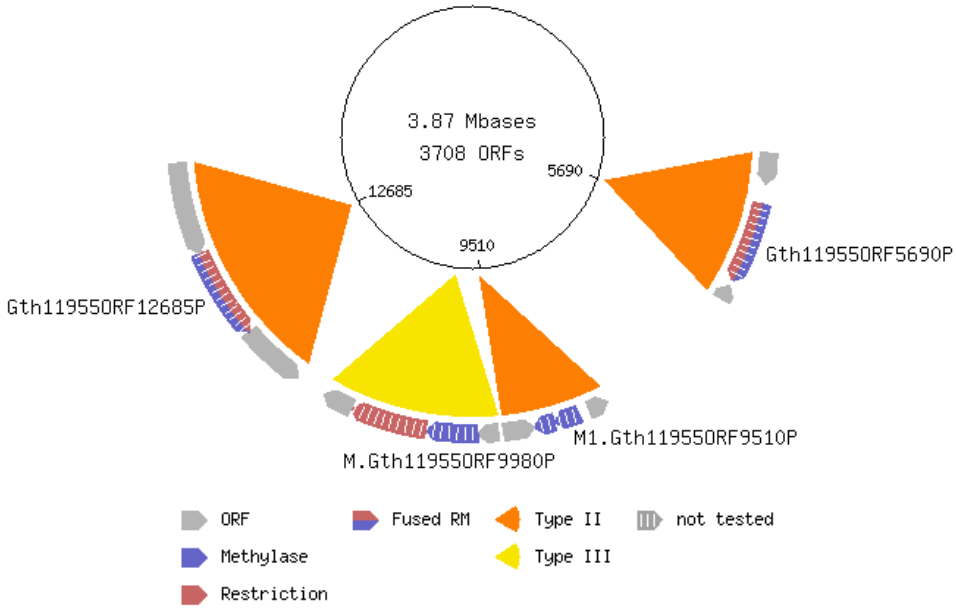


Figure 5.1 – A representation of the position and type of restriction modification (R-M) systems present on the *P. thermoglucosidasius* genome. Orange segments indicates Type II systems; Yellow segments indicates Type III systems. Gene colour: Red – Restriction activity, Blue – Methylase activity, Blue/Red – Fused activity, Grey – a new open reading frame (ORF). Image obtained from REBASE (Roberts, 2015; http://rebase.neb.com).

5.1.2 Chapter Aims

This chapter aims to investigate the currently unexplored R-M systems present on the genome of *P. thermoglucosidasius*. The initial focus is to establish the transformation efficiency of WT *P. thermoglucosidasius*, prior to the deletion of the three identified R-M systems. The aim is to achieve KO of these R-M systems, two putative Type IIG systems and one putative Type III system, using the CRISPR/Cas9 method developed in chapter 4. The successful KO of these native systems will also serve to validate the use of the CRISPR/Cas9 system as a successful genome editing tool for *P. thermoglucosidasius*. Through the production of R-M deficient strains, this research hopes to both elucidate the role of these R-M systems and increase the transformation efficiency of *P. thermoglucosidasius*. This will enable an increase in the efficiency of AE and ultimately genetic engineering of *P. thermoglucosidasius* as a whole.

5.2 Results

5.2.1 Single Knock-out of Restriction Modification Systems

5.2.1.1 Transformation efficiency of Wild-Type P. thermoglucosidasius

The initial stage of this work was to establish the transformation efficiencies for WT *P. thermoglucosidasius* to act as the baseline for comparing the efficiencies in subsequently generated R-M mutants. The shuttle vector pMTL61110 was transformed into *P. thermoglucosidasius* competent cells. Transformations were performed using 500 ng of plasmid DNA, added to an aliquot (60 µL) of competent cells. The time constants of each transformation were consistently between 5.1 – 5.3 ms. Cells were immediately added to pre-warmed 2SPY and allowed to recover at 52 °C at 250 RPM for 3.5 hours. After recovery, the cells were plated onto TSA km plates and left to grow O/N. For each transformation two dilutions, specifically 1 x 10^{-2} and 1 x 10^{-1}, together with an undiluted sample were grown. The colonies of individual plates were then manually counted to determine the transformation efficiency of each sample, calculated by the number of CFUs per µg of DNA (Figure 5.2). The transformation efficiency of WT *P. thermoglucosidasius* was shown to be exceedingly variable, with a two order of magnitude difference between the highest and the lowest transformations. The highest transformation

observed for a single colony was measured at ~ 4 x 10^4 CFUs per µg of DNA, with the lowest at ~ 6 x 10^2 CFUs per µg of DNA. This unexpected result indicates a need to not only increase the efficiency but also the consistency of *P. thermoglucosidasius* transformations, which may be achieved through the removal of the R-M systems.

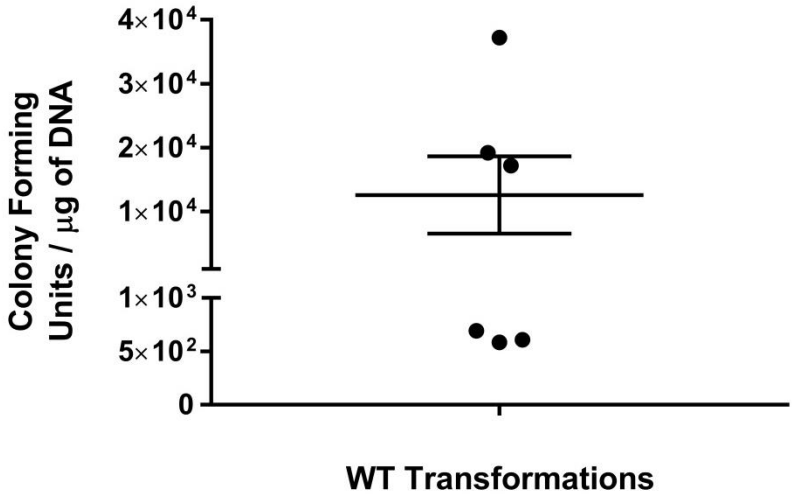


Figure 5.2 – The vector pMTL61110 (500 ng) was transformed into WT *P. thermoglucosidasius* competent cells. After recovery, the cells were grown and the produced colonies were counted to determine the number of colony forming units per µg of DNA. The error bar represents the standard error of the mean (n = 6).

5.2.1.2 Construction of R-M Knock-out vectors and Single Knock-out R-M Systems

As discussed in section 5.1.1, REBASE indicated the presence of three Type II and one Type III R-M systems in *P. thermoglucosidasius*. However, one of the Type II R-M systems, appeared to lack a restriction enzyme within the gene and therefore this R-M system was excluded from the study. The

remaining three R-M systems, named RM1, RM2 and RM3 for the purpose of this chapter in the order they appear on the genome, were targeted for deletion using the CRISPR/Cas9 technology developed in chapter 4. The construction of the CRISPR/Cas9 vectors followed the previously developed methodology (section 4.2.2). Primers for the PCR of each sgRNA were designed and cloned into the *stcas9-3* vectors for targeting of each R-M system. Primers for SOE PCR of the HAs were designed for each of the three R-M systems. The HAs, to facilitate AE, were selected from upstream and downstream of the gene of interest and SOEing PCR was used to produce the complete editing template. The completed constructs were sequenced, before transformation into *P. thermoglucosidasius* electrocompetent cells was achieved. For each R-M system, three different sgRNA sequences were selected for testing and for each R-M sgRNA combination three replicates were grown. Colonies were then picked and cultured in 2SPYNG km (10 mL). After 6 x 12 hour passages, each culture was screened for possible mutants. To avoid mixed cultures, the samples exhibiting mutants were diluted to 1 x 10^{-5} and plated onto TSA km plates and incubated O/N (52 °C). Individual colonies were then screened by PCR and ran on an agarose gel, in order to identify clean mutants (Figure 5.3). These mutants were then cultured at an increased temperature (60 °C) to facilitate the loss of the CRISPR/Cas9 plasmid. Colonies were screened for successful plasmid loss through PCR.

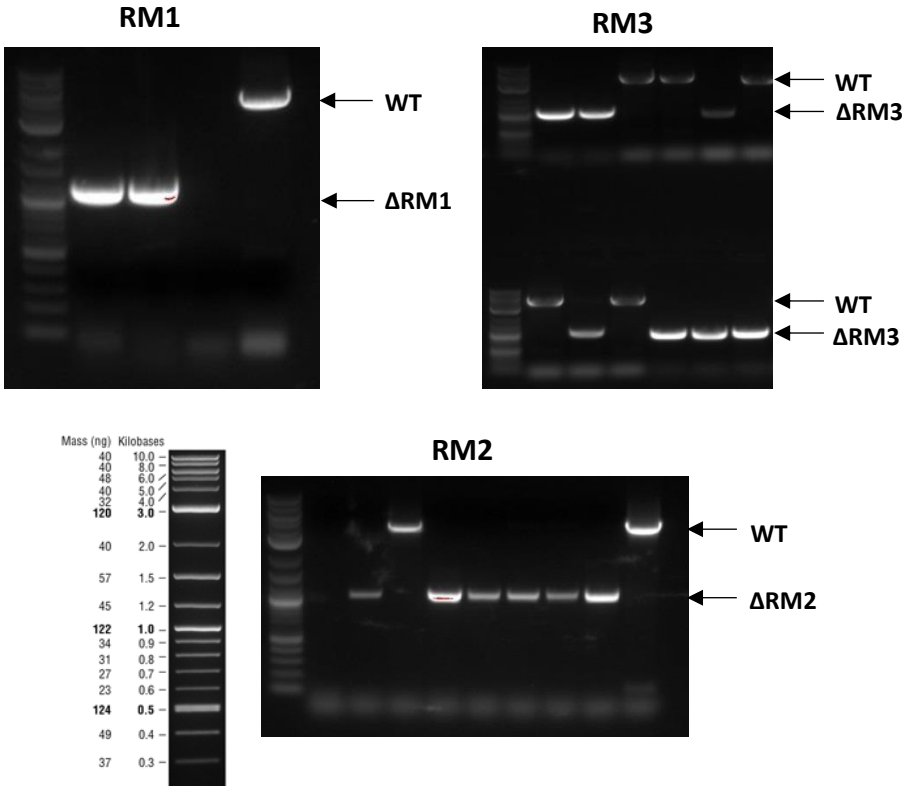


Figure 5.3 – A colony PCR was performed to assess whether the R-M genes, RM1, RM2, and RM3, had been knocked out of the *P. thermoglucosidasius* genome. The primers used were (RM1) CP_RM1_F and CP_RM1_R, (RM2) CP_RM2_F and CP_RM2_R, and (RM3) CP_RM3_F and CP_RM3_R. A PCR of each WT R-M system was performed as a control (~4.6 kB, ~4.0 kB and ~4.3 kB respectively). Colonies that have the R-M system deleted would yield a band at ~1.1 kB. A 2-log DNA ladder was used to determine the size of each band.

5.2.1.3 Genome Sequencing of Restriction Modification Knock-out Strains

The genome of each strain, ΔRM1, ΔRM2 and ΔRM3, was subjected to Illumina MiSeq (DeepSeq, Nottingham, UK) sequencing to confirm the successful mutations created through the CRISPR/Cas9 mediated

deletions. The genomic DNA was prepared via phenol:chloroform extraction. The strains sequenced were mapped against the reference genome (263). Single nucleotide polymorphisms (SNPs) identified in each of the samples are indicated in table 5.1. The genome sequencing of each mutant shows clear, clean deletions of the gene of interest, with no major deletions produced at other points across the chromosome.

Table 5.1 – A list of all the SNPs identified from the *stcas9-3* deletions of ΔRM1, ΔRM2 and ΔRM3. SNP – single nucleotide polymorphism, SNV – single nucleotide variation, CDS – coding sequence.

ΔRM1					
Reference Position	**Type of SNP**	**Length**	**Reference**	**Allele**	**Gene**
1128033	SNV	1	T	A	*adh*
1349479	SNV	1	T	A	CDS
1884695	SNV	1	T	G	CDS Membrane
2723791	SNV	1	C	A	Mobile Element
3069936	Deletion	1	A	–	Non-coding
3268970	Deletion	5	TTGCA	–	CDS Transporter
ΔRM2					
Reference Position	**Type of SNP**	**Length**	**Reference**	**Allele**	**Gene**
1358124	SNV	1	G	T	CDS ATPase
1605217	Deletion	1	T	–	Threonine dehydratase
1966875	SNV	1	G	A	Spore Germination

2279429	SNV	1	T	A	CDS Exporter
2719108	Insertion	1	–	A	Non-coding
ΔRM3					
Reference Position	**Type of SNP**	**Length**	**Reference**	**Allele**	**Gene**
1162676	SNV	1	C	T	Nitrate Reductase
1394477	SNV	1	C	T	Cytochrome C Biogenesis
1605026	SNV	1	C	T	Threonine dehydratase
2278524	SNV	1	T	A	CDS Exporter
2529810	SNV	1	G	A	Non-coding
2606228	SNV	1	A	G	Non-coding
2719108	Insertion	1	–	A	Non-coding
2723814	SNV	1	A	G	Mobile Element
2826489	Insertion	1	–	A	Non-coding
3519511	SNV	1	C	G	rRNA
3756419	Deletion	2	TA	–	Non-coding

5.2.2 Creation of a Restriction Modification Deficient Strain

Following the production of the single mutants, three double mutants and one triple mutant was produced. These were created using the same process, by transforming the already constructed R-M KO plasmids into the single and double mutants respectively (Section 5.2.1.2). As previously, after screening for clean mutants, the plasmid was cured (Figures 5.4 and 5.5).

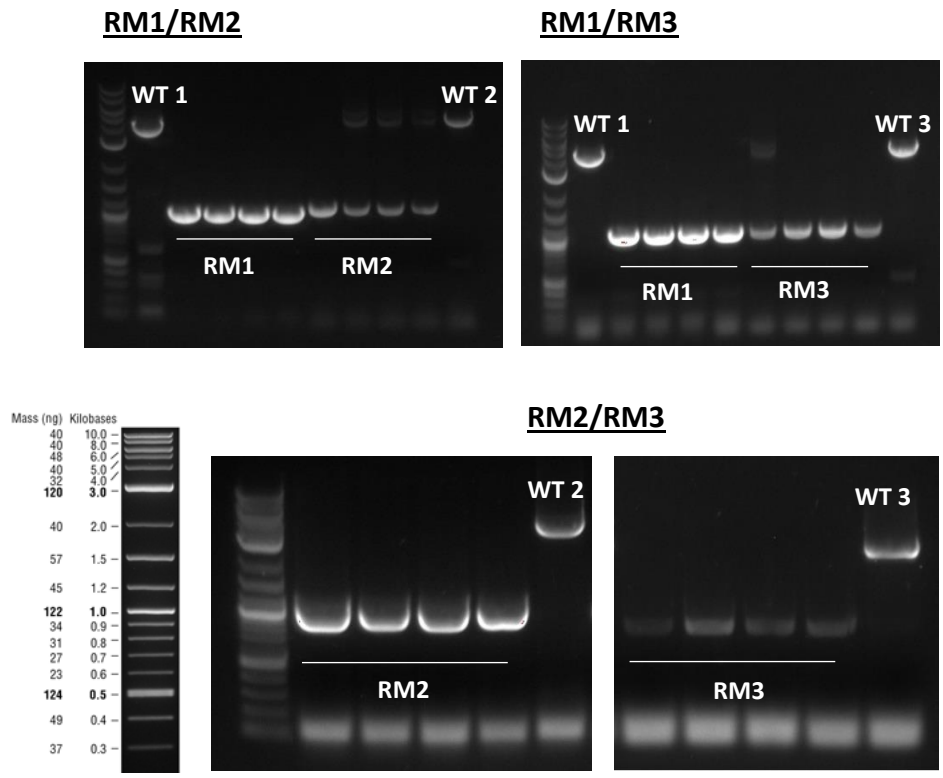


Figure 5.4 – A colony PCR was performed to assess whether the R-M genes, RM1, RM2, and RM3, had been knocked out of *P. thermoglucosidasius* genome. The primers used were (RM1) CP_RM1_F and CP_RM1_R, (RM2) CP_RM2_F and CP_RM2_R, and (RM3) CP_RM3_F and CP_RM3_R. A PCR of each WT R-M system was performed as a control

(~4.6 kB, ~4.0 kB and ~4.3 kB respectively). Colonies that have the R-M system deleted would yield a band at ~1.1 kB. A 2-log DNA ladder was used to determine the size of each band.

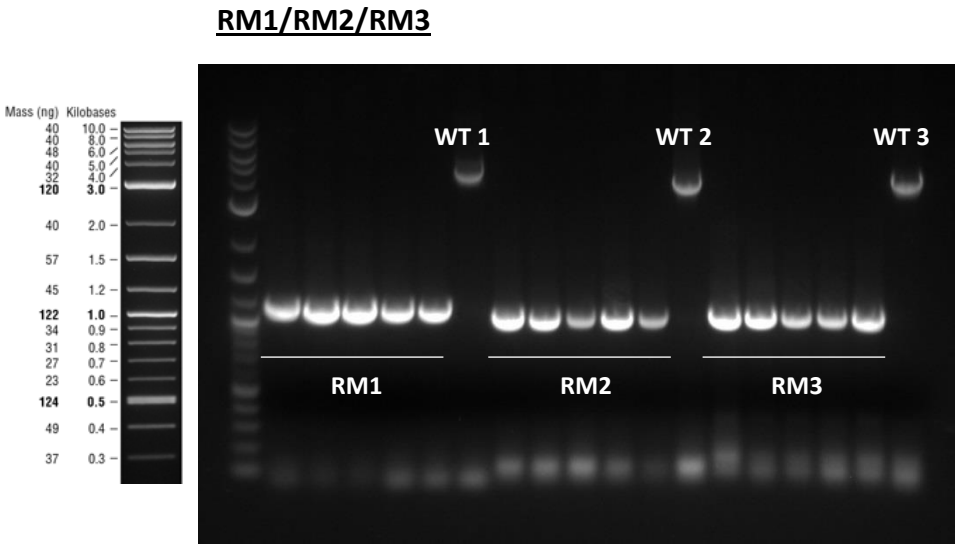


Figure 5.5 – A colony PCR was performed to assess whether the R-M genes, RM1, RM2, and RM3, had been knocked out of *P. thermoglucosidasius* genome. The primers used were (RM1) CP_RM1_F and CP_RM1_R, (RM2) CP_RM2_F and CP_RM2_R, and (RM3) CP_RM3_F and CP_RM3_R. A PCR of each WT R-M system was performed as a control (~4.6 kB, ~4.0 kB and ~4.3 kB respectively). Colonies that have the R-M system deleted would yield a band at ~1.1 kB. A 2-log DNA ladder was used to determine the size of each band.

After producing the double and triple mutants of the R-M systems, the plasmid pMTL61110 was transformed into all the strains, following the method detailed in section 5.2.1.1. As previously, for each transformation two dilutions, specifically 1 x 10^{-2}, 1 x 10^{-1}, and an undiluted sample were plated onto TSA km and grown O/N (52 °C). The colonies of each plate were manually counted to determine the transformation efficiency of each sample (Figure 5.6).

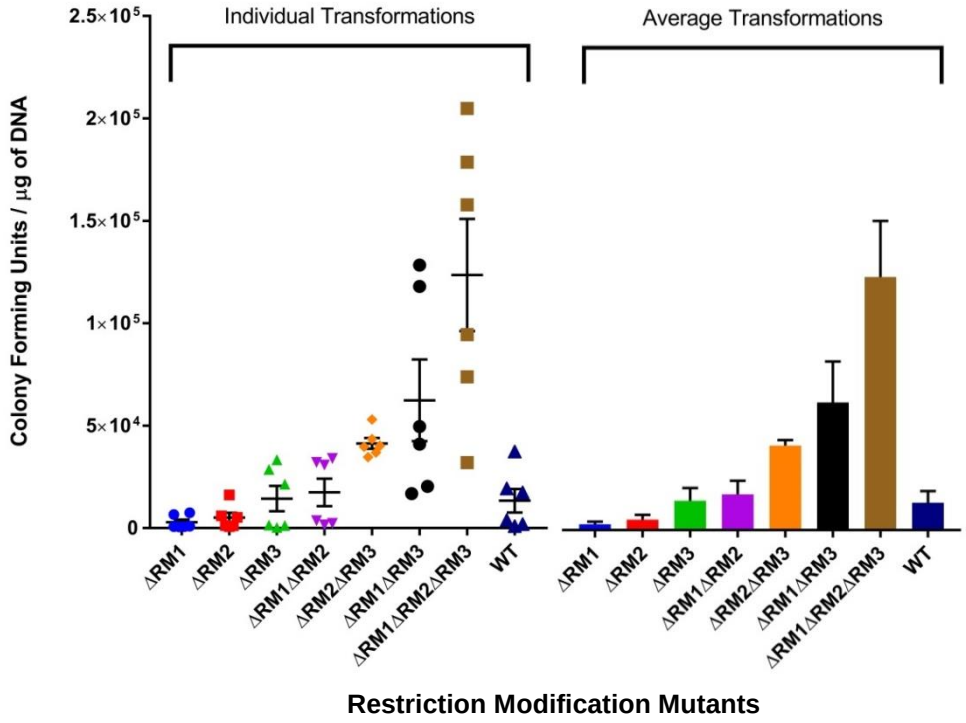


Figure 5.6 – The vector pMTL61110 was transformed into competent cells of each strain, alongside a WT control. The colonies produced on each plate were counted to determine the average variable transformants per µg of DNA for each mutant. The left side of the graph displays each individual transformation. The right side of the graph represents the average number of colony forming units manually counted for each mutant. The error bar represents standard error of the mean (n = 6).

The variation in transformation is demonstrated to be generally high amongst the majority of the R-M mutants, with the exception of ΔRM2ΔRM3, which actually exhibits increased consistency between replicates. A decrease in transformation efficiency is observed for the mutants ΔRM1 and ΔRM2 compared to WT; whereas no significant difference is apparent for ΔRM3 and ΔRM1ΔRM2. The mutants ΔRM3 and ΔRM1ΔRM2, were the only ones to appear unaltered in both their transformation efficiency and reproducibility, compared to WT. The

mutants ΔRM2ΔRM3, ΔRM1ΔRM3 and ΔRM1ΔRM2ΔRM3 do appear to show a general increase in transformation efficiency, compared to WT. The triple R-M mutant, ΔRM1ΔRM2ΔRM3, showed the highest increase in transformation efficiency, with the highest single transformation producing over 2 x 10^5 CFUs per μg of DNA. Consistency of transformations was poor for this mutant, which produced a large range of CFUs across the replicates tested. However, the lowest transformation efficiency recorded was similar to the highest transformation of WT, confirming that this mutant did offer an improvement in efficiency over WT. Interestingly, for the majority of the strains, the replicates were split into distinct groups of transformation efficiencies, rather than being normally distributed around the mean.

5.3 Discussion

R-M systems are described as a form of primitive bacterial immunity. These systems often encode two proteins, or one protein with dual activity, and carry out digestion and methylation of DNA. Digestion of foreign DNA occurs through a REase domain, which recognises a specific recognition sequence. Similarly, the encoded methyltransferase (MTase) also recognises the same sequence, of which methylation inhibits digestion. This methylation allows the bacteria to distinguish its own DNA from foreign DNA to be digested. R-M systems are often a major barrier to genome engineering in new organisms. *P. thermoglucosidasius* contains three R-M systems of unknown function, which may be the cause of the previously observed low transformation efficiency of this species. Removal of the activity of these genes can be achieved through methylation of exogenous DNA, or by the removal of the R-M systems, which may lead to increased transformation efficiency. Currently, neither of these methods has been achieved in *P. thermoglucosidasius*.

This work describes the first successful deletion of three R-M genes in *P. thermoglucosidasius*. These genes were knocked-out using the previously developed CRISPR/Cas9 system, isolated from *S. thermophilus* (Chapter 4). Single, double and triples mutants were successfully produced. The results demonstrated that transformation efficiency was affected by the removal of the R-M genes. It also highlighted an observed variability between replicates for the transformation efficiency of both WT *P.*

thermoglucosidasius and the resulting R-M deficient strains. In addition, this work has demonstrated that genome modification by *stcas9-3* produces clean deletions without producing any clear Off-target effects.

The deletion of all three R-M systems within a single strain of *P. thermoglucosidasius* resulted in an increase in the average number of CFUs produced after transformation with the plasmid pMTL61110. It has been demonstrated that this mutant produces the highest transformation efficiency observed in *P. thermoglucosidasius* thus far. Previous experiments have demonstrated transformation efficiencies of between 1 x 10^3 – 1 x 10^4 CFUs per µg of DNA (51,118,180). This confirms that the R-M systems do appear to have some discriminatory effect against foreign DNA. The data produced suggests that the removal of RM3 appears to have the most significant effect on transformation efficiency, with all mutants containing ΔRM3 exhibiting an increase in CFUs. Overall, the increase in transformation efficiency observed following the removal of the R-M systems is expected and has been shown previously in many different bacterial species (314,321-324).

The generation of the single mutants resulted in a counterintuitive decrease in transformation efficiency. RM1 and RM3 have been suggested to be Type IIG R-M systems, which are a class of Type II R-M systems defined by the fusion of the MTase and the REase into one single polypeptide (69,325). However, it could be feasible that in *P. thermoglucosidasius* these two R-M systems are in fact separated MTase

and REase genes that are positioned on different loci but recognise the same restriction sequence (326). Therefore, it could be theorised that RM1 is an MTase that has the same recognition sequence as RM3, the REase, which would explain the reduction in transformation efficiency for the strain ΔRM1 and the increase in CFUs for the ΔRM3 mutant. This phenomenon has been observed in other species, such as *S. aureus* and *Lactococcus lactis*, which both possess R-M systems containing genes that are located separately (327,328).

It is also possible for R-M systems to display features not associated with a traditional immunity response. Transcription regulation of gene expression has been associated with R-M systems, through epigenetic methylation by associated MTases (317,318). *Helicobacter pylori* was shown to contain a large number of R-M systems, of which several retained no activity (329,330). These R-M systems displayed truncated REases, whilst still maintaining functional MTases, the deleting of which resulted in reduced growth (331). In addition, it has been previously observed, through transcriptomics, that these R-M systems are upregulated in strains within environments requiring pH regulation, such as the gut (332,333). Furthermore, the MTase, CcrM, from *Caulobacter crescentus* is involved in regulating the cell cycle, through the methylation of several key promoters (334). Indeed, gene regulation through MTases has been shown to occur in several bacteria including *E. coli* (335,336). The evolutionary pressures associated with R-M systems differ for each of the associate genes. REases are toxic in the absence of MTases and

therefore cannot change their recognition sequence. However, MTases are able to modify their recognition sequences and still maintain their activity, at no detrimental effect to the cell. MTases have the ability to be maintained after the loss of REase, an event that cannot occur in the reverse, and therefore MTases can evolve to exhibit a broad spectrum of activities. Therefore, it could be suggested that the Type IIG R-M systems removed from *P. thermoglucosidasius* within this study, may have functionally redundant REase domains and only retain the MTase function of which the removal would lead to a reduction in transformation efficiency, as was observed.

In addition, R-M systems in general display similar qualities to that of toxin-antitoxin systems. Both these systems exhibit characteristics such as encoding a form of stable toxin and a labile antitoxin and both are associated with mobile elements and exhibit selfish behaviour (65). Similar to toxin-antitoxin systems, it has also been shown that R-M systems have the added ability to stabilise neighbouring regions on a genome (275), such as the toxin-antitoxin system that stabilises the plasmid pNCI001, as demonstrated by the work in chapter 4. Therefore, it can be suggested that the R-M systems investigated within this chapter could also be exhibiting a similar role, explaining the reduction in transformation efficiency. However, as transformation efficiency does appear to increase with the simultaneous deletion of all three R-M systems, it is more likely that the R-M systems either lack a functional REase or that RM1 and RM3 are a single system separated on the genome.

With the exception of ΔRM2ΔRM3, variation between replicates did not improve with the deletion of the R-M systems. This indicates that other factors are involved in the determination of transformation efficiency. Often, a major factor affecting transformation efficiency is the competency of cells (296,316). This can be affected during cell preparation, as often the growth phase of the bacteria is important, with the cells typically required to be in the exponential phase of growth for the optimum transformation. This is shown through observed erratic competency when cells are grown to either saturation or during nutrient limitation (316). The current in-house protocol for transformation of *P. thermoglucosidasius* is relatively broad, indicating a range of growth conditions for both the preparation and recovery of cells, which may introduce error into the competency of the cells. The inconsistent transformation efficiencies observed may also point to a lack of an optimised growth protocol.

No clear Off-target effects can be extrapolated from the genome sequences obtained. The SNPs that were observed within the genomes of each mutant did not correlate to sequences that resembled PAM sites, suggesting no link to the production of the mutants. However, Off-target effects have been observed within human cells using spCas9, which demonstrated targeting of an alternative PAM sequence. Though the canonical sequence for the PAM of *S. pyogenes* is NGG, a noncanonical PAM, NRG (where R is G or A), has been identified (231,337). Nevertheless, it has also been suggested that *P. thermoglucosidasius* is

susceptible to SNPs, through the reproduction of the strain TM242 by Sheng *et al.* (118). This study demonstrated that LS242, the reproduction of TM242, did not produce ethanol at levels observed for TM242. Genome sequencing of TM242 indicated that NCIMB 11955 is prone to the emergence of random mutations, suggesting that the SNPs observed here may also be a product of this.

5.4 Conclusions and Future Work

This work describes the successful removal of three R-M systems, two Type IIG systems and one Type III system, from *P. thermoglucosidasius* using CRISPR/Cas9 technology. The initial data has demonstrated that an increase in transformation efficiency is achievable, recording 1 x 10^5 CFUs for the triple mutant strain, which is an order of magnitude higher than any transformation previously observed. This increase in transformation efficiency should therefore enable an increase in the efficiency of the CRISPR/Cas9 method, allowing greater efficiency of AE and the number of transformants produced. The data produced has confirmed the use of *stcas9-3* as a specific and efficient genome editing tool, demonstrating the production of clean mutants with no clear Off-target effects. This study has also highlighted the observation of an unwanted variation in CFUs produced between replicates that appear to occur in *P. thermoglucosidasius*, both in WT and in many of the tested mutant strains. For future investigations, the primary focus should be to tackle this inconsistency, which may be improved through optimising the transformation protocol.

Another key next step would be to conduct phenotypic analysis on the R-M system mutants, in order to determine whether they have any metabolic or transcriptional effect. Many studies have suggested a potential fitness cost to encoding R-M systems (338-341). For example, a phenomenon called restriction site avoidance can be observed in bacteria,

whereby the particular recognition sequence has a lower frequency on the genome to protect the host bacteria from attacking itself. This avoidance can be due to the occurrence of mutations that could inadvertently affect fitness and other cellular functions (342). This restriction site avoidance has also been observed through the highly susceptible deamination of 5-methyl-cytosine, as often MTases modify the genome through methylation of either adenine or cytosine. An increased mutation frequency of C – T is observed and the organism may therefore incur a potential fitness cost. This phenomenon is even more prevalent within thermophiles, as the frequency of deamination of 5-methyl-cytosine is considerably higher (343). Furthermore, some R-M systems utilise ATP to allow for translocation along the foreign DNA. This consumption of energy means that the enzyme concentration is controlled intracellularly and therefore not always active (339). Therefore, future work should explore the effect of deleting the R-M systems on the growth of *P. thermoglucosidasius*.

5.5 Key Outcomes

- Production of an R-M system deficient mutant, ΔRM1ΔRM2ΔRM3, that exhibited an increased transformation efficiency, >1 x 10^5 CFUs, greater than any transformation observed thus far.

- Identification of the variation in CFUs for the transformation of WT *P. thermoglucosidasius*.

- Confirmation that *stcas9-3* produces no Off-target effects and can be confidently utilised as a genome editing tool in *P. thermoglucosidasius*.

Chapter 6

A View Towards Metabolic Engineering of *P. thermoglucosidasius* for Improved Succinic Acid Production

6.1 Introduction

Platform chemicals such as succinic acid, lactic acid, 3-HP and ethanol are essential to the production of many important materials, but are still largely produced by the petrochemical industry through fossil fuel use. Therefore, using organisms such as *P. thermoglucosidasius* to metabolise biomass, offers the potential to produce high value platform chemicals from carbon-neutral, renewable sources. These chemicals are often produced through the mixed acid fermentation pathway, which is a potential target for genetic engineering using CRISPR/Cas9 technology.

6.1.1 Mixed Acid Fermentation

Mixed acid fermentation is the biological process whereby glucose is broken down into a variety of acids during anaerobic digestion, a common process within many types of bacteria. The partial oxidation of glucose generates the production of pyruvate, ATP, and nicotinamide adenine dinucleotide (NADH). In the absence of oxygen as an electron acceptor, pyruvate is used to regenerate NAD^+, which produces waste products such as lactic acid, acetic acid, succinic acid, and ethanol (141) (Figure 6.1).

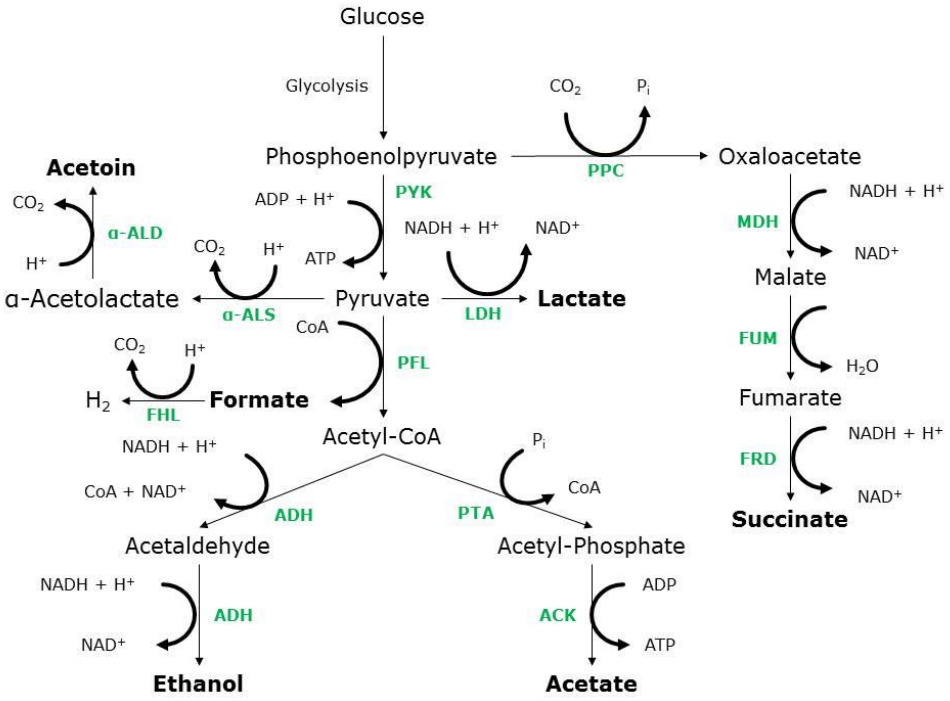


Figure 6.1 – A schematic representation of the mixed acid fermentation pathway showing the reoxidation of NAD^+ from NADH, using pyruvate as the final electron acceptor to produce a variety of acids.

The concentrations of these waste products can be altered through beneficial genetic engineering of the metabolic pathways. There are a variety of specialised species of bacteria that only produce one specific organic acid as an end-product. Examples include, *A. succinogenes* (344) and *Acetobacter aceti* (345), which produce succinic acid and acetic acid respectively, in addition to *Lactobacillales*, an entire order of lactic acid producing bacteria (346). These specialised bacteria are hugely diverse in their metabolite profiles and also operate under a range of different conditions, from microaerophilic fermenting microorganisms to strict anaerobic microorganisms (347).

6.1.2 Succinic Acid Production

As previously discussed in section 1.6, succinic acid is an industrially important, high value, platform chemical, that can be used to produce many useful products for industries such as pharmaceutical, agricultural, and food produce. Furthermore, succinic acid can also be used as an intermediate for many other important chemicals, such as 1,4-butanediol and tetrahydrofuran (27,138,140) which are commonly used in the chemical industry as solvents and polymer-precursors. The importance of this chemical and its future potential has led to it being identified by the U.S. Department of Energy as one of twelve high value products targeted for biomass production (137,139). This designation has resulted in increased research into succinic acid production using microorganisms (31,348,349).

There are many advantages to using thermophiles as whole cell biocatalysts to produce high value products from waste, discussed in more detail in section 1.2.5. However, currently, no thermophiles have been employed to produce succinic acid. Platform chemical production in general is lacking in thermophiles, largely due to limited tool availability (49). Metabolic engineering in thermophiles has to date been mainly focused on the production of ethanol (24,350,351). However, tools for use in thermophiles are continually improving and it is clear that research into the production of platform chemicals from these species is an area of great interest (352).

6.1.3 Chapter Aims

The ultimate aim of this work is to investigate the potential of *P. thermoglucosidasius* for the production of important platform chemicals via carbon-neutral, renewable sources. The mixed acid fermentation pathway is crucial to the production of many metabolites and therefore is a key target to genetically manipulate. Using the CRISPR/Cas9 technology, and the tools developed within the first two chapters, the aim is therefore to target this pathway within *P. thermoglucosidasius.* This chapter aims to produce initial data to demonstrate the viability of this and to start to build a library of KO vectors, which can be utilised in future investigations.

6.2 Results

6.2.1 Using CRISPR/Cas9 for Genetic Engineering for Strain Development

After developing the CRISPR/Cas9 technology in *P. thermoglucosidasius*, it was decided that strain development should be a priority. Initial mutations were made through the knock-out of *ldh*, aldehyde-alcohol dehydrogenase (*adhE*) enzymes and the cytochrome C oxidase (*cox*) complex. For each gene, a CRSIPR/Cas9 KO vector was produced using *stcas9-3* and used to KO each selected gene to produce a clean mutant, following the developed method detailed in section 4.2.2. Each strain was then screened to confirm both the production of a clean mutation, shown in figure 6.2, and the subsequent successful loss of the plasmid.

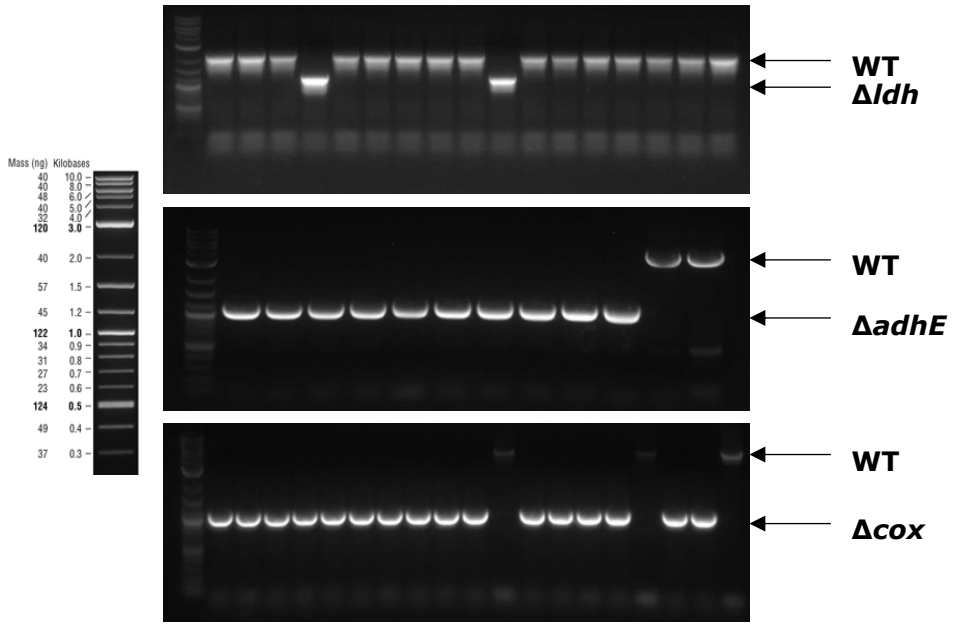


Figure 6.2 – A colony PCR was performed to assess whether the genes encoding lactate dehydrogenase (*ldh*), aldehyde-alcohol dehydrogenase (*adhE*) and cytochrome C oxidase (*cox*), had been knocked out of the *P. thermoglucosidasius* genome. The primers used were (*ldh*) CP_LDH_F and CP_LDH_R, (*adhE*) CP_ADH_F and CP_ADH_R, and (*cox*) CP_COX_F and CP_COX_R. A PCR of each WT gene was performed as a control (~2.1 kB, ~3.7 kB and ~5.0 kB respectively). Colonies that had the successful gene deletion would yield a band at ~1.1 kB. A 2-log DNA ladder was used to determine the size of each band.

The colony PCR shows that clear differences were observed in the number of KOs achieved for each gene. Both Δ*cox* and Δ*adhE* demonstrated a high success rate, with 16/18 and 10/11 colonies screened showing successful KOs respectively. Conversely, the KO rate for Δ*ldh* was significantly lower, with only 2/16 colonies displaying the KO. In addition to the KO colony PCR, internal primers for each gene were designed, to further confirm that the genes were removed from the genome. A PCR of each mutant clearly demonstrated that the gene had been deleted, as no band was observed in each strain (Figure 6.3).

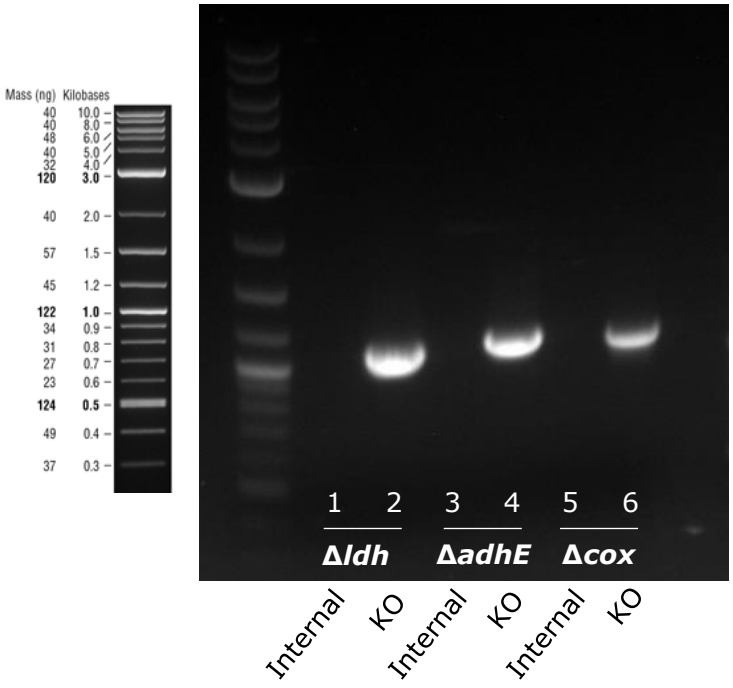


Figure 6.3 – A colony PCR was performed to confirm loss of the genes, lactate dehydrogenase (*ldh*), aldehyde-alcohol dehydrogenase (*adhE*) and cytochrome C oxidase (*cox*), from the *P. thermoglucosidasius* genome. The internal gene primers were (*ldh*) LDH_Inter_F LDH_Inter_R, (*adhE*) ADH_Inter_F ADH_Inter_R and (*cox*) Cox_Inter_F and Cox_Inter_R. Amplification of the internal primers yields bands at ~0.9 kb, ~1.3 kb and ~1.2 kb respectively. The primers (*ldh*) CP_LDH_F and CP_LDH_R, (*adhE*) CP_ADH_F and CP_ADH_R, and (*cox*) CP_COX_F and CP_COX_R were used as controls.

Each mutant was subsequently grown in ASYE media, supplemented with 1% glucose under both aerobic (Ae) and micro-aerobic (MA) conditions. Ae growth was achieved using a baffled conical flask (250 ml) containing media (50 mL), for 8 hours at 250 RPM. MA growth was achieved using a parafilm-sealed falcon tube, containing media (10 mL), for 24 hours at 250 RPM. Following growth, the supernatant of each sample was collected and prepared for HPLC-UV analysis, to determine the metabolite profile of each mutant. The concentrations of the metabolites in each sample were measure in mM and are displayed in figure 6.4.

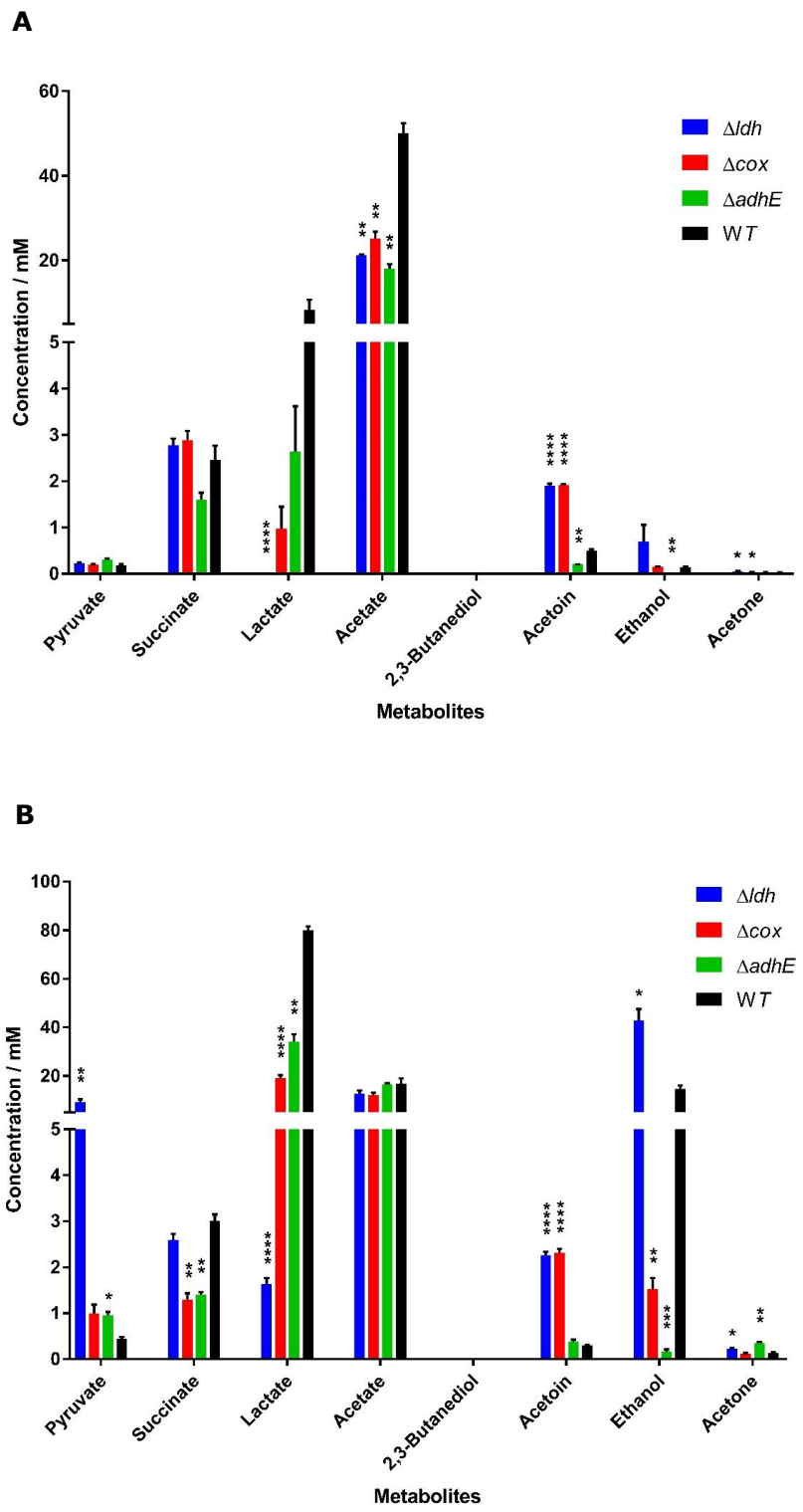


Figure 6.4 – The concentration of metabolites (mM) produced by the strains Δ*ldh*, Δ*cox*, Δ*adhE* and WT *P. thermoglucosidasius*, as measured by HPLC-UV. All strains were grown under both aerobic (A) and micro-aerobic (B) conditions to determine the difference in metabolite production. The error bars on the graph denote the standard error of the mean (n = 3). *, **, *** and **** denote levels of statistical significance from the negative

control equal to the p values, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$, respectively, as determined by a one-way analysis of variance (ANOVA) assuming unequal variances.

It was apparent that, for both growth conditions, all three strains exhibited clear differences in their metabolite profile compared to WT. When grown under MA conditions, an increase of pyruvate was observed for all strains compared to the WT, with the most dramatic increase seen for the Δ*ldh* strain, which exhibited a > 20-fold increase. However, under Ae conditions the strains were comparable to WT. There was no noticeable increase for any of the strains over the WT in succinate production, under both growth conditions. However, the Δ*adhE* strain showed a decrease in production for both conditions, as did Δ*cox* under MA growth. As expected, the Δ*ldh* strain saw a decrease in lactate production compared to WT for both growth conditions, completely eliminating this metabolite under Ae conditions. However, the other two strains also showed a decrease in lactate production compared to the WT across both growth conditions. Under MA growth, the acetate production was comparable between all three strains and the WT. However, under Ae conditions, all strains exhibited a similar reduction in acetate production compared to the WT. The production of acetoin showed and improvement for both the Δ*ldh* and Δ*cox* strains, under both the Ae and MA growth conditions, with levels of acetate that were 7-fold and 12-fold higher than the WT, respectively. During Ae growth, limited ethanol production in the WT was observed. An equivalent low level of ethanol was obtained with all of the mutant strains with the exception of the Δ*ldh* mutant which

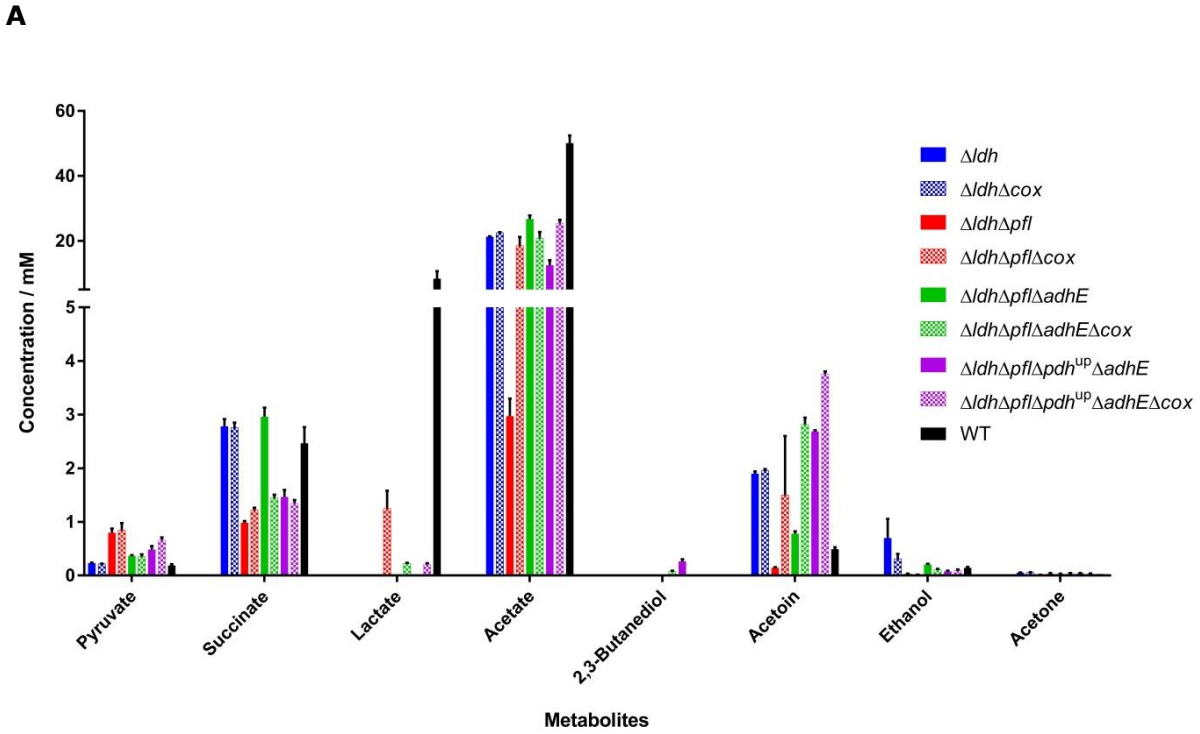
produced marginally more ethanol. Conversely, comparatively high levels of ethanol were produced by the WT strain under MA conditions which were exceeded by the Δ*ldh* strain. Both Δ*adhE* and Δ*cox* produced significantly lower amounts of ethanol compared to WT under MA conditions. Neither the WT nor the three strains produced measurable amounts of 2,3-butanediol and the levels of acetone produced were also minimal.

6.2.2 Effect of Cytochrome C Oxidase Deletion on Mixed Acid Fermentation

The products of mixed acid fermentation detected within this study can be changed through genetic manipulation of specific genes, as previously discussed. Multiple strains have been produced in-house which contain a variety of gene deletions and upregulations of genes encoding the pivotal enzymes of the mixed acid fermentation pathway. These strains include Δ*ldh*Δ*pfl*, Δ*ldh*Δ*pfl*Δ*adhE* and Δ*ldh*Δ*pfl*Δ*pdh*^{up}Δ*adhE* (*pfl*, pyruvate formate lyase; *pdh*^{up}, pyruvate dehydrogenase upregulated).

In collaboration with Oxford Brookes University (David Fell), a computer model of *P. thermoglucosidasius* metabolism was used to model succinic acid production. The model indicated numerous deletions that could be carried out to deliver a potential increase in succinic acid production. One such deletion was that of the *cox* complex, an electron transport chain,

membrane bound protein, which aids in the conversion of oxygen to water and has been knocked-out previously in other microbes (section 6.2.1). Therefore, it was decided that the *cox* CRISPR/Cas9 KO plasmid would be used to KO the *cox* complex in electrocompetent *P. thermoglucosidasius* strains, those detailed above in addition to Δ*ldh*. Mutants were produced, screened and the plasmid lost as previously described in section 6.2.1. The mutants were then grown in both Ae and MA conditions and HPLC-UV samples were prepared from the supernatant, as described previously (Section 6.2.1). Results from the metabolite profile analysis are displayed in figure 6.5.


A
Concentration / mM
60
40
20
5
4
3
2
1
0
Pyruvate
Succinate
Lactate
Acetate
2,3-Butanediol
Acetoin
Ethanol
Acetone
Metabolites
Δldh
Δldh∆cox
Δldh∆pfl
Δldh∆pfl∆cox
Δldh∆pfl∆adhE
Δldh∆pfl∆adhE∆cox
Δldh∆pfl∆pdh^up∆adhE
Δldh∆pfl∆pdh^up∆adhE∆cox
WT

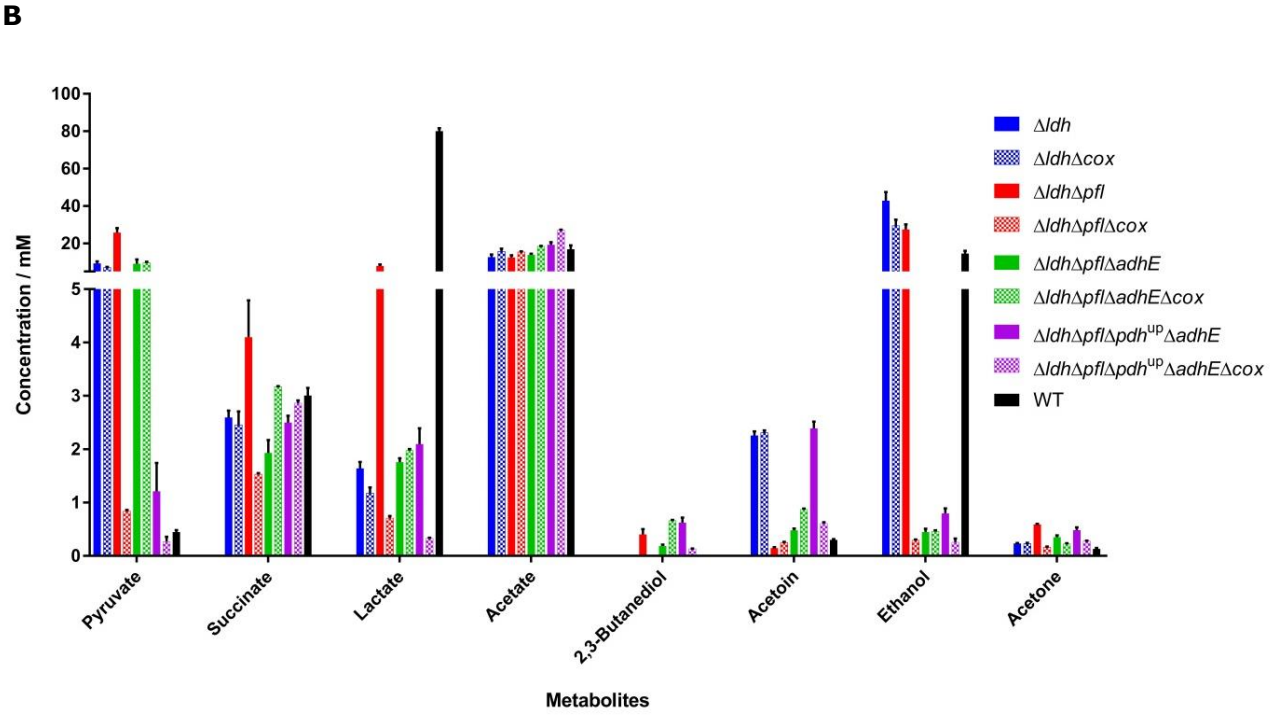


Figure 6.5 – The concentration of metabolites (mM) produced by the strains with and without the Δ*cox* mutation as measured by HPLC-UV. The strains tested, with and without their Δ*cox* counterpart, are Δ*ldh*, Δ*ldh*Δ*pfl*, Δ*ldh*Δ*pfl*Δ*adhE* and Δ*ldh*Δ*pfl*Δ*pdh*^{up}Δ*adhE* compared to WT *P. thermoglucosidasius*. All strains were grown under both (A) aerobic and (B) micro-aerobic conditions to determine the difference in metabolite production. The error bars on the graph denote the standard error of the mean (n = 3).

In general, the addition of a Δ*cox* mutation to each strain did not result in a noticeable increase in the metabolite profile, with some significant decreases in metabolite production observed. However, there were a few exceptions that did display an improvement. Under MA conditions, Δ*ldh*Δ*pfl*Δ*adhE*Δ*cox* exhibited improved levels of succinate production, yet only producing levels comparable to the WT. However, both 2,3-Butanediol and acetoin production was increased compared to both the *cox*-containing counterpart and the WT. Surprisingly, in the case of a Δ*ldh* mutant, the addition of the Δ*cox* mutation returned the production of lactate in three strains under Ae conditions, most notably Δ*ldh*Δ*pfl*Δ*cox*. However, this was still at a considerably reduced level compared to WT. Under Ae conditions there was also a notable increase in acetoin production, over both the cox-containing equivalents and the WT, for all strains except Δ*ldh*Δ*cox*. The most significant increase was observed for the strain Δ*ldh*Δ*pfl*Δ*pdh*^{up}Δ*adhE*Δ*cox*, which exhibited a 1.4-fold increase compared to Δ*ldh*Δ*pfl*Δ*pdh*^{up}Δ*adhE* and an improvement of > 7-fold compared to WT.

6.2.3 Effect of Succinate Dehydrogenase Deletion on Succinic Acid Production

Succinate dehydrogenase (Sdh, encoded by the *sdh* gene) is an enzyme present in the TCA cycle that converts succinate into fumarate. It was decided to attempt to KO *sdh* to observe if this would increase production of succinate. Both a single mutant, Δ*sdh*, and a double mutant Δ*ldh*Δ*sdh* were produced to observe the effects on metabolite production. Clean mutants were successfully produced and the plasmid lost, confirmed by PCR (Figure 6.6), as previously described in section 6.2.1. The mutants were then grown in both Ae and MA conditions and HPLC-UV analysis was conducted on samples prepared from the supernatant. Results from the metabolite profile analysis are displayed in figure 6.7.

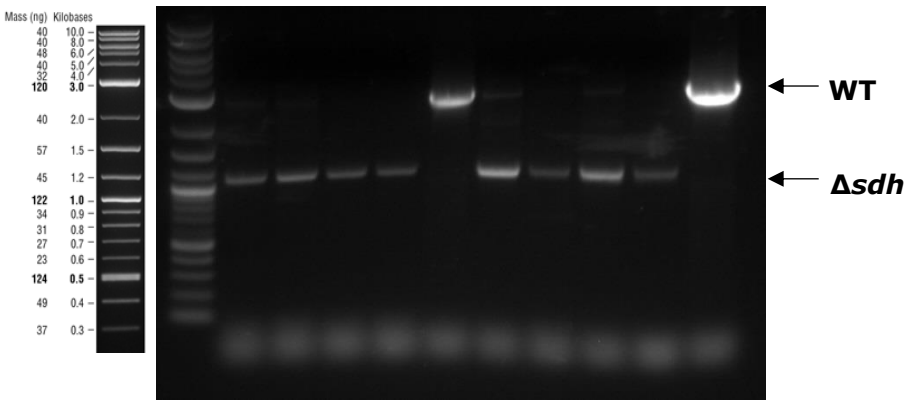


Figure 6.6 – A colony PCR was performed to assess whether the gene encoding succinate dehydrogenase (*sdh*), had been knocked out of the *P. thermoglucosidasius* genome. The primers used were CP_SDH_F and CP_SDH_R. A PCR of the WT gene was performed as a control (~2.9 kB). Colonies that have the gene deleted would yield a band at ~1.1 kB. A 2-log DNA ladder was used to determine the size of each band.

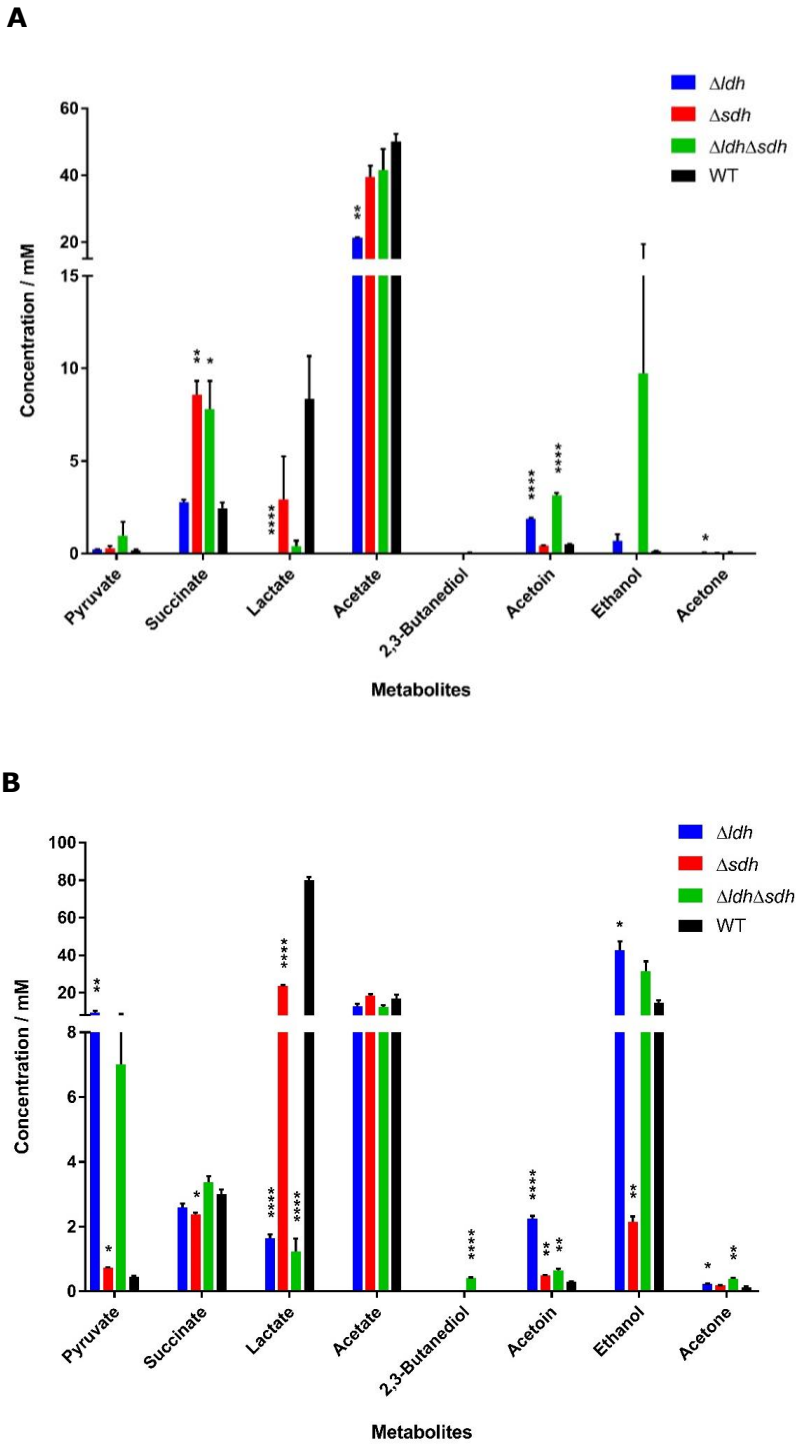


Figure 6.7 – The concentration of metabolites (mM) produced by the strains Δ*ldh*, Δ*sdh*, Δ*ldh*Δ*sdh* and WT *P. thermoglucosidasius* as measured by HPLC-UV. All strains were grown under both (A) aerobic and (B) micro-aerobic conditions to determine the difference in metabolite production. The error bars on the graph denote the standard error of the mean (n = 3). *, **, *** and **** denote levels of statistical significance from the negative control equal to the p values, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$,

respectively, as determined by a one-way analysis of variance (ANOVA) assuming unequal variances.

Under MA conditions the Δ*sdh* mutation alone did not have a positive effect on the production of any metabolites compared to the WT. The addition of this mutation to the Δ*ldh* strain did not demonstrate any significant advantages, except to afford a small increase in 2,3-butanediol production. Conversely, during aerobic growth the Δ*sdh* mutation produced some noticeable changes to metabolite production. For both the Δ*sdh* and Δ*ldh*Δ*sdh* strain, succinate production was significantly increased, with both strains exhibiting a 3.5-fold compared to WT. One unusual, yet interesting result observed for the Δ*ldh*Δ*sdh* strain is that it improved ethanol production significantly, with one colony producing almost 30 mM, a >200-fold increase compared to the WT.

Of all the strains tested, the Δ*sdh* and Δ*ldh*Δ*sdh* strains were the highest producers of succinate (Figure 6.8). Besides these two strains, the other strains tested exhibited reduced production, or at best equivalent to, the levels seen in the WT strain.

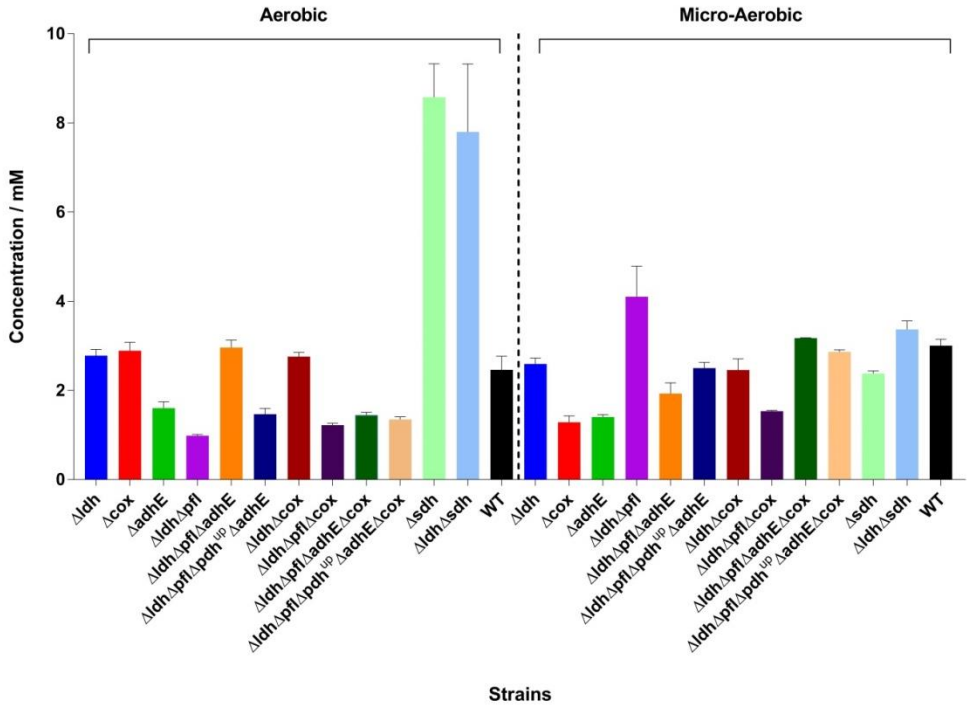


Figure 6.8 – The concentration of succinate produced (mM) by all the strains constructed in this study, as determined by HPLC-UV. All strains were grown under both aerobic and micro-aerobic conditions to observe the differences in succinate production. The error bars on the graph denote the standard error of the mean (n = 3).

6.3 Discussion

This chapter details an investigation into the potential of *P. thermoglucosidasius* for the production of important platform chemicals. The metabolism of biomass by microorganisms is already being utilised as a renewable, carbon neutral source. However, currently both production and research are focused on the use of mesophiles and so the advantages that thermophiles offer are not being exploited. The mixed acid fermentation pathway is a crucial target for genetic manipulation and so using CRISPR/Cas9 technology initial mutations within this pathway have been explored within *P. thermoglucosidasius.*

This work describes the successful deletion of the *sdh* gene which resulted in a 3.5-fold increase in succinate production compared to the WT. This strain can now be used as a baseline for the creation of strains with improved succinate production through further mutational changes to genes encoding enzymes involved in the glyoxylate pathway. The deletion of *sdh* was carried out using *stcas9-3*, further demonstrating its usefulness as a genome editing tool in *P. thermoglucosidasius*. Previous deletion of the equivalent gene in other bacterial species has met with variable success. In those cases where KO of the gene has not been achieved, the possibility exists that it is an essential gene (353,354). The role of Sdh in the central carbon metabolism pathway, specifically the TCA cycle, is to reversibly catalyse the oxidation of succinate to fumarate. Nevertheless, successful deletion of *sdh* has been observed with an aim to increase succinate

production (143,355). This is due to the fact that although it can convert fumarate to succinate, the enzyme has been shown to preferentially converts succinate to fumarate at a ratio of 25:1 in *E. coli* (356).

One interesting observation relates to the fact that of the three Δ*sdh* biological replicates, one produced abnormally high levels of ethanol compared to the others. The levels obtained represented the highest ethanol levels obtained under Ae respiration of any of the strains tested. Normally, under Ae conditions, oxygen utilisation enables the complete breakdown of glucose and therefore ethanol production is not required. Hence, further investigation needs to be conducted to determine whether this was an artefact of the data or an actual phenomenon. If this strain consistently produces high levels of ethanol under Ae conditions, sequencing can be carried out to determine the underlying reason behind this. As discussed in chapter 5, it has been suggested that *P. thermoglucosidasius* is prone to SNPs and, as with the study regarding TM242, this can result in increased ethanol production due to unintended mutations (118). SNPs were also identified in the creation of the restriction modification mutants, detailed in chapter 5. However, it was unconfirmed whether this was due to Off-target effects produced by using the CRISPR/Cas9 system, which also may be occurring with this strain.

In order to fully utilise the efficiency of CRISPR/Cas9 technology, it is important to have pre-constructed libraries of gene editing plasmids available for use, to facilitate rapid, high-throughput strain generation.

However, currently these libraries do not exist for *P. thermoglucosidasius.* The work in this chapter describes the initiation of a library of CRISPR/Cas9 KO vectors designed to alter the metabolite production of the mixed acid fermentation pathway. The specific targets identified to KO were the genes coding for *ldh*, *adhE* and the cytochrome C oxidase (*cox*) complex. The central involvement of these enzymes in the mixed acid fermentation pathway have been either suggested or shown to affect the carbon flux and the termination or enhancement of their synthesis can be used to improve production of a desired metabolite (150,357,358). Importantly, the KO of *ldh* has already been shown to improve ethanol production in *P. thermoglucosidasius* (24,118).

The data produced in this chapter also highlights a variability in KO efficiency regarding the use of CRISPR/Cas9 for the three genes tested. Mutation of *ldh* produced 2/16 successful mutants, in contrast to the more efficient mutations of *cox* and *adhE*, which produced 16/18 and 10/11 successful mutants respectively. This variation can possibly be explained through the difference in the On-target scores of each gene. For each gene selected the highest three On-target scores were 61.5, 55.6 and 30.8 for *ldh*, 92.2, 76.7 and 66.9 for *cox* and 96.8, 54.9 and 48.2 for *adhE*. The On-target score is the specificity of the sgRNA to target its marker (359) and therefore, the lower the On-target score, the lower the efficiency of deletion, an observation confirmed by the KOs produced in this work. This has also been corroborated through previous studies, which have shown that there is a significant variation between the cutting efficiency of

different sgRNAs (359-362). However, these On-target scores are not considered to be a comprehensive predictor of efficiency. The context of the DNA at the specific location, such as the secondary structure or methylation pattern (363,364), may also play a factor in the ability of the sgRNA to bind to the DNA, a feature which is currently not considered in most models. Furthermore, it has been suggested that the structure of the sgRNA itself may play a role in the efficiency of KOs (363). For example, it has been suggested that the secondary structure and internal interactions of the sgRNA may interfere with *cas9* interactions, with Thyme *et al.* indicating that hair-pin loops cause inhibition (365). Another consideration is that the optimum temperature for *P. thermoglucosidasius* growth may have an effect on the structure of the sgRNA and therefore the ability to successfully guide *cas9*. It has been shown that the free energy of the sgRNA does indeed positively correlate with KO efficiency (366). This whole area of sgRNA On- and Off-targeting is currently under heavy investigation, due to the importance of gene editing at precise locations, which is especially crucial within medical CRISPR/Cas9 applications (359-363,367,368).

One final consideration with regards to gene KO efficiency, is that *P. thermoglucosidasius* has been shown to be prone to mutations (118). The specificity of CRISPR/Cas9 may cause point mutations, which would lead to resistance. In future work, it would be useful to sequence the region of non-mutant transformants, in order to observe any SNPs.

Previous work (unpublished) by Mougiakos *et al.* has shown translocation of genes, within *B. smithii*, when targeted using *cas9* from *S. pyogenes* (369). Therefore, in order to conclusively confirm the successful gene KOs, the whole genome was screened for a potential translocation event. This could potentially occur through transposons, a common method of horizontal gene transfer (370,371). However, through an internal PCR of the genes deleted, it was confirmed that no translocation event had occurred and the genes had been successfully deleted.

Following KO of the genes, the metabolite profile of the resulting strains was determined using HPLC-UV. The strains were grown in both Ae and MA conditions to observe the differences in the metabolites produced. All strains exhibited a different metabolic profile from each other and also demonstrated clear differences from WT *P. thermoglucosidasius*, which further confirms successful deletion of the target genes. The *cox* complex is an electron transport chain protein involved in the Ae production of ATP (372), which suggests that the loss of *cox* should affect Ae growth. During Ae conditions, the *cox* mutation led to a reduction in the production of lactate and acetate compared to the WT; whereas, no increase or decrease was observed for pyruvate, succinate and ethanol. The reduction of acetate and lactate under Ae conditions may be caused by the reduction of ATP production, due to the disruption of the electron transport chain. Therefore, it is conceivable that an increase in the carbon flux is needed to produce the same amount of ATP through the Ae pathways. This could suggest a reduction in carbon overflow, leading to a

reduction in acetate and lactate. However, similar to Ae conditions, inactivation of *cox* largely resulted in a reduction or maintenance of metabolite concentration compared to the WT, suggesting that oxidation of NADH is not occurring through the reduction of pyruvate to produce the metabolites tested.

Under both Ae and MA conditions, only acetoin was increased in the Δ*cox* strain compared to WT. Acetoin is a neutralising molecule that is often produced in response to over acidification of the cytoplasm and the surrounding media, by accumulation of acidic compounds such as lactic acid and TCA cycle intermediates (373). Although an overall reduction was observed for lactate and acetate, there are numerous other acidic metabolites that have not been tested. Furthermore, acetoin can be utilised as an energy store, by conversion into acetyl-CoA by acetoin dehydrogenase (373). Therefore, the metabolic stress caused by the *cox* deletion may have activated a survival response to produce an external energy store. This is also suggested by the lack of 2,3-butanediol, which can be produced from acetoin, which is an intermediate of the pathway producing 2,3-butanediol (374,375). In addition, the metabolic stress, caused by the *cox* deletion, could be responsible for the overall reduction in metabolic products, under both Ae and MA conditions. Therefore, to further understand this, a phenotypic analysis of the strain should be considered. In general, bacterial electron transport chains are exceedingly complex and contain large, diverse pathways, multiple possible electron acceptors and can adapt to environmental change (376). Therefore,

further investigation is ultimately needed to fully understand these metabolic changes.

The enzyme AdhE is responsible for catalysing the conversion of acetyl-CoA to acetaldehyde and also that of acetaldehyde to ethanol. In addition, AdhE has been shown to regulate pyruvate formate lyase (Pfl, encoded by the *pfl* gene) activity. Thus, this enzyme acts as an alcohol dehydrogenase (377), a CoA-dependent acetaldehyde dehydrogenase (378,379). For the metabolites tested in this study, the Δ*adhE* mutant produced either equivalent or reduced levels of metabolites compared to the WT, under both growth conditions. A reduction in lactate and acetate under Ae conditions is consistent with the deletion of *adhE* as the encoded enzyme is known in some microorganisms to play a role in their generation via the conversion of acetaldehyde (381). Furthermore, AdhE can also catalyse the reverse reaction and convert ethanol to aldehyde, which is then converted to acetyl-CoA. This then feeds into the TCA cycle, which could account for the reduction in succinate that is observed under Ae conditions (382).

Ldh is a common enzyme, found in both eukaryotes and prokaryotes, which bidirectionally catalyses the reaction between pyruvate and lactate. Under Ae conditions, lactate production was completely eliminated in the Δ*ldh* mutant strain, as expected. However, the production of ethanol was increased in this mutant under Ae conditions, a result which is not expected under Ae conditions. The effect of the *ldh* deletion may lead to

carbon overflow proceeding through different pathways, hence the production of ethanol. The reduction in other metabolites may be due to the cross-linking of many different pathways, for example, lactate feeds into propanoate metabolism. Therefore, a reduction in lactate would negatively affect the metabolites produced by this pathway. Alternative routes into propanoate metabolism are therefore needed to compensate for this reduction in lactate. Acetyl-CoA also feeds into propanoate and therefore this pathway may be used to compensate for the consequences of the *ldh* deletion. This would then affect any other products derived from by acetyl-CoA, which may explain the reduction in acetate. Under MA conditions, the Δ*ldh* mutant produces increased concentrations of pyruvate, which is expected as without Ldh the cell cannot convert pyruvate into lactate. An increase in ethanol production was also observed for this strain. This lack of pyruvate conversion leads to a bottleneck and to compensate, pyruvate is converted to ethanol, a result also observed by Cripps *et al.* (24) in *P. thermoglucosidasius*.

Acetoin was also increased compared to the WT, under both growth conditions, which may be explained as an indirect effect of the deletion of *ldh*. In addition to the small number of metabolites tested, a multitude of others may have been affected by the genetic modifications made. Changes to these effected metabolites may then exert a downstream effect on the metabolites tested in this study. In the case of acetoin, the *ldh* deletion may have caused an increase in the levels of acetolactate, as this is produced from two molecules of pyruvate, which was shown to be

increased for this strain. Acetolactate can then be decarboxylated to produce acetoin (383). As deletions affect the carbon flux, the metabolite profile maybe affected in ways that are difficult to predict, with so many different enzymes affecting this central system. Nevertheless, this study has further confirmed *stcas9-3* as a simple and effective way to produce gene deletions in *P. thermoglucosidasius*.

Using the CRISPR/cas9 KO technology, *cox* mutations were applied to a series of in-house strains that already contained mutations in the mixed acid fermentation pathway. A computer model produced at Oxford Brookes University suggested that this *cox* mutation may lead to enhanced succinate production. For both growth conditions, the *cox* mutation generally resulted in no increase in metabolite concentration in comparison to the progenitor strains. Under Ae conditions, only one strain, Δ*ldh*Δ*pfl*Δ*adhE*, showed any change in succinate levels, yet this was a reduction. Interestingly, this strain is the only one to show an increase in succinate under MA conditions compared to its progenitor strain; however, this level of succinate was still lower than produced by WT *P. thermoglucosidasius*.

The fact that no strains demonstrated an increase in succinate production under Ae conditions was unexpected. It was hypothesised that Δ*cox* would cause the strain to act as if under MA conditions, in which the activity of the electron transport chain is lower, and hence produce increased metabolite concentrations. This does appear to be the case for acetoin

production, as this metabolite is increased in all of the *cox* deficient strains compared to their progenitor during Ae conditions. In some organisms, glucose is converted using the butanediol pathway, of which acetoin is an intermediate. Therefore, it can be suggested that the *cox* deletion is forcing carbon flux down this pathway, producing higher levels of acetoin, as opposed to through the TCA cycle, which is normal for Ae respiration (374). Alternatively, as discussed earlier, acetoin production may be linked to the metabolic stress caused by the Δ*cox* mutation. Interestingly, lactate production in the *cox* deficient mutants, under Ae conditions, increased compared to the progenitor strains. It is plausible that the acetyl-CoA, which is produced from pyruvate, is shuttled towards lactate production instead of being use in the TCA cycle. However, this increase was still significantly lower compared to the concentration of lactate produced by WT.

6.4 Conclusions and Future Work

This work describes the successful utilisation of CRISPR/Cas9 technology to genetically engineer *P. thermoglucosidasius* through the deletion of multiple genes. It demonstrated the successful generation of a Sdh deficient *P. thermoglucosidasius* strain, which displayed increased production of succinate. Furthermore, variances in the KO efficiency of different genes has been highlighted for further investigation. This chapter has also demonstrated that a Cox mutation has little or no positive effect on products generated from the mixed acid fermentation pathway, with the exception of acetoin.

A crucial part of strain development involves the creation and testing of many multiple combinations of gene deletions and upregulations, to achieve the maximum production of desired metabolites, whilst still allowing viable cell growth. Therefore, the construction of a library containing successful CRISPR/Cas9 KO plasmids, that can be applied to strain development of any metabolite, is highly desirable. Many gene targets have been suggested in this study to increase the production of succinate; therefore, the production of KO and KI plasmids of these targets would make desirable initial vectors in a future potential library. Additional future work should involve the creation of a Δ*sdh*Δ*cox* double mutant. This would be an interesting combination as, similar to *cox*, *sdh* is also involved in the electron transport chain. It has previously been shown in *P. thermoglucosidasius* that cell growth can be increased through the

addition of acetic acid, which provides additional acetyl-CoA necessary for the regeneration of NAD^+ (118). Therefore, this method may prove useful to further improve metabolite production for the strains tested. Furthermore, investigations should be conducted regarding the single colony of the Δ*ldh*Δ*sdh* strain, which produced high ethanol levels under Ae conditions.

Ae production has many advantages, which include an increased uptake of the carbon source and also increased rate of cell growth (348). Although a 3.5-fold increase was achieved with both Δ*sdh* and Δ*ldh*Δ*sdh* strains under Ae conditions, to achieve higher levels of succinate production, the glyoxylate pathway must be manipulated. Isocitrate lyase regulator (IclR) is the transcriptional regulator for the glyoxylate pathway (146) and by targeting the IclR, along with genes from competing pathways, such as *adhE*, *ldh, ack* and phosphate acetyltransferase (*pta*), succinate production has been shown to increase (145-148). In addition to deleting genes in competing pathways, upregulation of specific genes is also often necessary. To achieve this, the knock-in abilities of CRISPR/Cas9 can be utilised. Two key targets for upregulation are the enzymes pyruvate carboxylase (Pyc, encoded by the *pyc* gene) and phosphoenolpyruvate carboxylase (Ppc, encoded by the *ppc* gene). These enzymes convert pyruvate and phosphoenolpyruvate to oxaloacetate respectively; therefore, upregulation of these would increase oxaloacetate levels, which is a key intermediary of succinate production (145,147,148). Additionally, overexpression of the enzymes within the glyoxylate pathway, such as

isocitrate lyase, malate synthase, and citrate synthase has been shown to direct the carbon flux through this pathway, (148) as well as upregulating acetyl-CoA synthase for the breakdown of the carbon source (147). The enzymes, and their encoded genes, discussed here represent ideal targets, which can be exploited using *stcas9-3*, to increase the expression of succinate.

One common problem with deleting competing pathways of mixed acid fermentation, is the lack of NAD^+ which is produced through oxidation of NADH by pyruvate (384-386). This may explain why the majority of strains unexpectedly exhibited no increase in succinate production. In order to overcome this limitation, it could be useful to facilitate the upregulation of malate dehydrogenase, which has been previously demonstrated to circumvent this problem (149). In addition to increasing metabolite production, the transportation of molecules in and out of the cell is just as important. Therefore, targeting import and export mechanisms could also be an efficient way to increase succinate production. Succinate transports are often general C4-dicarboxylic acid transporters. Beauprez *et al.* (387) have recently shown success in deletion of the succinate importer, *dctA*, and the overexpression of the succinate exporter, *dcuC*, to improve succinate production, achieving a 50 % increase in yield.

6.5 Key Outcomes

- Production of several mutants in WT *P. thermoglucosidasius*, which included the deletion of those encoding lactate dehydrogenase, cytochrome C oxidase and aldehyde-alcohol dehydrogenase, increasing the library of CRISPR/Cas9 KO vectors available for selection of specific KOs.

- Clarification that a cytochrome C oxidase mutation produces no significant advantage in the production of valuable metabolites and its detrimental effect on the metabolism of glucose.

- Production and conformation of a succinate dehydrogenase deficient strain in *P. thermoglucosidasius*.

- Enhancement of succinate production through the deletion of succinate dehydrogenase.

Chapter 7

General Discussion

The current environmental situation facing this planet, a rapidly increasingly population and a reliance on fossil fuels, has led to efforts to develop alternate, sustainable routes for the generation of important chemicals and fuels. One of the most obvious feedstocks for a sustainable process is biomass and, to avoid competition with food production, the emphasis is on using lignocellulosic material derived from non-food crops and agricultural wastes and residues. However, current studies have demonstrated limitations in production, an increase in associated costs and restrictions in carbon source utilisation. Thermophiles, notably *Parageobacillus* species, have been suggested as the solution to resolve these observed shortcomings. Yet, unlike model organisms such as *E. coli* and *S. cerevisiae*, these species are currently inaccessible due to limitations in the availability of genetic tools. Through the course of this study, the constraint of limited tools has been successfully eliminated, through the production of multiple libraries and the application of CRISPR/Cas9 for efficient genetic engineering. This study details the identification and subsequent testing of these tools, in addition to demonstrating their applicability for successful genome engineering in *P. thermoglucosidasius*.

7.1 Development of New and Existing Genetic Tools for *P. thermoglucosidasius*

The overall aim of this thesis was to identify and validate new genetic tools and to subsequently improve their utility. This initially involved the identification and validation of two stable fluorescent proteins, superfolder green fluorescent protein (sGFP) and enhanced consensus green protein variant 123 (eCGP123), for use in *P. thermoglucosidasius.* The successful implementation of sGFP was subsequently followed by two further published studies also demonstrating the success of sGFP in *P. thermoglucosidasius* (180,183), validating the results shown in this work. Nevertheless, the work described in this thesis is the first reported use of eCGP123 in a thermophilic bacillus, where it has higher thermostability than sGFP. The successful application of these fluorescent reporters has increased the possibilities available for future investigation of *P. thermoglucosidasius*. This has been demonstrated through protein localisation using fluorescence microscopy and a simpler characterisation of gene expression tools. The successful employment of additional reporter genes would be of future interest, in particular, those which produce fluorescence under different wavelengths. This would be advantageous, as it would allow further elucidation of processes within *P. thermoglucosidasius*, through the observation of colocalisation of proteins (388). Furthermore, additional fluorescent genes would allow methods

such as förster resonance energy transfer (FRET) to be used to elucidate protein interactions (389-391).

Following the successful application of fluorescent reporter genes, these were then exploited for the identification and testing of bioparts needed to tailor gene expression levels for future metabolic engineering projects. Specifically, they were used to validate and quantify the relative strengths of various promoters, RBS, transcriptional terminators and an inducible expression system, a significant advancement on what was currently available. These libraries facilitate a move towards producing a synthetic system, allowing tailored expression of endogenous and exogenous DNA, which is currently unattainable (7,24,159). These tools have now been successfully utilised for the genetic engineering of *P. thermoglucosidasius*, in both expression of further tools and the production of valuable products, both within this project and others (278) (Sheng, unpublished). Nevertheless, though this describes a significant advancement on the current tools available, a number of improvements to these tools are apparent. For example, the promoter library consists of endogenous sequences, which can be susceptible to homologous recombination (HR). Reeve *et al.* (180) and Pogrebnyakov *et al.* (183) have since produced promoter libraries that have mitigated this problem.

In addition to expression elements, the efficiency of exogenous gene expression can be influenced by translation factors. Differences in codon usage systems between organisms can dramatically effect expression

(81,86,209). Early methods for optimisation included the use of the Codon Adaptive Index (CAI) (82) and tRNA Adaptive Index (tAI) (204), to determine highly expressed codons and tRNA, similar to the manual codon optimisation evaluated in this study. However, several studies have now shown that alternative factors affect efficient expression and it has even been suggested that actually CAI plays no role in optimisation (86,178,209). Therefore, initial steps have been taken within this study towards identifying and developing a codon optimised usage system in *P. thermoglucosidasius*, using an algorithm produced by ATUM (178). Further optimisation of the *pheB* gene variants is now required, in order to produce a complete usage system optimised for protein expression. The data obtained within this study will be used to guide the design and synthesis of two further sets of variant genes, which can again be tested in terms of expression levels in *P. thermoglucosidasius* (178). This will facilitate the synthesis of codon optimised genes, which will allow the efficient expression of a desired protein.

7.2 Development of CRISPR/Cas-based Knock-In and Knock-Out Technology

The emergence of CRISPR/Cas9 technology for gene editing represents a disruptive technology. Clustered regularly interspaced short palindromic repeats, more commonly referred to as CRISPR, and CRISPR-associated proteins (Cas), allow the accurate targeting of precise locations within the genome. This facilitates the efficient addition or removal of specific gene sequences through allelic exchange (AE). First observed by Ishino (120) within the genome of *Escherichia coli* and identified as a bacterial adaptive immune system, this technology has been harnessed as the primary method for genome editing (135,136,246-248). This chapter details the successful implementation of two CRISPR/Cas9 systems from *S. thermophilus*, as genome editing tools in *P. thermoglucosidasius*. The successful implementation of these two *cas9* genes demonstrates the first expression of any Cas9 gene in thermophiles at temperatures above 45 °C. This work is also the first use of CRISPR/Cas9 technology to enable efficient and selective knock-ins (KI) and knock-outs (KO) in *P. thermoglucosidasius*. Successful utilisation of this technology was further demonstrated through the KO of multiple genes, in addition to the curing of the two native megaplasmids, pNCI001 and pNCI002. Curing demonstrated the nonessential nature of both the megaplasmids, but also implied the presence of a carbon utilisation pathway, for degradation of aromatic hydrocarbons, which was unnecessary under the conditions tested. Furthermore, this study demonstrates the first use of *stcas9-3* to KI

exogenous DNA into *P. thermoglucosidasius*, both onto the chromosome and the extrachromosomal megaplasmids. The efficiency of *stcas9-3* KI was also demonstrated to be much higher than spCas9 (237). However, as discussed, it is imperative to screen a much larger sample size in order to fully assess the efficiency of *stcas9-3*. This will be facilitated by the described optimisation of the in-house CRISPR/Cas9 protocol developed during the course of this study. Following the successful exemplification of the use of stCas9-1 and -3, two further examples of genome editing in thermophiles using Cas9 systems were published (237,269).

Further characterisation to establish the efficiency of each *cas9* is a necessary next step, which can be achieved through screening an increased number of colonies (237). Initial experimentation indicates that *stcas9-3* exhibits better efficiency than *stcas9-1*, a result also demonstrated by Muller *et al.* (247) in human cells. Another area of future interest is CRISPR interference (CRISPRi), which utilises a catalytically dead *cas9* (*dcas9*), through the inactivation of the nuclease domains RuvC and HNH (264), to affect gene transcription (297). CRISPRi can both repress and activate transcription, via the sequence specific binding of a promoter and allow the downregulation of essential genes and observe the coresponding effect (298-300). In previous studies using, *E. coli*, CRISPRi resulted in up to 99.9% repression of target gene transcription (392).

7.3 Increasing Transformation Efficiency through KO of R-M Systems

R-M systems are described as a form of primitive bacterial immunity. These systems often encode two proteins and carry out digestion and methylation of DNA. Digestion of foreign DNA occurs through a REase domain, which recognises a specific recognition sequence. Similarly, the encoded methyltransferase (MTase) also recognises the same sequence, of which methylation inhibits digestion. This methylation allows the bacteria to distinguish its own DNA from foreign DNA to be digested. R-M systems can present an insurmountable barrier to the effective genetic modification of a particular chassis. *P. thermoglucosidasius* contains three R-M systems of unknown function, which may be the cause of the previously observed low transformation efficiency of this species.

This work presented in this chapter demonstrated the successful deletion of three R-M genes in *P. thermoglucosidasius* using the newly developed CRISPR/Cas9 system to produce single, double and triples mutants. This study has demonstrated that an increase in transformation efficiency is achievable, recording 1 x 10^5 CFUs/µg of DNA for the triple mutant strain, which is an order of magnitude higher than any transformation previously observed (51,118,180). This increase in transformation efficiency should therefore enable an increase in the efficiency of the CRISPR/Cas9 method, allowing greater efficiency of AE and the number of transformants produced.

The generation of several single mutants resulted in a counterintuitive decrease in transformation efficiency. It was theorised that RM1 and RM3 represented separated MTase and REase genes, positioned on different loci (326), demonstrated by the reduction in transformation efficiency for ΔRM1 and the increase in ΔRM3. This phenomenon has been observed in other species, such as *Staphylococcus aureus* and *Lactococcus lactis*, which both possess R-M systems containing genes that are located separately (327,328). The work within this chapter has also highlighted an observed variability between replicates for the transformation efficiency of both WT *P. thermoglucosidasius* and the resulting R-M deficient strains. For future investigations, the primary focus should be to tackle this inconsistency, which may be improved through optimisation of the transformation protocol.

It is also necessary to conduct a phenotypic analysis on the R-M system mutants, in order to determine whether they have any metabolic or transcriptional effect. Many studies have suggested a potential fitness cost to encoding R-M systems (338-341). Restriction site avoidance (342), deamination of 5-methyl-cytosine (343), utilisation ATP for translocation (339), stabilisation of genomic islands (275) and epigenetic methylation (317,318,331,335,336) are all mechanisms that may exhibit a metabolic or transcriptional effect on *P. thermoglucosidasius*.

7.4 A View towards Metabolic Engineering of *P. thermoglucosidasius* for Improved Succinic Acid Production

As previously discussed, the environmental damage and increase in Green House Gas (GHG) emissions caused by the use of fossil fuels is one of the biggest problems currently facing the world. Therefore, the focus is now on reproducing products via carbon-neutral, renewable processes. One such route is the production of high value platform chemicals from microorganism metabolism of biomass, which aims to replace the current production by the petrochemical industry. Many of these platform chemicals, which are used in the pharmaceutical, and agricultural industries and also in food production, are naturally occurring metabolic intermediates of microorganisms and can be synthetically produced, or enhanced, through metabolic engineering. One such chemical is succinate which is produced naturally through the tricarboxylic acid (TCA) cycle, the phosphoenolpyruvate (PEP) pathway and the glyoxylate pathway.

The work presented in this chapter demonstrated how the tools developed throughout the course of this thesis could be utilised to explore the use of *P. thermoglucosidasius* for the production of industrially useful platform chemicals. More specifically, this work successfully demonstrated the deletion of the succinate dehydrogenase (Sdh) gene

(*sdh*) and observed a 3.5-fold increase in the concentration of succinate, compared to WT. This produces an initial strain that can be used to create further mutations from with the glyoxylate pathway. It has been shown that isocitrate lyase regulator (IclR) is responsible for transcriptional regulation of the glyoxylate pathway (146). This demonstrates a key initial target in the production of succinate, with the removal of pivotal genes of competing pathways, such as *adhE*, *ldh, ack* and *pta*, also showing an increase in succinate production (143,146-148). Furthermore, it has also been demonstrated that an increase in oxaloacetate levels, a key intermediary of succinate production, also increases the level of succinate produced. This can be achieved by the upregulation of genes such as *pyc* and *ppc* (145,147,148). One common problem often observed during succinate production is a limitation of NAD^+, which often causes growth aberrations, affecting the production of succinate (Singh, 2011; Lui, 2012; Lee, 2002). This limitation can be overcome through the upregulation of malate dehydrogenase, which has been previously been used to circumvent this problem (Wang, 2009).

7.5 Conclusions

The relative ease with which *P. thermoglucosidasius* can now be genetically modified has been demonstrated through the successful generation of multiple deletions and integrations produced throughout the course of this work. These developments have increased the attractiveness of this chassis for the industrial production of chemicals and fuels. This study has successfully achieved the remit of the project, by effectively expanding the available tools for genetic engineering of *P. thermoglucosidasius*. The initial subset of libraries, containing gene expression elements, were successfully utilised to produce more complex tools, such as CRISPR/Cas9, in order to facilitate genome modifications. These tools were then further validated through the deletion of the native R-M systems, integration of exogenous DNA and also the demonstration of increased succinate production, through deletion of genes from within the central carbon utilisation pathways. Further suggested enhancements of these tools will lead to more effective metabolic engineering of this chassis. Both the tools and strains presented here can be built upon to synthetically produce an industrially useful strain. This demonstrates promise for the potential of producing of biofuels and platform chemicals from lignocellulosic biomass using *P. thermoglucosidasius*.

Chapter 8

Bibliography

1. Worldometer.info. (2017) Global Population.
2. Matharu, A. S., de Melo, E. M., and Houghton, J. A. (2016) Opportunity for high value-added chemicals from food supply chain wastes. *Bioresource Technology* **215**, 123-130
3. Jager, G., and Buchs, J. (2012) Biocatalytic conversion of lignocellulose to platform chemicals. *Biotechnology Journal* **7**, 1122-1136
4. Keasling, J. D. (2010) Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358
5. Shafiee, S., and Topal, E. (2009) When will fossil fuel reserves be diminished? *Energy Policy* **37**, 181-189
6. Bloch, H., Rafiq, S., and Salim, R. (2015) Economic growth with coal, oil and renewable energy consumption in China: Prospects for fuel substitution. *Economic Modelling* **44**, 104-115
7. Lin, P. P., Rabe, K. S., Takasumi, J. L., Kadisch, M., Arnold, F. H., and Liao, J. C. (2014) Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabolic Engineering* **24**, 1-8
8. REN21. (2017) Renewables 2017 Global Status Report.
9. Haines, A., and Patz, J. A. (2004) Health effects of climate change. *Jama* **291**, 99-103
10. Brunekreef, B., and Holgate, S. T. (2002) Air pollution and health. *Lancet (London, England)* **360**, 1233-1242
11. Kampa, M., and Castanas, E. (2008) Human health effects of air pollution. *Environmental Pollution* **151**, 362-367
12. Klare, M. T. (2001) The new geography of conflict. *Foreign Affairs* **80**, 49
13. Li, C. B., Shi, H. Q., Cao, Y. J., Wang, J. H., Kuang, Y. H., Tan, Y., and Wei, J. (2015) Comprehensive review of renewable energy curtailment and avoidance: A specific example in China. *Renewable & Sustainable Energy Reviews* **41**, 1067-1079
14. House of Commons, E. A. C. (2008) Are biofuels sustainable?, House of Commons
15. Goldemberg, J. (2007) Ethanol for a sustainable energy future. *Science* **315**, 808-810
16. Sims, R., Taylor, M., Saddler, J., Mabee, W. (2008) From 1st- to 2nd-generation biofuel technologies. *IEA Bioenergy*
17. Taylor, M. P., Eley, K. L., Martin, S., Tuffin, M. I., Burton, S. G., and Cowan, D. A. (2009) Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends in Biotechnology* **27**, 398-405
18. van Maris, A. J. A., Abbott, D. A., Bellissimi, E., van den Brink, J., Kuyper, M., Luttik, M. A. H., Wisselink, H. W., Scheffers, W. A., van Dijken, J. P., and Pronk, J. T. (2006) Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek* **90**, 391-418
19. Corporation, I. (2005) Cellulose ethanol: clean fuel for today and tomorrow. Iogen Corporation
20. Nordhoff, S. (2007) Food vs fuel - the role of biotechnology. *Biotechnology Journal* **2**, 1451
21. Tenenbaum, D. J. (2008) Food vs. fuel: diversion of crops could cause more hunger. *Environmental Health Perspectives* **116**, A254-A257
22. Havlík, P., Schneider, U. A., Schmid, E., Böttcher, H., Fritz, S., Skalský, R., Aoki, K., Cara, S. D., Kindermann, G., Kraxner, F., Leduc, S., McCallum, I., Mosnier, A., Sauer, T., and Obersteiner, M.

(2011) Global land-use implications of first and second generation biofuel targets. *Energy Policy* **39**, 5690-5702
23. Hamelinck, C. N., Hooijdonk, G. v., and Faaij, A. P. C. (2005) Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass Bioenergy* **28**, 384-410
24. Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., and Atkinson, T. (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering* **11**, 398-408
25. Aro, E.-M. (2016) From first generation biofuels to advanced solar biofuels. *Ambio* **45**, 24-31
26. Tsao, G. T., Cao, N. J., Du, J., and Gong, C. S. (1999) Production of multifunctional organic acids from renewable resources. *Advances in Biochemical Engineering and Biotechnology* **65**, 243-280
27. Zeikus, J. G., Jain, M. K., and Elankovan, P. (1999) Biotechnology of succinic acid production and markets for derived industrial products. *Applied Microbiology and Biotechnology* **51**, 545-552
28. McKinlay, J. B., Vieille, C., and Zeikus, J. G. (2007) Prospects for a bio-based succinate industry. *Applied Microbiology and Biotechnology* **76**, 727-740
29. Black, M., and Miller, R. (2006) Platform chemicals from crops. *Journal of Chemical Technology and Biotechnology* **81**, 1725-1728
30. Jang, Y. S., Kim, B., Shin, J. H., Choi, Y. J., Choi, S., Song, C. W., Lee, J., Park, H. G., and Lee, S. Y. (2012) Bio-based production of C2-C6 platform chemicals. *Biotechnology and Bioengineering* **109**, 2437-2459
31. Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A., and Wang, Z. (2013) Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnology Advances* **31**, 1200-1223
32. Ingram, L. O., Conway, T., Clark, D. P., Sewell, G. W., and Preston, J. F. (1987) Genetic engineering of ethanol production in *Escherichia coli*. *Applied and Environmental Microbiology* **53**, 2420-2425
33. Portnoy, V. A., Bezdan, D., and Zengler, K. (2011) Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology* **22**, 590-594
34. Sonderegger, M., and Sauer, U. (2003) Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose. *Applied and Environmental Microbiology* **69**, 1990-1998
35. Dien, B. S., Cotta, M. A., and Jeffries, T. W. (2003) Bacteria engineered for fuel ethanol production: current status. *Applied Microbiology and Biotechnology* **63**, 258-266
36. Rogers, P. L., Lee, K. J., Skotnicki, M. L., and Tribe, D. E. (1982) Ethanol production by *Zymomonas mobilis*. in *Microbial Reactions*, Springer Berlin Heidelberg. pp 37-84
37. Zhang, M., Eddy, C., Deanda, K., Finkestein, M., and Picataggio, S. (1995) Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* **267**, 240-243
38. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804-807
39. Alterthum, F., and Ingram, L. O. (1989) Efficient ethanol production from glucose, lactose, and xylose by recombinant *Escherichia coli*. *Applied and Environmental Microbiology* **55**, 1943-1948

40. Zaldivar, J., Nielsen, J., and Olsson, L. (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology* **56**, 17-34
41. Mills, T. Y., Sandoval, N. R., and Gill, R. T. (2009) Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnology for Biofuels* **2**, 26
42. Jeffries, T. W. (1983) Utilization of xylose by bacteria, yeasts, and fungi. *Advances in Biochemical Engineering/Biotechnology* **27**, 1-32
43. Stulke, J., and Hillen, W. (1999) Carbon catabolite repression in bacteria. *Current Opinion in Microbiology* **2**, 195-201
44. Gorke, B., and Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews. Microbiology* **6**, 613-624
45. Lin, L., Song, H., Tu, Q., Qin, Y., Zhou, A., Liu, W., He, Z., Zhou, J., and Xu, J. (2011) The *Thermoanaerobacter* glycobiome reveals mechanisms of pentose and hexose co-utilization in bacteria. *PLoS Genetics* **7**, e1002318
46. Bischoff, K. M., Skinner-Nemec, K. A., and Leathers, T. D. (2007) Antimicrobial susceptibility of *Lactobacillus* species isolated from commercial ethanol plants. *Journal of Industrial Microbiology and Biotechnology* **34**, 739-744
47. Skinner, K. A., and Leathers, T. D. (2004) Bacterial contaminants of fuel ethanol production. *Journal of Industrial Microbiology and Biotechnology* **31**, 401-408
48. Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., and Sani, R. K. (2013) Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology* **128**, 751-759
49. Turner, P., Mamo, G., and Karlsson, E. N. (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories* **6**, 9
50. Haki, G. D., and Rakshit, S. K. (2003) Developments in industrially important thermostable enzymes: a review. *Bioresource Technology* **89**, 17-34
51. Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J., and Cowan, D. A. (2011) Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microbial Biotechnology* **4**, 438-448
52. Kananaviciute, R., and Citavicius, D. (2015) Genetic engineering of *Geobacillus spp. Journal of Microbiological Methods* **111**, 31-39
53. Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M., and Abe, S. (1983) *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic bacilli. *Systematic and Applied Microbiology* **4**, 487-495
54. Ash, C., Farrow, J. A. E., Wallbanks, S., and Collins, M. D. (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative-analysis of small-subunit-ribosomal RNA sequences. *Letters in Applied Microbiology* **13**, 202-206
55. Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., Lysenko, A. M., Petrunyaka, V. V., Osipov, G. A., Belyaev, S. S., and Ivanov, M. V. (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp nov and *Geobacillus uzenensis* sp nov from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus*

thermoglucosidasius and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* **51**, 433-446

56. Bartosiak-Jentys, J., Hussein, A. H., Lewis, C. J., and Leak, D. J. (2013) Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology* **159**, 1267-1275
57. Raita, M., Ibenegbu, C., Champreda, V., and Leak, D. J. (2016) Production of ethanol by thermophilic oligosaccharide utilising *Geobacillus thermoglucosidasius* TM242 using palm kernel cake as a renewable feedstock. *Biomass Bioenergy* **95**, 45-54
58. Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nature Reviews. Microbiology* **2**, 241-249
59. Panja, S., Saha, S., Jana, B., and Basu, T. (2006) Role of membrane potential on artificial transformation of *E. coli* with plasmid DNA. *Journal of Biotechnology* **127**, 14-20
60. Rahimzadeh, M., Sadeghizadeh, M., Najafi, F., Arab, S., and Mobasheri, H. (2016) Impact of heat shock step on bacterial transformation efficiency. *Molecular Cell Biology Research Communications* **5**, 257-261
61. Tsong, T. Y. (1991) Electroporation of cell membranes. *Biophysical Journal* **60**, 297-306
62. Imanaka, T., Fujii, M., Aramori, I., and Aiba, S. (1982) Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. *Journal of Bacteriology* **149**, 824-830
63. Wu, L. J., and Welker, N. E. (1989) Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. *Journal of General Microbiology* **135**, 1315-1324
64. Monod, M., Denoya, C., and Dubnau, D. (1986) Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. *Journal of Bacteriology* **167**, 138-147
65. Vasu, K., and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and Molecular Biology Reviews : MMBR* **77**, 53-72
66. Dryden, D. T., Murray, N. E., and Rao, D. N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Research* **29**, 3728-3741
67. Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cellular and Molecular Life Sciences : CMLS* **62**, 685-707
68. McClelland, S. E., and Szczelkun, M. D. (2004) The Type I and III restriction endonucleases: structural elements in molecular motors that process DNA. in *Restriction Endonucleases* (Pingoud, A. M. ed.), Springer Berlin Heidelberg, Berlin, Heidelberg. pp 111-135
69. Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Kruger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker,

E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G., and Xu, S. Y. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research* **31**, 1805-1812

70. Stahl, M. M., Kobayashi, I., Stahl, F. W., and Huntington, S. K. (1983) Activation of Chi, a recombinator, by the action of an endonuclease at a distant site. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 2310-2313
71. Dillingham, M. S., and Kowalczykowski, S. C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and Molecular Biology Reviews : MMBR* **72**, 642-671, Table of Contents
72. Bickle, T. A., and Kruger, D. H. (1993) Biology of DNA restriction. *Microbiological Reviews* **57**, 434-450
73. Tock, M. R., and Dryden, D. T. (2005) The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**, 466-472
74. Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M., Edwardson, J. M., and Dryden, D. T. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Molecular Cell* **9**, 187-194
75. Levin, B. R. (1993) The accessory genetic elements of bacteria: existence conditions and (co)evolution. *Current Opinion in Genetics & Development* **3**, 849-854
76. Guiziou, S., Sauveplane, V., Chang, H. J., Clerte, C., Declerck, N., Jules, M., and Bonnet, J. (2016) A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Research* **44**, 7495-7508
77. Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of Molecular Biology* **180**, 549-576
78. Sorensen, M. A., Kurland, C. G., and Pedersen, S. (1989) Codon usage determines translation rate in *Escherichia coli*. *Journal of Molecular Biology* **207**, 365-377
79. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* **2**, 13-34
80. Stoletzki, N., and Eyre-Walker, A. (2007) Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Molecular Biology and Evolution* **24**, 374-381
81. Gouy, M., and Gautier, C. (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research* **10**, 7055-7074
82. Sharp, P. M., and Li, W. H. (1987a) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15**, 1281-1295
83. Sharp, P. M., and Li, W. H. (1987b) The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution* **4**, 222-230
84. Duret, L., and Mouchiroud, D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4482-4487
85. Goetz, R. M., and Fuglsang, A. (2005) Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from

Escherichia coli. *Biochemical and Biophysical Research Communications* **327**, 4-7
86. Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255-258
87. Shearwin, K. E., Callen, B. P., and Egan, J. B. (2005) Transcriptional interference – a crash course. *Trends in Genetics : TIG* **21**, 339-345
88. Adhya, S., and Gottesman, M. (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**, 939-944
89. Wang, J. P., Yeh, C. M., and Tsai, Y. C. (2006) Improved subtilisin YaB production in *Bacillus subtilis* using engineered synthetic expression control sequences. *Journal of Agricultural and Food Chemistry* **54**, 9405-9410
90. Wu, S. C., and Wong, S. L. (1999) Development of improved pUB110-based vectors for expression and secretion studies in *Bacillus subtilis*. *Journal of Biotechnology* **72**, 185-195
91. Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *Journal of Cellular and Comparative Physiology* **59**, 223-239
92. Arun, K. H., Kaul, C. L., and Ramarao, P. (2005) Green fluorescent proteins in receptor research: an emerging tool for drug discovery. *Journal of Pharmacological and Toxicological Methods* **51**, 1-23
93. Chudakov, D. M., Matz, M. V., Lukyanov, S., and Lukyanov, K. A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* **90**, 1103-1163
94. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994) Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805
95. Jugder, B. E., Welch, J., Braidy, N., and Marquis, C. P. (2016) Construction and use of a *Cupriavidus necator* H16 soluble hydrogenase promoter (PSH) fusion to gfp (green fluorescent protein). *PeerJ* **4**, e2269
96. Phillips, G. J. (2001) Green fluorescent protein--a bright idea for the study of bacterial protein localization. *FEMS Microbiology Letters* **204**, 9-18
97. Stepanenko, O. V., Verkhusha, V. V., Kuznetsova, I. M., Uversky, V. N., and Turoverov, K. K. (2008) Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes. *Current Protein & Peptide Science* **9**, 338-369
98. Ronald, G. K., and Kindell, S. M. (2005) Quantitative assay of senescence-associated β-galactosidase activity in mammalian cell extracts. *Analytical Biochemistry* **343**, 329-334
99. Ludwicka, A., Switalski, L. M., Lundin, A., Pulverer, G., and Wadström, T. (1985) Bioluminescent assay for measurement of bacterial attachment to polyethylene. *Journal of Microbiological Methods* **4**, 169-177
100. de Wet, J. R., Wood, K. V., Helinski, D. R., and DeLuca, M. (1985) Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 7870-7873
101. Smale, S. T. (2010) Chloramphenicol acetyltransferase assay. *Cold Spring Harbor protocols* **2010**, pdb prot5422
102. Gossen, M., and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.

Proceedings of the National Academy of Sciences of the United States of America **89**, 5547-5551
103. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769
104. Jacob, F., and Monod, J. (1961a) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318-356
105. Jacob, F., and Monod, J. (1961b) On regulation of gene activity. *Cold Spring Harb Symp Quant Biol* **26**, 193-+
106. Friehs, K. (2004) Plasmid copy number and plasmid stability. *Advances in biochemical engineering/biotechnology* **86**, 47-82
107. Kohara, Y., Akiyama, K., and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495-508
108. Henry, M. F., and Cronan, J. E., Jr. (1991) Direct and general selection for lysogens of *Escherichia coli* by phage lambda recombinant clones. *Journal of Bacteriology* **173**, 3724-3731
109. Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P., and Kushner, S. R. (1989) New method for generating deletions and gene replacements in *Escherichia coli*. *Journal of Bacteriology* **171**, 4617-4622
110. Slater, S., and Maurer, R. (1993) Simple phagemid-based system for generating allele replacements in *Escherichia coli*. *Journal of Bacteriology* **175**, 4260-4262
111. Yang, X. W., Model, P., and Heintz, N. (1997) Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nature Biotechnology* **15**, 859-865
112. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33**, 25-35
113. Russell, C. B., Thaler, D. S., and Dahlquist, F. W. (1989) Chromosomal transformation of *Escherichia coli recD* strains with linearized plasmids. *Journal of Bacteriology* **171**, 2609-2613
114. Itaya, M., and Tanaka, T. (1990) Gene-directed mutagenesis on the chromosome of *Bacillus subtilis* 168. *Molecular & General Genetics : MGG* **223**, 268-272
115. Itaya, M., Fujita, K., Kuroki, A., and Tsuge, K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nature methods* **5**, 41-43
116. Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K., and Minton, N. P. (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research* **40**, e59
117. Boeke, J. D., LaCroute, F., and Fink, G. R. (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular & General Genetics : MGG* **197**, 345-346
118. Sheng, L., Kovács, K., Winzer, K., Zhang, Y., and Minton, N. P. (2017) Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnology for Biofuels* **10**, 5
119. Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., Liew, F., Kelly, M. L., Sheng, L., Schwarz, K., and Zhang, Y. (2016) A roadmap for gene system development in *Clostridium*. *Anaerobe* **41**, 104-112

120. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology* **169**, 5429-5433
121. Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002a) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43**, 1565-1575
122. Jansen, R., van Embden, J. D., Gaastra, W., and Schouls, L. M. (2002b) Identification of a novel family of sequence repeats among prokaryotes. *Omics : a journal of integrative biology* **6**, 23-33
123. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561
124. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* **60**, 174-182
125. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712
126. Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S., and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239-1259
127. Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J., and Koonin, E. V. (2011) Evolution and classification of the CRISPR-Cas systems. *Nature Reviews. Microbiology* **9**, 467-477
128. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R., and Koonin, E. V. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews. Microbiology* **13**, 722-736
129. Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013) The basic building blocks and evolution of CRISPR-CAS systems. *Biochemical Society Transactions* **41**, 1392-1400
130. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740
131. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607
132. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Molecular Microbiology* **77**, 1367-1379
133. Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H., and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71

134. Marraffini, L. A., and Sontheimer, E. J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845
135. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821
136. Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology* **16**, 253
137. Bozell, J. J., and Petersen, G. R. (2010) Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chemistry* **12**, 539-554
138. Beauprez, J. J., De Mey, M., and Soetaert, W. K. (2010) Microbial succinic acid production: Natural versus metabolic engineered producers. *Process Biochemistry* **45**, 1103-1114
139. Werpy, T., Petersen, G., Aden, A., Bozell, J., Holladay, J., White, J., Manheim, A., Eliot, D., Lasure, L., and Jones, S. (2004) *Top value added chemicals from biomass. volume 1 - results of screening for potential candidates from sugars and synthesis gas*,
140. Sauer, M., Porro, D., Mattanovich, D., and Branduardi, P. (2008) Microbial production of organic acids: expanding the markets. *Trends in Biotechnology* **26**, 100-108
141. Clark, D. P. (1989) The fermentation pathways of *Escherichia coli*. *FEMS Microbiology Letters* **5**, 223-234
142. Kornberg, H. L. (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochemical Journal* **99**, 1-11
143. Lin, H., Bennett, G. N., and San, K. Y. (2005a) Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. *Metabolic Engineering* **7**, 116-127
144. Lin, H., Bennett, G. N., and San, K. Y. (2005b) Fed-batch culture of a metabolically engineered *Escherichia coli* strain designed for high-level succinate production and yield under aerobic conditions. *Biotechnology and Bioengineering* **90**, 775-779
145. Lin, H., Bennett, G. N., and San, K. Y. (2005c) Genetic reconstruction of the aerobic central metabolism in *Escherichia coli* for the absolute aerobic production of succinate. *Biotechnology and Bioengineering* **89**, 148-156
146. Balzer, G. J., Thakker, C., Bennett, G. N., and San, K. Y. (2013) Metabolic engineering of *Escherichia coli* to minimize byproduct formate and improving succinate productivity through increasing NADH availability by heterologous expression of NAD(+)-dependent formate dehydrogenase. *Metabolic Engineering* **20**, 1-8
147. Zhu, N., Xia, H., Wang, Z., Zhao, X., and Chen, T. (2013) Engineering of acetate recycling and citrate synthase to improve aerobic succinate production in *Corynebacterium glutamicum*. *PloS One* **8**, e60659
148. Zhu, N., Xia, H., Yang, J., Zhao, X., and Chen, T. (2014) Improved succinate production in *Corynebacterium glutamicum* by engineering glyoxylate pathway and succinate export system. *Biotechnology Letters* **36**, 553-560
149. Wang, W., Li, Z., Xie, J., and Ye, Q. (2009) Production of succinate by a *pflB ldhA* double mutant of *Escherichia coli* overexpressing malate dehydrogenase. *Bioprocess and Biosystems Engineering* **32**, 737

150. Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., Moore, J. C., Shanmugam, K. T., and Ingram, L. O. (2008) Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnology and Bioengineering* **99**, 1140-1153
151. Millard, C. S., Chao, Y. P., Liao, J. C., and Donnelly, M. I. (1996) Enhanced production of succinic acid by overexpression of phosphoenolpyruvate carboxylase in *Escherichia coli*. *Applied and Environmental Microbiology* **62**, 1808-1810
152. Guettler, M. V., Rumler, D., and Jain, M. K. (1999) *Actinobacillus succinogenes* sp. nov., a novel succinic-acid-producing strain from the bovine rumen. *International Journal of Systematic Bacteriology* **49 Pt 1**, 207-216
153. Lee, S. J., Song, H., and Lee, S. Y. (2006) Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. *Applied and Environmental Microbiology* **72**, 1939-1948
154. Meynial-Salles, I., Dorotyn, S., and Soucaille, P. (2008) A new process for the continuous production of succinic acid from glucose at high yield, titer, and productivity. *Biotechnology and Bioengineering* **99**, 129-135
155. Solovyev, V. (2011) *Automatic Annotation of Microbial Genomes and Metagenomic Sequences*, Nova Science
156. Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vazquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., and Scalbert, A. (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* **46**, D608-D617
157. Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009) A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods* **78**, 79-85
158. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **28**, 292
159. Bartosiak-Jentys, J., Eley, K., and Leak, D. J. (2012) Application of *pheB* as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength. *Applied and Environmental Microbiology* **78**, 5945-5947
160. Isaacs, F. J., Dwyer, D. J., and Collins, J. J. (2006) RNA synthetic biology. *Nature Biotechnology* **24**, 545-554
161. Benenson, Y. (2012) Synthetic biology with RNA: progress report. *Current Opinion in Biotechnology* **16**, 278-284
162. Cui, W., Han, L., Cheng, J., Liu, Z., Zhou, L., Guo, J., and Zhou, Z. (2016) Engineering an inducible gene expression system for *Bacillus subtilis* from a strong constitutive promoter and a theophylline-activated synthetic riboswitch. *Microbial Cell Factories* **15**, 199
163. Tsien, R. Y. (1998) The green fluorescent protein. *Annual Review of Biochemistry* **67**, 509-544
164. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24**, 79-88
165. Cava, F., de Pedro, M. A., Blas-Galindo, E., Waldo, G. S., Westblade, L. F., and Berenguer, J. (2008) Expression and use of superfolder green fluorescent protein at high temperatures *in vivo*:

a tool to study extreme thermophile biology. *Applied and Environmental Microbiology* **10**, 605-613
166. Blanchard, K., Robic, S., and Matsumura, I. (2014) Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. *Applied Microbiology and Biotechnology* **98**, 6715-6723
167. Ulijasz, A. T., Cornilescu, G., von Stetten, D., Kaminski, S., Mroginski, M. A., Zhang, J. R., Bhaya, D., Hildebrandt, P., and Vierstra, R. D. (2008) Characterization of two thermostable cyanobacterial phytochromes reveals global movements in the chromophore-binding domain during photoconversion. *Journal of Biological Chemistry* **283**, 21251-21266
168. Don Paul, C., Traore, D. A. K., Byres, E., Rossjohn, J., Devenish, R. J., Kiss, C., Bradbury, A., Wilce, M. C. J., and Prescott, M. (2011) Expression, purification, crystallization and preliminary X-ray analysis of eCGP123, an extremely stable monomeric green fluorescent protein with reversible photoswitching properties. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* **67**
169. Kiss, C., Temirov, J., Chasteen, L., Waldo, G. S., and Bradbury, A. R. M. (2009) Directed evolution of an extremely stable fluorescent protein. *Protein Engineering, Design and Selection* **22**, 313-323
170. Song, X. F., Wang, Y. F., Shu, Z. Y., Hong, J. B., Li, T., and Yao, L. S. (2013) Engineering a more thermostable blue light photo receptor *Bacillus subtilis* YtvA LOV domain by a computer aided rational design method. *PLOS Computational Biology* **9**
171. Tan, P. S., van Alen-Boerrigter, I. J., Poolman, B., Siezen, R. J., de Vos, W. M., and Konings, W. N. (1992) Characterization of the *Lactococcus lactis pepN* gene encoding an aminopeptidase homologous to mammalian aminopeptidase N. *FEBS Letters* **306**, 9-16
172. Boot, H. J., Kolen, C. P., van Noort, J. M., and Pouwels, P. H. (1993) S-layer protein of *Lactobacillus acidophilus* ATCC 4356: purification, expression in *Escherichia coli*, and nucleotide sequence of the corresponding gene. *Journal of Bacteriology* **175**, 6089-6096
173. Eidhin, D. N., Ryan, A. W., Doyle, R. M., Walsh, J. B., and Kelleher, D. (2006) Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of *Clostridium difficile*. *Journal of Medical Microbiology* **55**, 69-83
174. Barthelemy, I., Salas, M., and Mellado, R. P. (1987) *In vivo* transcription of bacteriophage phi 29 DNA: transcription termination. *Journal of Virology* **61**, 1751-1755
175. Grundy, F. J., Moir, T. R., Haldeman, M. T., and Henkin, T. M. (2002) Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Research* **30**, 1646-1655
176. Lampe, M. F., and Bott, K. F. (1984) Cloning the gyrA gene of *Bacillus subtilis*. *Nucleic Acids Research* **12**, 6307-6323
177. Graves, M. C., Mullenbach, G. T., and Rabinowitz, J. C. (1985) Cloning and nucleotide sequence determination of the *Clostridium pasteurianum* ferredoxin gene. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 1653-1657
178. Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009) Design parameters to

control synthetic gene expression in *Escherichia coli*. *PloS One* **4**, e7002
179. Henche, A. L., Koerdt, A., Ghosh, A., and Albers, S. V. (2012) Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environmental Microbiology* **14**, 779-793
180. Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. J., and Ellis, T. (2016) The *Geobacillus* plasmid set: A modular toolkit for thermophile engineering. *ACS Synthetic Biology* **5**, 1342-1347
181. Lakowicz, J. R. (2006) Quenching of Fluorescence. in *Principles of Fluorescence Spectroscopy* (Lakowicz, J. R. ed.), Springer US, Boston, MA. pp 277-330
182. Ware, W. R. (1962) Oxygen quenching of fluorescence in solution: an experimental study of the diffusion process. *The Journal of Physical Chemistry* **66**, 455-458
183. Pogrebnyakov, I., Jendresen, C. B., and Nielsen, A. T. (2017) Genetic toolbox for controlled expression of functional proteins in *Geobacillus* spp. *PloS One* **12**, e0171313
184. Jensen, P. R., and Hammer, K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and Environmental Microbiology* **64**, 82-87
185. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013a) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods* **10**, 354-360
186. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G., Keasling, J. D., Endy, D., and Arkin, A. P. (2013b) Quantitative estimation of activity and quality for collections of functional genetic elements. *Nature Methods* **10**, 347-353
187. Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2013) Promoters maintain their relative activity levels under different growth conditions. *Molecular Systems Biology* **9**, 701
188. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *The Glycolytic Pathway Is Tightly Controlled. Biochemistry. 5th edition.*, New York: W H Freeman
189. Mor, I., Cheung, E. C., and Vousden, K. H. (2011) Control of glycolysis through regulation of PFK1: old friends and recent additions. *Cold Spring Harb Symp Quant Biol* **76**, 211-216
190. Lee, K., Holland-Staley, C. A., and Cunningham, P. R. (1996) Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *RNA (New York, N.Y.)* **2**, 1270-1285
191. Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001) Anatomy of *Escherichia coli* ribosome binding sites. *Journal of Molecular Biology* **313**, 215-228
192. Komarova, A. V., Tchufistova, L. S., Supina, E. V., and Boni, I. V. (2002) Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA (New York, N.Y.)* **8**, 1137-1147
193. Vimberg, V., Tats, A., Remm, M., and Tenson, T. (2007) Translation initiation region sequence preferences in *Escherichia coli*. *BMC Molecular Biology* **8**, 100

194. de Smit, M. H., and van Duin, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 7668-7672
195. de Smit, M. H., and van Duin, J. (1994) Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *Journal of Molecular Biology* **244**, 144-150
196. Kozak, M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 2850-2854
197. Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994a) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Research* **22**, 4953-4957
198. Shine, J., and Dalgarno, L. (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 1342-1346
199. Gusarov, I., and Nudler, E. (1999) The Mechanism of Intrinsic Transcription Termination. *Molecular Cell* **3**, 495-504
200. Schwartz, A., Rahmouni, A., and Boudvillain, M. (2003) The functional anatomy of an intrinsic transcription terminator. *EMBO J* **22**, 3385-3394
201. Mohr, G., Kang, S. Y.-S., Park, S. K., Qin, Y., Grohman, J., Yao, J., Stamos, J. L., and Lambowitz, A. M. (2018) A highly proliferative group IIC intron from *Geobacillus stearothermophilus* reveals new features of group II Intron mobility and splicing. *Journal of Molecular Biology* **430**, 2760-2783
202. Moretz, S. E., and Lampson, B. C. (2010) A group IIC-type intron interrupts the rRNA methylase gene of *Geobacillus stearothermophilus* strain 10. *Journal of Bacteriology* **192**, 5245
203. Gingold, H., and Pilpel, Y. (2011) Determinants of translation efficiency and accuracy. *Molecular Systems Biology* **7**, 481
204. dos Reis, M., Wernisch, L., and Savva, R. (2003) Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Research* **31**, 6976-6985
205. Van Zyl, L. J., Taylor, M. P., Eley, K., Tuffin, M., and Cowan, D. A. (2013) Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host *Geobacillus thermoglucosidasius*. *Applied Microbiology and Biotechnology* **98**, 1247-1259
206. Angov, E. (2011b) Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology Journal* **6**, 650-659
207. Angov, E., Hillier, C. J., Kincaid, R. L., and Lyon, J. A. (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS One* **3**, e2189
208. Angov, E., Legler, P. M., and Mease, R. M. (2011a) Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods in Molecular Biology* **705**, 1-13
209. Allert, M., Cox, J. C., and Hellinga, H. W. (2010) Multifactorial determinants of protein expression in prokaryotic open reading frames. *Journal of Molecular Biology* **402**, 905-918

210. Wu, X., Jornvall, H., Berndt, K. D., and Oppermann, U. (2004) Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochemical and biophysical research communications* **313**, 89-96
211. Maertens, B., Spriestersbach, A., von Groll, U., Roth, U., Kubicek, J., Gerrits, M., Graf, M., Liss, M., Daubert, D., Wagner, R., and Schafer, F. (2010) Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. *Protein science* **19**, 1312-1326
212. Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., and Welch, M. (2012) Engineering genes for predictable protein expression. *Protein Expression and Purification* **83**, 37-46
213. Kurland, C., and Gallant, J. (1996) Errors of heterologous protein expression. *Current Opinion in Biotechnology* **7**, 489-493
214. Kane, J. F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 494-500
215. Hayes, C. S., Bose, B., and Sauer, R. T. (2002) Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3440-3445
216. Spanjaard, R. A., and van Duin, J. (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7967-7971
217. Kramer, E. B., and Farabaugh, P. J. (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA (New York, N.Y.)* **13**, 87-96
218. Olivares-Trejo, J. J., Bueno-Martinez, J. G., Guarneros, G., and Hernandez-Sanchez, J. (2003) The pair of arginine codons AGA AGG close to the initiation codon of the lambda int gene inhibits cell growth and protein synthesis by accumulating peptidyl-tRNAArg4. *Molecular Microbiology* **49**, 1043-1049
219. Chen, G. T., and Inouye, M. (1994b) Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes & Development* **8**, 2641-2652
220. Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., and Gottesman, M. M. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525-528
221. Widmann, M., Clairo, M., Dippon, J., and Pleiss, J. (2008) Analysis of the distribution of functionally relevant rare codons. *BMC Genomics* **9**, 207
222. Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., and Nussinov, R. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology* **383**, 281-291
223. Suzuki, H., Yoshida, K., and Ohshima, T. (2013) Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426. *Applied and Environmental Microbiology* **79**, 5151-5158
224. Rudolph, M. M., Vockenhuber, M. P., and Suess, B. (2015) Conditional control of gene expression by synthetic riboswitches in *Streptomyces coelicolor*. *Methods in Enzymology* **550**, 283-299
225. Barnes, P. J. (2013) Theophylline. *American Journal of Respiratory and Critical Care Medicine* **188**, 901-906

226. Raj, C. V. S., and Dhala, S. (1965) Effect of Naturally Occurring Xanthines on Bacteria: I. Antimicrobial Action and Potentiating Effect on Antibiotic Spectra. *Applied Microbiology* **13**, 432-436
227. Algharrawi, K. H., Summers, R. M., Gopishetty, S., and Subramanian, M. (2015) Direct conversion of theophylline to 3-methylxanthine by metabolically engineered *E. coli*. *Microbial Cell Factories* **14**, 203
228. Klumpp, S., Zhang, Z., and Hwa, T. (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375
229. Shahrezaei, V., and Marguerat, S. (2015) Connecting growth with gene expression: of noise and numbers. *Current Opinion in Microbiology* **25**, 127-135
230. Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012) CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host & Microbe* **12**, 177-186
231. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233-239
232. Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., Van Nostrand, J. D., He, Z., and Zhou, J. (2015) Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology* **81**, 4423-4431
233. Altenbuchner, J. (2016) Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Applied and Environmental Microbiology* **82**, 5421-5427
234. Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016) Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. *Applied and Environmental Microbiology* **82**, 6109-6119
235. Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., and Gu, Y. (2016) CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synthetic Biology* **5**, 1355-1361
236. Wasels, F., Jean-Marie, J., Collas, F., Lopez-Contreras, A. M., and Lopes Ferreira, N. (2017) A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of Microbiological Methods* **140**, 5-11
237. Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., and van Kranenburg, R. (2017a) Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS Synthetic Biology* **6**, 849-861
238. Tamime, A. Y., and Deeth, H. C. (1980) Yogurt: technology and biochemistry. *Journal of Food Protection* **43**, 939-977
239. Fox, P. F. (1993) Cheese : Chemistry, Physics and Microbiology, Chapman & Hall, London; New York
240. Lier, C., Baticle, E., Horvath, P., Haguenoer, E., Valentin, A.-S., Glaser, P., Mereghetti, L., and Lanotte, P. (2015) Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Frontiers in Genetics* **6**
241. Serbanescu, M. A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., Yakunin, A. F., Senadheera, D. B., and Cvitkovitch, D. G. (2015) Role of the *Streptococcus mutans* CRISPR-Cas Systems in immunity and cell physiology. *Journal of Bacteriology* **197**, 749-761

242. Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G. D., Fonstein, M., Overbeek, R., Kyprides, N., Purnelle, B., Prozzi, D., Ngui, K., Masuy, D., Hancy, F., Burteau, S., Boutry, M., Delcour, J., Goffeau, A., and Hols, P. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nature Biotechnology* **22**, 1554-1558
243. Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology* **190**, 1401-1412
244. Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., and Church, G. M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116-1121
245. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P., Li, Z., Peterson, R. T., Yeh, J. R., Aryee, M. J., and Joung, J. K. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485
246. Xu, K., Ren, C., Liu, Z., Zhang, T., Zhang, T., Li, D., Wang, L., Yan, Q., Guo, L., Shen, J., and Zhang, Z. (2015) Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cellular and Molecular Life Sciences : CMLS* **72**, 383-399
247. Muller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., Bao, G., Cathomen, T., and Mussolino, C. (2016) *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Molecular Therapy : The Journal of the American Society of Gene Therapy* **24**, 636-644
248. Steinert, J., Schiml, S., Fauser, F., and Puchta, H. (2015) Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *The Plant Journal: For cell and Molecular Biology* **84**, 1295-1305
249. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4576-4579
250. Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P., and de la Cruz, F. (2010) Mobility of plasmids. *Microbiology and Molecular Biology Reviews : MMBR* **74**, 434-452
251. Holmes, M. L., Pfeifer, F., and Dyall-Smith, M. L. (1995) Analysis of the halobacterial plasmid pHK2 minimal replicon. *Gene* **153**, 117-121
252. Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M., and Kristjansson, J. K. (1993) Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Systematic and Applied Microbiology* **16**, 609-628
253. Groisman, E. A., and Ochman, H. (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**, 791-794
254. Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., and Rocha, E. P. (2009) Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Current Biology : CB* **19**, 1683-1691
255. Stouthamer, A. H., and Kooijman, S. A. L. M. (1993) Why it pays for bacteria to delete disused DNA and to maintain megaplasmids. *Antonie van Leeuwenhoek* **63**, 39-43

256. Romanchuk, A., Jones, C. D., Karkare, K., Moore, A., Smith, B. A., Jones, C., Dougherty, K., and Baltrus, D. A. (2014) Bigger is not always better: Transmission and fitness burden of ~1MB *Pseudomonas syringae* megaplasmid pMPPla107. *Plasmid* **73**, 16-25
257. Ohtani, N., Tomita, M., and Itaya, M. (2016) Curing the megaplasmid pTT27 from *Thermus thermophilus* HB27 and maintaining exogenous plasmids in the plasmid-free strain. *Applied and Environmental Microbiology* **82**, 1537-1548
258. Oresnik, I. J., Liu, S. L., Yost, C. K., and Hynes, M. F. (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. *Journal of Bacteriology* **182**, 3582-3586
259. Pelicic, V., Reyrat, J. M., and Gicquel, B. (1996) Expression of the *Bacillus subtilis* sacB gene confers sucrose sensitivity on mycobacteria. *Journal of Bacteriology* **178**, 1197-1199
260. Hynes, M. F., Quandt, J., O'Connell, M. P., and Pühler, A. (1989) Direct selection for curing and deletion of *Rhizobium* plasmids using transposons carrying the *Bacillus subtilis* sacB gene. *Gene* **78**, 111-120
261. Ni, B., Du, Z., Guo, Z., Zhang, Y., and Yang, R. (2008) Curing of four different plasmids in *Yersinia pestis* using plasmid incompatibility. *Letters in Applied Microbiology* **47**, 235-240
262. Trevors, J. T. (1986) Plasmid curing in bacteria. *FEMS Microbiology Reviews* **1**, 149-157
263. Sheng, L., Zhang, Y., and Minton, N. P. (2016) Complete genome sequence of *Geobacillus thermoglucosidasius* NCIMB 11955, the progenitor of a bioethanol production strain. *Genome Announcements* **4**
264. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* **39**, 9275-9282
265. Gentz, R., Langner, A., Chang, A. C., Cohen, S. N., and Bujard, H. (1981) Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4936-4940
266. Phue, J. N., Lee, S. J., Kaufman, J. B., Negrete, A., and Shiloach, J. (2010) Acetate accumulation through alternative metabolic pathways in ackA (-) pta (-) poxB (-) triple mutant in *E. coli* B (BL21). *Biotechnology Letters* **32**, 1897-1903
267. Kuit, W., Minton, N. P., Lopez-Contreras, A. M., and Eggink, G. (2012) Disruption of the acetate kinase (*ack*) gene of *Clostridium acetobutylicum* results in delayed acetate production. *Applied Microbiology and Biotechnology* **94**, 729-741
268. Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., Hogsett, D. A., and Lynd, L. R. (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
269. Mougiakos, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., Finger Bou, M., Naduthodi, M. I. S., Gussak, A., Brinkman, R. B. L., van Kranenburg, R., and van der Oost, J. (2017b) Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature Communications* **8**, 1647
270. Gerdes, K., Rasmussen, P. B., and Molin, S. (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-

free cells. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3116-3120
271. Van Melderen, L., and Saavedra De Bast, M. (2009) Bacterial toxin–antitoxin systems: more than selfish entities? *PLoS Genetics* **5**, e1000437
272. Thisted, T., Sorensen, N. S., Wagner, E. G., and Gerdes, K. (1994) Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region. *EMBO J* **13**, 1960-1968
273. Tsilibaris, V., Maenhaut-Michel, G., Mine, N., and Van Melderen, L. (2007) What is the benefit to *Escherichia coli* of having multiple toxin-antitoxin systems in its genome? *Journal of Bacteriology* **189**, 6101-6108
274. Van Melderen, L. (2010) Toxin-antitoxin systems: why so many, what for? *Current Opinion in Microbiology* **13**, 781-785
275. Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D. A. (2007) Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Molecular Microbiology* **63**, 1588-1605
276. Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I., and Hazan, R. (2006) Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS genetics* **2**, e135
277. Hazan, R., and Engelberg-Kulka, H. (2004) *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Molecular Genetics and Genomics : MGG* **272**, 227-234
278. Habgood, R. (2018) Investigating the potential of producing alkanes and other fatty acid-derived biofuels using the thermophilic chassis *Geobacillus thermoglucosidasius*.
279. Wang, D., Gao, Z., Wang, H., Feng, E., Zhu, L., Liu, X., and Wang, H. (2015) Curing both virulent mega-plasmids from *Bacillus anthracis* wild-type strain A16 simultaneously using plasmid incompatibility. *Journal of Microbiology and Biotechnology* **25**, 1614-1620
280. Leak, D. J. (2018) Personal Communication.
281. Milunovic, B., diCenzo, G. C., Morton, R. A., and Finan, T. M. (2014) Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of *Sinorhizobium meliloti*. *Journal of Bacteriology* **196**, 811-824
282. Harayama, S., Lehrbach, P. R., and Timmis, K. N. (1984) Transposon mutagenesis analysis of meta-cleavage pathway operon genes of the TOL plasmid of *Pseudomonas putida* mt-2. *Journal of Bacteriology* **160**, 251-255
283. Harayama, S., Rekik, M., Ngai, K. L., and Ornston, L. N. (1989) Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via meta cleavage in *Pseudomonas putida*. *Journal of Bacteriology* **171**, 6251-6258
284. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J. (1991) Chemical and enzymic ketonization of 2-hydroxymuconate, a conjugated enol. *Journal of the American Chemical Society* **113**, 3154-3162
285. Gurujeyalakshmi, G., and Oriel, P. (1989) A thermophilic process for catechol production from phenol. *Biotechnology Letters* **11**, 689-694

286. Arenghi, F. L., Berlanda, D., Galli, E., Sello, G., and Barbieri, P. (2001) Organization and regulation of meta cleavage pathway genes for toluene and o-xylene derivative degradation in *Pseudomonas stutzeri* OX1. *Applied and Environmental Microbiology* **67**, 3304-3308
287. Burlage, R. S., Hooper, S. W., and Sayler, G. S. (1989) The TOL (pWW0) catabolic plasmid. *Applied and Environmental Microbiology* **55**, 1323-1328
288. Omokoko, B., Jantges, U. K., Zimmermann, M., Reiss, M., and Hartmeier, W. (2008) Isolation of the phe-operon from *G. stearothermophilus* comprising the phenol degradative meta-pathway genes and a novel transcriptional regulator. *BMC Microbiology* **8**, 197
289. Chen, L. H., Kenyon, G. L., Curtin, F., Harayama, S., Bembenek, M. E., Hajipour, G., and Whitman, C. P. (1992) 4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *Journal of Biological Chemistry* **267**, 17716-17721
290. Kirchner, U., Westphal, A. H., Muller, R., and van Berkel, W. J. (2003) Phenol hydroxylase from *Bacillus thermoglucosidasius* A7, a two-protein component monooxygenase with a dual role for FAD. *Journal of Biological Chemistry* **278**, 47545-47553
291. Kado, C. I. (1998) Origin and evolution of plasmids. *Antonie Van Leeuwenhoek* **73**, 117-126
292. Yarmolinsky, M. B. (1995) Programmed cell death in bacterial populations. *Science* **267**, 836-837
293. Cooper, T. F., and Heinemann, J. A. (2000) Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12643-12648
294. Jakociunas, T., Jensen, M. K., and Keasling, J. D. (2016) CRISPR/Cas9 advances engineering of microbial cell factories. *Metabolic Engineering* **34**, 44-59
295. Ranganathan, V., Wahlin, K., Maruotti, J., and Zack, D. J. (2014) Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nature Communications* **5**, 4516
296. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557-580
297. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183
298. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451
299. Li, X. T., Jun, Y., Erickstad, M. J., Brown, S. D., Parks, A., Court, D. L., and Jun, S. (2016) tCRISPRi: tunable and reversible, one-step control of gene expression. *Scientific Reports* **6**, 39076
300. Peters, J. M., Colavin, A., Shi, H., Czarny, T. L., Larson, M. H., Wong, S., Hawkins, J. S., Lu, C. H. S., Koo, B. M., Marta, E., Shiver, A. L., Whitehead, E. H., Weissman, J. S., Brown, E. D., Qi, L. S., Huang, K. C., and Gross, C. A. (2016) A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* **165**, 1493-1506

301. Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* **43**, D298-299
302. Oultram, J. D., Loughlin, M., Swinfield, T.-J., Brehm, J. K., Thompson, D. E., and Minton, N. P. (1988) Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiology Letters* **56**, 83-88
303. Scott, P. T., and Rood, J. I. (1989) Electroporation-mediated transformation of lysostaphin-treated *Clostridium perfringens*. *Gene* **82**, 327-333
304. Purdy, D., O'Keeffe, T. A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N. P. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology* **46**, 439-452
305. Herbert, M., O'Keeffe, T. A., Purdy, D., Elmore, M., and Minton, N. P. (2003) Gene transfer into *Clostridium difficile* CD630 and characterisation of its methylase genes. *FEMS Microbiology Letters* **229**, 103-110
306. Meaney, C. A., Cartman, S. T., McClure, P. J., and Minton, N. P. (2016) The role of small acid-soluble proteins (SASPs) in protection of spores of *Clostridium botulinum* against nitrous acid. *International Journal of Food Microbiology* **216**, 25-30
307. Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2013) Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnology for Biofuels* **6**, 50
308. Yang, X., Xu, M., and Yang, S. T. (2016) Restriction modification system analysis and development of *in vivo* methylation for the transformation of *Clostridium cellulovorans*. *Applied Microbiology and Biotechnology* **100**, 2289-2299
309. Cui, G. Z., Hong, W., Zhang, J., Li, W. L., Feng, Y., Liu, Y. J., and Cui, Q. (2012) Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *Journal of Microbiological Methods* **89**, 201-208
310. Croux, C., Nguyen, N. P., Lee, J., Raynaud, C., Saint-Prix, F., Gonzalez-Pajuelo, M., Meynial-Salles, I., and Soucaille, P. (2016) Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. *Biotechnology for Biofuels* **9**, 23
311. Lesiak, J. M., Liebl, W., and Ehrenreich, A. (2014) Development of an in vivo methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. *Journal of Biotechnology* **188**, 97-99
312. Mermelstein, L. D., and Papoutsakis, E. T. (1993) *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **59**, 1077-1081
313. Marinus, M. G., and Morris, N. R. (1973) Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *Journal of Bacteriology* **114**, 1143-1150
314. Dong, H., Zhang, Y., Dai, Z., and Li, Y. (2010) Engineering *Clostridium* strain to accept unmethylated DNA. *PloS One* **5**, e9038

315. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research* **16**, 6127-6145
316. Hanahan, D., Jessee, J., and Bloom, F. R. (1991) Plasmid transformation of *Escherichia coli* and other bacteria. in *Methods in Enzymology*, Academic Press. pp 63-113
317. Karyagina, A., Shilov, I., Tashlitskii, V., Khodoun, M., Vasil'ev, S., Lau, P. C., and Nikolskaya, I. (1997) Specific binding of sso II DNA methyltransferase to its promoter region provides the regulation of sso II restriction-modification gene expression. *Nucleic Acids Research* **25**, 2114-2120
318. Som, S., and Friedman, S. (1994) Regulation of EcoRII methyltransferase: effect of mutations on gene expression and *in vitro* binding to the promoter region. *Nucleic Acids Research* **22**, 5347-5353
319. Handa, N., Nakayama, Y., Sadykov, M., and Kobayashi, I. (2001) Experimental genome evolution: large-scale genome rearrangements associated with resistance to replacement of a chromosomal restriction-modification gene complex. *Molecular Microbiology* **40**, 932-940
320. Sadykov, M., Asami, Y., Niki, H., Handa, N., Itaya, M., Tanokura, M., and Kobayashi, I. (2003) Multiplication of a restriction-modification gene complex. *Molecular Microbiology* **48**, 417-427
321. Ohshima, H., Matsuoka, S., Asai, K., and Sadaie, Y. (2002) Molecular organization of intrinsic restriction and modification genes BsuM of *Bacillus subtilis* Marburg. *Journal of Bacteriology* **184**, 381-389
322. Iwai, M., Katoh, H., Katayama, M., and Ikeuchi, M. (2004) Improved genetic transformation of the thermophilic cyanobacterium, *Thermosynechococcus elongatus* BP-1. *Plant & Cell Physiology* **45**, 171-175
323. Kawabata, H., Norris, S. J., and Watanabe, H. (2004) BBE02 disruption mutants of *Borrelia burgdorferi* B31 have a highly transformable, infectious phenotype. *Infection and Immunity* **72**, 7147-7154
324. Chung, D., Farkas, J., and Westpheling, J. (2013) Overcoming restriction as a barrier to DNA transformation in *Caldicellulosiruptor* species results in efficient marker replacement. *Biotechnology for Biofuels* **6**, 82
325. Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S., and Butkus, V. (1992) Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Research* **20**, 6043-6049
326. Kobayashi, I. (2004) Restriction-Modification systems as minimal forms of life. in *Restriction Endonucleases* (Pingoud, A. M. ed.), Springer Berlin Heidelberg, Berlin, Heidelberg. pp 19-62
327. Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet (London, England)* **357**, 1225-1240

328. Schouler, C., Gautier, M., Ehrlich, S. D., and Chopin, M. C. (1998) Combinational variation of restriction modification specificities in *Lactococcus lactis*. *Molecular Microbiology* **28**, 169-178
329. Kong, H., Lin, L.-F., Porter, N., Stickel, S., Byrd, D., Posfai, J., and Roberts, R. J. (2000) Functional analysis of putative restriction–modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Research* **28**, 3216-3223
330. Lin, L.-F., Posfai, J., Roberts, R. J., and Kong, H. (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2740-2745
331. Kumar, R., Mukhopadhyay, A. K., and Rao, D. N. (2010) Characterization of an N6 adenine methyltransferase from *Helicobacter pylori* strain 26695 which methylates adjacent adenines on the same strand. *The FEBS Journal* **277**, 1666-1683
332. Oh, J. D., Kling-Backhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J., Mardis, E. R., Engstrand, L. G., and Gordon, J. I. (2006) The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9999-10004
333. Baldwin, D. N., Shepherd, B., Kraemer, P., Hall, M. K., Sycuro, L. K., Pinto-Santini, D. M., and Salama, N. R. (2007) Identification of *Helicobacter pylori* genes that contribute to stomach colonization. *Infection and Immunity* **75**, 1005-1016
334. Gonzalez, D., Kozdon, J. B., McAdams, H. H., Shapiro, L., and Collier, J. (2014) The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach. *Nucleic Acids Research* **42**, 3720-3735
335. Lobner-Olesen, A., Marinus, M. G., and Hansen, F. G. (2003) Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4672-4677
336. Lobner-Olesen, A., Skovgaard, O., and Marinus, M. G. (2005) Dam methylation: coordinating cellular processes. *Current Opinion in Microbiology* **8**, 154-160
337. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**, 827-832
338. Rocha, E. P., Danchin, A., and Viari, A. (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Research* **11**, 946-958
339. Seidel, R., Bloom, J. G. P., Dekker, C., and Szczelkun, M. D. (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J* **27**, 1388-1398
340. Kelleher, J. E., and Raleigh, E. A. (1994) Response to UV damage by four *Escherichia coli* K-12 restriction systems. *Journal of Bacteriology* **176**, 5888-5896
341. Roberts, G. A., Cooper, L. P., White, J. H., Su, T. J., Zipprich, J. T., Geary, P., Kennedy, C., and Dryden, D. T. (2011) An investigation of the structural requirements for ATP hydrolysis and DNA cleavage by the EcoKI Type I DNA restriction and modification enzyme. *Nucleic Acids Research* **39**, 7667-7676

342. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715
343. Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C., and Gehrke, C. W. (1985) DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Research* **13**, 1399-1412
344. van Heerden, C. D., and Nicol, W. (2013) Continuous succinic acid fermentation by *Actinobacillus succinogenes*. *Biochemical Engineering Journal* **73**, 5-11
345. Park, Y. S., Toda, K., Fukaya, M., Okumura, H., and Kawamura, Y. (1991) Production of a high concentration acetic acid by *Acetobacter aceti* using a repeated fed-batch culture with cell recycling. *Applied Microbiology and Biotechnology* **35**, 149-153
346. Narvhus, J. A., and Axelsson, L. (2003) Lactic acid bacteria. in *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (Caballero, B. ed.), Academic Press, Oxford. pp 3465-3472
347. Ciani, M., Comitini, F., and Mannazzu, I. (2013) Fermentation. in *Reference Module in Earth Systems and Environmental Sciences*, Elsevier. pp
348. Ahn, J. H., Jang, Y.-S., and Lee, S. Y. (2016) Production of succinic acid by metabolically engineered microorganisms. *Current Opinion in Biotechnology* **42**, 54-66
349. Forster, A. H., and Gescher, J. (2014) Metabolic engineering of *Escherichia coli* for production of mixed-acid fermentation end products. *Frontiers in Bioengineering and Biotechnology* **2**, 16
350. Basen, M., Schut, G. J., Nguyen, D. M., Lipscomb, G. L., Benn, R. A., Prybol, C. J., Vaccaro, B. J., Poole, F. L., 2nd, Kelly, R. M., and Adams, M. W. (2014) Single gene insertion drives bioalcohol production by a thermophilic archaeon. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17618-17623
351. Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., and Guss, A. M. (2015) Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum*. *Biotechnology for Biofuels* **8**, 20
352. Koendjbiharie, J. G., Wiersma, K., and van Kranenburg, R. (2018) Investigating the central metabolism of *Clostridium thermosuccinogenes*. *Applied and Environmental Microbiology* **84**
353. Pecsi, I., Hards, K., Ekanayaka, N., Berney, M., Hartman, T., Jacobs, W. R., Jr., and Cook, G. M. (2014) Essentiality of succinate dehydrogenase in *Mycobacterium smegmatis* and its role in the generation of the membrane potential under hypoxia. *mBio* **5**
354. Cook, G. M., Greening, C., Hards, K., and Berney, M. (2014) Energetics of pathogenic bacteria and opportunities for drug development. in *Advances in Microbial Physiology* (Poole, R. K. ed.), Academic Press. pp 1-62
355. Kubo, Y., Takagi, H., and Nakamori, S. (2000) Effect of gene disruption of succinate dehydrogenase on succinate production in a sake yeast strain. *Journal of Bioscience and Bioengineering* **90**, 619-624
356. Hirsch, C. A., Rasminsky, M., Davis, B. D., and Lin, E. C. (1963) A fumarate reductase in *Escherichia coli* distinct from succinate dehydrogenase. *Journal of Biological Chemistry* **238**, 3770-3774
357. Zhang, X., Jantama, K., Shanmugam, K. T., and Ingram, L. O. (2009) Reengineering *Escherichia coli* for succinate production in mineral salts medium. *Applied and Environmental Microbiology* **75**, 7807-7813

358. Cheng, K.-K., Wang, G.-Y., Zeng, J., and Zhang, J. (2013) Improved succinate production by metabolic engineering. *BioMed Research International* **2013**
359. Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J., and Root, D. E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* **32**, 1262-1267
360. Cui, Y., Xu, J., Cheng, M., Liao, X., and Peng, S. (2018) Review of CRISPR/Cas9 sgRNA Design Tools. *Interdisciplinary Sciences, Computational Life Sciences* **10**, 455-465
361. Rahman, M. K., and Rahman, M. S. (2017) CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. *PloS One* **12**, e0181943
362. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., and Root, D. E. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**, 184-191
363. Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J., and Takumi, T. (2016) Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Scientific Reports* **6**, 19675
364. Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Akhmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C., and Schier, A. F. (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS One* **9**, e98186
365. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., and Schier, A. F. (2016) Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications* **7**, 11750
366. Yang, Y. J., Wang, Y., Li, Z. F., Gong, Y., Zhang, P., Hu, W. C., Sheng, D. H., and Li, Y. Z. (2017) Increasing on-target cleavage efficiency for CRISPR/Cas9-induced large fragment deletion in *Myxococcus xanthus*. *Microbial Cell Factories* **16**, 142
367. Wong, N., Liu, W., and Wang, X. (2015) WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biology* **16**, 218
368. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., and Joung, J. K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279
369. Mougiakos, I. (2018) Personal Communication.
370. McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* **36**, 344-355
371. Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. C. (2000) *Introduction to Genetic Analysis. 7th edition*,
372. von Ballmoos, C., Adelroth, P., Gennis, R. B., and Brzezinski, P. (2012) Proton transfer in ba(3) cytochrome c oxidase from *Thermus thermophilus*. *Biochimica et Biophysica Acta* **1817**, 650-657
373. Xiao, Z., and Xu, P. (2007) Acetoin metabolism in bacteria. *Critical Reviews in Microbiology* **33**, 127-140
374. Krieg, N. R., and Padgett, P. J. (2011) Phenotypic and physiological characterization methods. in *Methods in Microbiology* (Rainey, F., and Oren, A. eds.), Academic Press. pp 15-60
375. Härtig, E., and Jahn, D. (2012) Regulation of the anaerobic metabolism in *Bacillus subtilis*. in *Advances in Microbial Physiology* (Poole, R. K. ed.), Academic Press. pp 195-216

376. Ludwig, B. (1987) Cytochrome c oxidase in prokaryotes. *FEMS Microbiology Letters* **46**, 41-56
377. Clark, D. P., and Cronan, J. E., Jr. (1980) Acetaldehyde coenzyme A dehydrogenase of *Escherichia coli*. *Journal of Bacteriology* **144**, 179-184
378. Rudolph, F. B., Purich, D. L., and Fromm, H. J. (1968) Coenzyme A-linked aldehyde dehydrogenase from *Escherichia coli*. *Journal of Biological Chemistry* **243**, 5539-5545
379. Shone, C. C., and Fromm, H. J. (1981) Steady-state and pre-steady-state kinetics of coenzyme A-linked aldehyde dehydrogenase from *Escherichia coli*. *Biochemistry* **20**, 7494-7501
380. Kessler, D., Leibrecht, I., and Knappe, J. (1991) Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*. *FEBS Letters* **281**, 59-63
381. Reid, M. F., and Fewson, C. A. (1994) Molecular characterization of microbial alcohol dehydrogenases. *Critical Reviews in Microbiology* **20**, 13-56
382. Membrillo-Hernandez, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., and Lin, E. C. (2000) Evolution of the *adhE* gene product of *Escherichia coli* from a functional reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. *Journal of Biological Chemistry* **275**, 33869-33875
383. Wood, B. J. B. H., W. H. (1995) *Lactic Acid Bacteria*, Springer
384. Singh, A., Cher Soh, K., Hatzimanikatis, V., and Gill, R. T. (2011) Manipulating redox and ATP balancing for improved production of succinate in *E. coli*. *Metabolic Engineering* **13**, 76-81
385. Liu, R., Liang, L., Chen, K., Ma, J., Jiang, M., Wei, P., and Ouyang, P. (2012) Fermentation of xylose to succinate by enhancement of ATP supply in metabolically engineered *Escherichia coli*. *Applied Microbiology and Biotechnology* **94**, 959-968
386. Lee, S. Y., Hong, S. H., and Moon, S. Y. (2002) *In silico* metabolic pathway analysis and design: succinic acid production by metabolically engineered *Escherichia coli* as an example. *Genome Informatics. International Conference on Genome Informatics* **13**, 214-223
387. Beauprez, J. J., Foulquié-Moreno, M. R., Maertens, J., van Horen, E., Bekers, K., Baart, G. J. E., Cunin, R. M., Charlier, D., Heijnen, J. J., and Soetaert, W. (2011) Influence of C4-dicarboxylic acid transporters on succinate production. *Green Chemistry* **13**, 2179-2186
388. Giepmans, B. N., Adams, S. R., Ellisman, M. H., and Tsien, R. Y. (2006) The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224
389. Day, R. N., Periasamy, A., and Schaufele, F. (2001) Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus. *Methods* **25**, 4-18
390. Truong, K., and Ikura, M. (2001) The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Current Opinions in Structural Biology* **11**, 573-578
391. Sekar, R. B., and Periasamy, A. (2003) Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *Journal of Cell Biology* **160**, 629-633
392. Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8**, 2180-2196

Chapter 9

Supplementary Data

Supplementary Data 9.1 – A table of oligonucleotide primers used in this study.

Name of Primer	Sequence (5' – 3')	Description
General Primers		
M13F	TGTAAAACGACGGCCAGT	M13 sequencing primer
M13R	AGGAAACAGCTATGACCAT	M13 sequencing primer
sgRNA_R	caatgctagcAAAAGGCCATCCGT	The reverse primer for changing any sgRNA sequence in pCas91 and pCas93
Cas_F	GCCAGGCATCAAACTAAGC	Sequencing primer for both Cas9 genes
Cas_R	CATTATATTGAGGGAGGATGAATGCA	Sequencing primer for both Cas9 genes
RepB-R1	TGCTGTTTTATCCTTTACCTTGTCT	Sequencing primer for the gram-positive replicon
sgRNA_F	CAAACAACGGGAGCACTCAG	Sequencing primer for the sgRNA
ColE1+tra-F1	CTGTGGATAACCGTATTACC	Sequencing primer for the gram-negative replicon
PyrE F Genome	CCATTATTTAAAAGATCAGGGACACCGG	Sequencing primer for *pyrE* on the genome
PyrE R Genome	GATTTTCCTTCAAAACCATTACTAAAGGATTATAACG	Sequencing primer for *pyrE* on the genome

sGFP_Seq_F	CCGGGCTAATTAGGGGGTGTCGCCC	Sequencing primer for sGFP
sGFP_Seq_R	GCATTTAGAATAAATTTTGTGTCGCCC	Sequencing primer for sGFP
Fluorescent Gene Primers		
sGFP_XbaI_F	gatctctagaATGAGCAAAGGCGAGGAACTGT	Amplification of sGFP
sGFP_HindIII_R	gatcaagcttTTAATGGTGATGGTGGTGATGGCTAC	Amplification of sGFP
GAF_XbaI_F	gatctctagaATGGACACCGAAACGTGGGC	Amplification of GAF
GAF_HindIII_R	gatcaagcttTTAAAAGTGAATGCGCAGACGCG	Amplification of GAF
YtvA_XbaI_F	gatctctagaATGGCGAGCTTTCAATCCTTCGG	Amplification of YtvA LOV
YtvA_HindIII_R	gatcaagcttTTAGTGGTGGTGATGATGATGGTGATGC	Amplification of YtvA LOV
eCGP_XbaI_F	gatctctagaATGGGCGCACACGCG	Amplification of eCGP123
eCGP_HindIII_R	gatcaagcttTTAGTGATGATGGTGGTGGTGCGT	Amplification of eCGP123
Dasher_XbaI_F	gatctctagaATGGCTTCTTTAGTTAAGAAAGACATGTGC	Amplification of Dasher
Dasher_HindIII_R	gatcaagcttCTAATGGTGACCTAACTTTGATGGCG	Amplification of Dasher
XhoI_RFP_F	Cacaattcctcgag GGTTTCCCAGGTAAATGG	Amplification of RFP

HindIII_RFP_R	ccgggaagcttAGGGGGTGTCGCCC	Amplification of RFP
Promoter Library Primers		
Gpi_2_F_NotI	gatcgcggccgcTATTAAAAGAAGCGGCGATGGCTG	Amplification of the promoter of Glucose-6-phosphate isomerase
Gpi_2_R_EcoRI	gatcgaattcGTTCTTCCCCTCTTACACTTTAGCG	Amplification of the promoter of Glucose-6-phosphate isomerase
Fbpa_F_NotI	gatcgcggccgcGAAAGTTGTCACGTTTTCCTTGAGTTACG	Amplification of the promoter of Fructose-bisphosphate aldolase
Fbpa_R_EcoRI	gatcgaattcACGTCATCGCCTTTCTTAAAATAGCT	Amplification of the promoter of Fructose-bisphosphate aldolase
Gapd_F_NotI	gatcgcggccgcGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATCA	Amplification of the promoter of Glyceraldehyde 3-phosphate dehydrogenase
Gapd_R_EcoRI	gatcgaattcATGCTATACTAATTAACAGCATTTCACAAAAA	Amplification of the promoter of Glyceraldehyde 3-phosphate dehydrogenase
Gpd_F_NotI	gatcgcggccgcATGATGCAAGCGGAGCAAGG	Amplification of the promoter of Glycero-3-phosphate dehydrogenase

Gpd_R_EcoRI	gatcgaattcTGACTAACTATTAACTTAGTTTCATTATAACGCAAAACAG	Amplification of the promoter of Glycero-3-phosphate dehydrogenase
Pfk_F_NotI	gatcgcggccgcAAGCAAGGCTGTCTCATCTTAGAGAC	Amplification of the promoter of 6-Phosphofructokinase
Pfk_R_EcoRI	gatcgaattcCATTATTTCACAAATCATCAAAATGTCCGACC	Amplification of the promoter of 6-Phosphofructokinase
Pgk_F_NotI	gatcgcggccgcTTGATGTGATTGTGTTTATCGCTTCATCA	Amplification of the promoter of Phosphoglycerate kinase
Pgk_R_EcoRI	gatcgaattcTGTGATTTTTGACAAACGAAGAGGAGTG	Amplification of the promoter of Phosphoglycerate kinase
CTS_F_NotI	gactgcggccgcGATGTCCTATTTTTAAGTTATTTTTGTGTAATATAAAGAATATCAGTGG	Amplification of the promoter of Citrate synthase
CTS_R_EcoRI	gatcgaattcAATTCCCCTTCTCCCAATTGCC	Amplification of the promoter of Citrate synthase
ldh_F _NotI	gatcgcggcgccCCTTGCACCATGGATCCGC	Amplification of the promoter of Lactate dehydrogenase
ldh_R_EcoRI	gatcgaattcCTCAATATAATGCGAACACTTTCATATAAATATTTTACGCG	Amplification of the promoter of Lactate dehydrogenase
RBS Library Primers		
R1_SDM_F	aggaggtattATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R1

R1_SDM_R	tttaaaaacaATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R1
R2_SDM_F	aggtgctactATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R2
R2_SDM_R	cctttatataATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R2
R3_SDM_F	aggtgttgatATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R3
R3_SDM_R	cctttttttaATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R3
R4_SDM_F	gaggttatatATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R4
R4_SDM_R	ctttatttatATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R4
R5_SDM_F	ggtataagctATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R5

R5_SDM_R	tcctttaaaaATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R5
R6_SDM_F	aggtgagtatATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R6
R6_SDM_R	cctttatttcATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R6
R7_SDM_F	aggtggcattATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R7
R7_SDM_R	ccttttatgtATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R7
R8_SDM_F	gttttatattATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R8
R8_SDM_R	ctcctttaagATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R8
R9_SDM_F	aggtagaactATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R9

R9_SDM_R	cctttatctaATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R9
R10_SDM_F	gtatttaaatATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R10
R10_SDM_R	ctcctttattATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R10
R11_SDM_F	aggtgagtatATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R11
R11_SDM_R	cctttattccATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R11
R12_SDM_F	gaggtgttttATGAGCAAAGGCGAGGAAC	Primer for site directed mutagenesis of RBS sequences to produce R12
R12_SDM_R	ctttgcttttATAAGAATTCCTCAATATAATGCGAAC	Primer for site directed mutagenesis of RBS sequences to produce R12
Terminator Library Primers		
TpepN_F	gatcgaattctaatttataaataaaaatcacctttagaggtggttttttttatttataa	Terminator primers of TpepN. Fw and Rv primers were

	attagatcttatattgaaggaggatgtctagagatc	annealed to produce double stranded constructed before cloning into the modular vector
TpepN_R	gatctctagacatcctccttcaatataagatctaatttataaataaaaaaaccacctc taaaaggtgatttttatttataaattagaattcgatc	Terminator primers of TpepN. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TslpA_LA_F	gatcgaattctgaaaaaggcagagcgaaagctctgtctttttttgatcttatattgaag gaggatgtctagagatc	Terminator primers of TslpA_LA. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TslpA_LA_R	gatctctagacatcctccttcaatataagatcaaaaaagacagagctttcgctctgc ctttttcagaattcgatc	Terminator primers of TslpA_LA. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TslpA_CD_F	gatcgaattcaaatataaaaagacttctcagatgagaagtctttttttgtgaaagatc ttatattgaaggaggatgtctagagatc	Terminator primers of TslpA_CD. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TslpA_CD_R	gatctctagacatcctccttcaatataagatctttcacaaaaaagacttctcatctga gaagtctttttatatttgaattcgatc	Terminator primers of TslpA_CD. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector

EcoT1_F	gatcgaattcccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcgatcttatattgaaggaggatgtctagagatc	Terminator primers of EcoT1. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
EcoT1_R	gatctctagacatcctccttcaatataagatcgagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggaattcgatc	Terminator primers of EcoT1. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
EcoT2_F	gatcgaattcattaagcagaaggccatcctgacggatggccttttgcgtttcgatcttatattgaaggaggatgtctagagatc	Terminator primers of EcoT2. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
EcoT2_R	gatctctagacatcctccttcaatataagatcgaaacgcaaaaaggccatccgtcaggatggccttctgcttaatgaattcgatc	Terminator primers of EcoT2. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
phiTD1_F	gatcgaattcaacaatcaaaagaaaagcctatcgtctgaggaacggtaggctctttgtagcatatagttggatcttatattgaaggaggatgtctagagatc	Terminator primers of phiTD1. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
phiTD1_R	gatctctagacatcctccttcaatataagatccaactatatgctacaaaagagcctaccgttcctcagacgataggcttttcttttgattgttgaattcgatc	Terminator primers of phiTD1. Fw and Rv primers were annealed to produce double stranded constructed before

		cloning into the modular vector
TtyrS_F	gatcgaattcataatcaatcgtcccttcgtgtaaacgaaggggcgtttttttatttgatcttatattgaaggaggatgtctagagatc	Terminator primers of TtyrS. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TtyrS_R	gatctctagacatcctccttcaatataagatcaaataaaaaacgccccttcgtttacacgaagggacgattgattatgaattcgatc	Terminator primers of TtyrS. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TtyrA_F	gatcgaattcaagaagaagtgtgaaaaagcgcagctgaaatagctgcgcttttttgtgtcataagatcttatattgaaggaggatgtctagagatc	Terminator primers of TtyrA. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
TtyrA_R	gatctctagacatcctccttcaatataagatcttatgacacaaaaaagcgcagctattcagctgcgctttttcacacttcttcttgaattcgatc	Terminator primers of TtyrA. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
Fdx_F	gatcgaattcataaaaataagaagcctgcatttgcaggcttcttatttttatgatcttatattgaaggaggatgtctagagatc	Terminator primers of Fdx. Fw and Rv primers were annealed to produce double stranded constructed before cloning into the modular vector
Fdx_R	gatctctagacatcctccttcaatataagatcataaaaataagaagcctgcaaatgc	Terminator primers of Fdx. Fw and Rv primers were

	aggcttcttatttttatgaattcgatc	annealed to produce double stranded constructed before cloning into the modular vector
Riboswitch Library Primers		
Gapd_F_NotI	gatcgcggccgcGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATCA	Amplification of the promoter of Glyceraldehyde 3-phosphate dehydrogenase
Fbpa_F_NotI	gatcgcggccgcGAAAGTTGTCACGTTTTCCTTGAGTTACG	Amplification of the promoter of Fructose-bisphosphate aldolase
PFK_F_NotI	gatcgcggccgcAAGCAAGGCTGTCTCATCTTAGAGAC	Amplification of the promoter of 6-Phosphofructokinase
Gapd_R_BsaI	cccggtctctattgATGCTATACTAATTAACAGCATTTCACAAAAATAGTATATC	Reverse primer to amplify the promoter of Glyceraldehyde 3-phosphate dehydrogenase for cloning with the riboswitches
Fbpa_R_BsaI	cccggtctctattgACGTCATCGCCTTTCTTAAAATAGCTATGC	Reverse primer to amplify the promoter of Fructose-bisphosphate aldolase for cloning with the riboswitches
PFK_R_BsaI	cccggtctctattgCATTATTTCACAAATCATCAAAATGTCCGACC	Reverse primer to amplify the promoter of 6-Phosphofructokinase for cloning with the riboswitches
Rbx_F_BsaI	ggtctcaCAATACGACTCACTATAGGTGATACCAGC	General forward primer to amplify the riboswitches

RbE_R_BsaI	ccggtctcctcatCTTGTTGTTACCTCCTTAGCAGGG	Forward primer for amplification of riboswitch E
RbF_R_BsaI	ccggtctcctcatTAAGTTACCTCCTTAGCAGGGTGC	Forward primer for amplification of riboswitch F
RbG_R_BsaI	ccggtctcctcatTAACACACCTCCTTAGCAGGGTG	Forward primer for amplification of riboswitch G
sGFP_F_BsaI	tataggtctcaATGAGCAAAGGCGAGGAACTG	Forward primer for amplification of sGFP for cloning with the riboswitches
sGFP_R_XhoI	atatctcgagttaGCTACCCTTATACAATTCGTCCATACCATG	Reverse primer for amplification of sGFP
Cloning of Cas9 Vector Primers		
Cas9-1_F	ATGAGTGACTTAGTTTTAGGACTTGATATCGG	Amplification of *stcas9-1*
Cas9-1_NotI_R	gcggccgTTAAAAATCTAGCTTAGGCTTATCACCCTC	Amplification of *stcas9-1*
Cas9-3_F	ATGACTAAGCCATACTCAATTGGACTTG	Amplification of *stcas9-3*
Cas9-3_NotI_R	gcggccgTTAACCCTCTCCTAGTTTGGCAAGG	Amplification of *stcas9-3*
Cas9_1_F_ldl_ovl	ttcgcattatattgagggaggatgaatgcaATGAGTGACTTAGTTTTAG GACTTGATATCGG	Primer for amplification of *stcas9-1* for spliced-overlap extension (SOE) PCR with P_{ldh}
Cas9-1_R_SbfI	gatccctgcaggTTAAAAATCTAGCTTAGGCTTATCACCCTC	Amplification of *stcas9-1*

LDHrbs_F_XbaI	gatctctagaGAATTCGCGGCCGCACT	Amplification of P_{ldh}
LDHrbs_R_Cas91_ovl	ccgatatcaagtcctaaaactaagtcactcatTGCATTCATCCTCCCTCAATATAATGC	Primer for amplification of P_{ldh} for SOE PCR with *stcas9-1*
Gapd_F_XbaI	gatctctagaGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATCA	Amplification of P_{gapd}
Gapd_R_C1ack1_ovl	gtacaaaaacccattgacggtgatcgtaaaATGCTATACTAATTAACAGCATTTCACAAAAA	Primer for amplification of P_{gapd} for SOE PCR with sgRNA1 for *ack* using *stcas9-1*
C1gRNAack1_F_gapd_ovl	cgaaatatgatatactatttttgtgaaatgctgttaattagtatagcatTTTACGATCACCGTCAATGGGTTTTTG	Primer for amplification of sgRNA1 for *ack* using *stcas9-1* for SOE PCR with P_{gapd}
C1gRNAack1_R_SalI	gactgtcgacAAAACACCCTGCCATAAAATGACAGG	Amplification of P_{gapd} sgRNA1 for *ack* using *stcas9-1*
Cas9_3_F_ldl_ovl	ttcgcattatattgagggaggatgaatgcaATGACTAAGCCATACTCAATTGGACTTG	Primer for amplification of *stcas9-3* for SOE PCR with P_{ldh}
Cas9-3_R_SbfI	gatccctgcaggTTAACCCTCTCCTAGTTTGGCAAGG	Amplification of *stcas9-3*
LDHrbs_R_Cas93_ovl	atatcaagtccaattgagtatggcttagtcatTGCATTCATCCTCCCTCAATATAATGC	Primer for amplification of P_{ldh} for SOE PCR with *stcas9-3*
Gapd_R_C3ack1_ovl	gctctaaaaccgttgtgcatggtggcgaaaATGCTATACTAATTAACAG	Primer for amplification of P_{gapd} for SOE PCR with sgRNA1

	CATTTCACAAAAA	for *ack* using *stcas9-3*
C3gRNAack1_F_gapd_ovl	cgaaatatgatatactatttttgtgaaatgctgttaattagtatagcatTTTCGCC ACCATGCACAACG	Primer for amplification of sgRNA1 for *ack* using *stcas9-3* for SOE PCR with P_{gapd}
C3gRNAack1_R_SalI	gatcgtcgacAAAAACACCGAATCGGTGCCAC	Amplification of P_{gapd} sgRNA1 for *ack* using *stcas9-3*
Ack KO Primers		
Ack_CrisprLF	gatccgcgcgcgCATACGCCCACCATTTGTTCATTG	Amplification of *ack* left homology arm
Ack_CrisprLR	CACGTTTTTGTCAAAATTAAGCCATATTCCAAACAACTCCCT TAATCATC	Amplification of *ack* left homology arm
Ack_CrisprLF	GGGAGTTGTTTGGAATATGGCTTAATTTTGACAAAAACGTG ACCTAAATC	Amplification of *ack* right homology arm
Ack_CrisprLF	gatcggcgcgccGAATCAAGCATTCGAAAATATTACATCAGG	Amplification of *ack* right homology arm
Ack F Genome	GCTATCATTTTGTTTAATGAATGGATCATGCG	Primers for screening of the *ack* locus
Ack R Genome	GCTTTGTCTTAAAGGCGGAAGAAGG	Primers for screening of the *ack* locus
Megaplasmid KO Primers		
St3_BMPR_1_BamHI_F	gatcggatccGTGGCTGGTGAAAAGCACTGGTTTTAGAGCTGTG	Primers for annealed oligonucleotides to produce the

	TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	sgRNA1 for *repB*
St3_BMPR_1_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA CAAACAACACAGCTCTAAAACCAGTGCTTTTCAC	Primers for annealed oligonucleotides to produce the sgRNA1 for *repB*
St3_BMPR_2_BamHI_F	gatcggatccACTTTTCTCGCTCCTTGACGGTTTTAGAGCTGTG TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	Primers for annealed oligonucleotides to produce the sgRNA2 for *repB*
St3_BMPR_2_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA CAAACAACACAGCTCTAAAACCGTCAAGGAGCGA	Primers for annealed oligonucleotides to produce the sgRNA2 for *repB*
St3_SMPR_1_BamHI_F	gatcggatccTGGTTTCCTATAGAAACCGGGTTTTAGAGCTGTG TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	Primers for annealed oligonucleotides to produce the sgRNA1 for *repA*
St3_SMPR_1_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA CAAACAACACAGCTCTAAAACCCGGTTTCTATAG	Primers for annealed oligonucleotides to produce the sgRNA1 for *repA*

St3_SMPR_2_BamHI_F	gatcggatccTGGTTTCTGTAGGAAACCAGGTTTTAGAGCTGTG TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	Primers for annealed oligonucleotides to produce the sgRNA2 for *repA*
St3_SMPR_2_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA CAAACAACACAGCTCTAAAACCTGGTTTCCTACA	Primers for annealed oligonucleotides to produce the sgRNA2 for *repA*
St3_Tox_1_BamHI_F	gatcggatccCTATTACCATTCCATCGCGAGTTTTAGAGCTGTG TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	Primers for annealed oligonucleotides to produce the sgRNA1 for *tox*
St3_Tox_1_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA CAAACAACACAGCTCTAAAACTCGCGATGGAATG	Primers for annealed oligonucleotides to produce the sgRNA1 for *tox*
St3_Tox_2_BamHI_F	gatcggatccGTACGGTGATGCCGAACAAGGTTTTAGAGCTGTG TTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAG TCCGTACTCAACTTGAAAAGGTGGCACCGATTCGG	Primers for annealed oligonucleotides to produce the sgRNA2 for *tox*
St3_Tox_2_ SalI _R	gatcgtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTG AGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAA	Primers for annealed oligonucleotides to produce the sgRNA2 for *tox*

	CAAACAACACAGCTCTAAAACCTTGTTCGGCATC	
BMPR_HA_LF_NheI	gcatgctagcCGAGCTGTTGGTACTTGGTC	Amplification of *repB* left homology arm
BMPR_HA_LR	GGGGGATAGAGAGGTAATCAACACGCCCTTTTTCAGAGTAATGG	Amplification of *repB* left homology arm
BMPR_HA_RF	CATTACTCTGAAAAAGGGCGTGTTGATTACCTCTCTATCCCCC	Amplification of *repB* right homology arm
BMPR_HA_RR_AscI	gtacggcgcgccGTTTATAACTTGTTTATACTTGTTTACGGG	Amplification of *repB* right homology arm
SMPR_HA_LF_NheI	gctagctagcGCAAACTGGGATAAGATCAATACC	Amplification of *repA* left homology arm
SMPR_HA_LR	GGGAAACAGGAGGAGAATATATCCGATGAGACTAGCATCGATAAG	Amplification of *repA* left homology arm
SMPR_HA_RF	ATCGATGCTAGTCTCATCGGATATATTCTCCTCCTGTTTCCCC	Amplification of *repA* right homology arm
SMPR_HA_RR_AscI	gtacggcgcgccCAACCAACGACTGGAAGAACG	Amplification of *repA* right homology arm
Tox_HA_LF_NheI	gcatgctagcCCATAGCTTCACATTTTTTGGAAGA	Amplification of *tox* left homology arm

Tox_HA_LR	TTTTGCCGCCGTCTTTCGTTTTCCCCTGACTGCTCC	Amplification of *tox* left homology arm
Tox_HA_RF	GAGCAGTCAGGGGAAAACGAAAGACGGCGGCAAAAG	Amplification of *tox* right homology arm
Tox_HA_RR_AscI	gtacggcgcgccAAACAAGAGTATTGAAAGGGTGTATTG	Amplification of *tox* right homology arm
BMPR_CP_F	GGTATGAATATAATAACTTTGATTATATTAGAAATGTCGAAGG	Primers for screening of the *repB* locus
BMPR_CP_R	CCATTGTTTCACTTGCTAAAGAGGAAGC	Primers for screening of the *repB* locus
SMPR_CP_F	GCACGAATTCTTCCCACATTTGAGG	Primers for screening of the *repA* locus
SMPR_CP_R	GCTAGATCAACTAATCGAGCGAATCG	Primers for screening of the *repA* locus
Tox_CP_F	GGGTTCATCAGAGCCTACAGG	Primers for screening of the *tox* locus
Tox_CP_R	CCTTAATGATGCTCTTGCTCTGC	Primers for screening of the *tox* locus
eCGP123 Genome KI Primers		
gRNA_Genome_F	ctaggatcctaaaatgaaaaatccaggaaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting the genome downstream of *pyrE*
gRNA2_Genome_F	ctaggatcctgctctttgttttcaaagatGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting the genome downstream of *pyrE*

gRNA3_Genome_F	ctaggatccagttcaaagaatcatttccaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA3 targeting the genome downstream of *pyrE*
PyrEeC_KI_LF	gctagctagcGCCGGCACGGCAACAG	Amplification of left homology arm for the KI of eCGP123 downstream of *pyrE*
PyrEeC_KI_LR	CAGTAAAATCATAAAAAAAGCCCGACTGCTACTAGAAAAATGAAACAATATCATG	Amplification of left homology arm for the KI of eCGP123 downstream of *pyrE*
eC_KI_PyrE_F	GTTTCATTTTTCTAGTAGCAGTCGGGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATCACC	Amplification of *eCGP123* for SOE PCR between the left and right homology arms
eC_KI_PyrE_R	CTATAACAAGTCACATAGAAGGTTAGTGATGATGGTGGTGGTGC	Amplification of *eCGP123* for SOE PCR between the left and right homology arms
PyrEeC_KI_RF	CCACCACCATCATCACTAACCTTCTATGTGACTTGTTATAGATAGTGAGTTTTTCC	Amplification of right homology arm for the KI of eCGP123 downstream of *pyrE*
PyrEeC_KI_RR	gctaggcgcgccCGTTGAATTGTTTTCTCATTTACTCCAAAACG	Amplification of right homology arm for the KI of eCGP123 downstream of *pyrE*
CP_Genome_eC_F	GCATTCGCACCCATTTTCC	Primers for screening of downstream of the *pyrE* locus
CP_Genome_eC_R	CCTTTGAATACATATCTAATTAATATATTACCATCG	Primers for screening of downstream of the *pyrE* locus

eCGP123 BMP KI Primers		
gRNA_Ps490F_1	gtacggatccGTAGCTGACTGCGCACAAAAGTTTTAGAGCTGTGTTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAGTCCGTACTCAACTTGAAAA	Primers for annealed oligonucleotides to produce the sgRNA1 for the pseudo gene BCV53_19490
gRNA_Ps490R_1	gtcagtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTGAGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAACAAACAACACAGCTCTAA	Primers for annealed oligonucleotides to produce the sgRNA1 for the pseudo gene BCV53_19490
gRNA_Ps490F_2	gtacggatccTGGCGGGAAAACGAAATACGGTTTTAGAGCTGTGTTGTTTGTTAAAACAACACAGCGAGTTAAAATAAGGCTTAGTCCGTACTCAACTTGAAAA	Primers for annealed oligonucleotides to produce the sgRNA2 for the pseudo gene BCV53_19490
gRNA_Ps490R_2	gtcagtcgacAAAAACACCGAATCGGTGCCACCTTTTCAAGTTGAGTACGGACTAAGCCTTATTTTAACTCGCTGTGTTGTTTTAACAAACAACACAGCTCTAA	Primers for annealed oligonucleotides to produce the sgRNA2 for the pseudo gene BCV53_19490
BMP_KI_LF	gctagctagcAATAGGCCAAAATTCCTGTTATTTTCATCCATG	Amplification of left homology arm for the KI of eCGP123 in the pseudo gene BCV53_19490
BMPe_KI_LR	CAGTAAAATCATAAAAAAAGCAATGAGGACATCGGTCGTGAAAACTG	Amplification of left homology arm for the KI of eCGP123 in the pseudo gene BCV53_19490

PreCBMP_KI_F	CACGACCGATGTCCTCATTGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATC	Amplification of *eCGP123* for SOE PCR between the left and right homology arms
PreCBMP_KI_R	CCTGATACGTTTGTAAAATCTGTTAGTGATGATGGTGGTGGTGCG	Amplification of *eCGP123* for SOE PCR between the left and right homology arms
BMPe_KI_RF	CGCACCACCACCATCATCACTAACAGATTTTACAAACGTATCAGGGCAAGTATG	Amplification of right homology arm for the KI of eCGP123 in the pseudo gene BCV53_19490
BMP_KI_RR	gctaggcgcgccTTCCATTCTCTATATAGGATTATGCTAATAACCGC	Amplification of right homology arm for the KI of eCGP123 in the pseudo gene BCV53_19490
CP_BMP_eC_F	CCAAAAGGCATGGATGCG	Primers for screening of the pseudo gene BCV53_19490
CP_BMP_eC_R	CCATCGAAACGAATTTTGTAGAAAAAGC	Internal primers of the pseudo gene BCV53_19490 for screening
CP_BMP_eC_F_2	CCTACGAAGATCAAGGTATTTGC	Primers for screening of the pseudo gene BCV53_19490
CP_BMP_eC_R_2	CGTTTAGAAGTAGGACTTCTCGG	Internal primers of the pseudo gene BCV53_19490 for screening
Restriction Modification KO Primers		

RM1_gRNA1_F	ctaggatcccgttggtttttaattaacaaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting RM1
RM1_gRNA2_F	ctaggatccgcaggtccactaattgtagaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting RM1
RM1_LF	gctagctagcCCATATTCTATATCTATATCTTTCGAATTTAAAGGTAATAACC	Amplification of RM1 left homology arm
RM1_LR	CAAAGGAGTCGAAACGTCCAGATATAAAACGGAAAGAGGTAACGATTCC	Amplification of RM1 left homology arm
RM1_RF	CCTCTTTCCGTTTTATATCTGGACGTTTCGACTCCTTTGATAAGC	Amplification of RM1 right homology arm
RM1_RR	gctaggcgcgccCCCTTGAAGAAAAGTTAGTCGAATTAAAAGATTATTACC	Amplification of RM1 right homology arm
CP_RM1_F	CGCCCGTTAATACAGCC	Primers for screening of the RM1 locus
CP_RM1_R	GGTAAATGAACAACGTAAGGC	Primers for screening of the RM1 locus
RM2_gRNA1_F	ctaggatcctctggctgtcaaccaaaacgGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting RM2

RM2_gRNA2_F	ctaggatcctctcgatacgtacgttcccgGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting RM2
RM2_LF	gctagctagcGGTTATAAAAACATTTGCGAAATCGGG	Amplification of RM2 left homology arm
RM2_LR	CCTCCTGTAATGCATCACGGTTACACACTCCTAACATCCG	Amplification of RM2 left homology arm
RM2_RF	GTTAGGAGTGTGTAACCGTGATGCATTACAGGAGGCTGTCC	Amplification of RM2 right homology arm
RM2_RR	gctaggcgcgccCGCACGCCCCACATAATTTAAAGG	Amplification of RM2 right homology arm
CP_RM2_F	GAGAAAACGGATGAAAAATCGG	Primers for screening of the RM2 locus
CP_RM2_R	TATCTTGTGCAACGAGTTTTGG	Primers for screening of the RM2 locus
RM3_gRNA1_F	ctaggatccagagagttagaagaaaaacaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting RM3
RM3_gRNA2_F	ctaggatccttttttaatgaaagaactagGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting RM3
RM3_LF	gctagctagcGGATTGAGTAGAAATCACAGAAAATGG	Amplification of RM3 left homology arm

RM3_LR	GTATCTTTAAAAGATAGTCTGCCCCCTACCTTATTCTCTATTATATCTTATTAATTACATTGATTAAC	Amplification of RM3 left homology arm
RM3_RF	GAGAATAAGGTAGGGGGCAGACTATCTTTTAAAGATACAAAATAGTCTATAGAGTTTTTCTC	Amplification of RM3 right homology arm
RM3_RR	gctaggcgcgccGCTTTATACGAGATGAGAAAAACATAAAAGGC	Amplification of RM3 right homology arm
CP_RM3_F	GGTGAAGACCCTTTTTTAAAACAAAGGATTG	Primers for screening of the RM3 locus
CP_RM3_R	GCAAATACGGTGATTCAGAAGTATTATCGTTC	Primers for screening of the RM3 locus
Succinate Production KO Primers		
ADH_gRNA1_F	ctaggatccgcagtagaagaaacaaaacgGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting *adhE*
ADH_gRNA2_F	ctaggatccgttaatgcgccatcttcgcaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting *adhE*
ADH_gRNA3_F	ctaggatcctgaaggatctggctctacttGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA3 targeting *adhE*
ADH_LF	gctagctagcGGAAGTGCTTTTGTTAATGGCGC	Amplification of *adhE* left homology arm

ADH_LR	CTTTGGGGGAGTGGACGCATTCTCCCTCCTGATTG	Amplification of *adhE* left homology arm
ADH_RF	GAGGGAGAATGCGTCCACTCCCCCAAAGAAAGAGG	Amplification of *adhE* right homology arm
ADH_RR	gctaggcgcgccGCATCGTTCCAGCAAAAATTAAATGAATTAC	Amplification of *adhE* right homology arm
CP_ADH_F	CCTTCCTCACGAAGAATTGACAAACC	Primers for screening of the *adhE* locus
CP_ADH_R	GATCGATTCATCGGGAGGATGC	Primers for screening of the *adhE* locus
ADH_Inter_F	ATGGCTGTGGAGGAGAGAGTC	Primers for internal screening of *adhE*
ADH_Inter_R	CGCATCCCAGCGTTAATGACG	Primers for internal screening of *adhE*
Cox_gRNA1_F	ctaggatcccattttctatggaaaatgcgGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting *cox*
Cox_gRNA2_F	ctaggatccgcgattgtaccgttgcaaatGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting *cox*
Cox_gRNA3_F	ctaggatccgccaaccgcaggaatccagaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA3 targeting *cox*
Cox_LF	gctagctagcGGTGATAGGATGGGCGGC	Amplification of *cox* left homology arm
Cox_LR	GATCCCCTTTTTGCAGCCTAATCAAACCCCTCTTTCTC	Amplification of *cox* left homology arm

Cox_RF	GGGGTTTGATTAGGCTGCAAAAAGGGGATCAGCTTTTTCCG	Amplification of *cox* right homology arm
Cox_RR	gctaggcgcgccCGAGAACAATCCGTTAAATAAAACGAGTGC	Amplification of *cox* right homology arm
CP_Cox_F	CGTTGAATACGGTTGTTGGAAGC	Primers for screening of the *cox* locus
CP_Cox_R	GCTTTTCACTAAATCAAAAACAAATGGAACG	Primers for screening of the *cox* locus
Cox_Inter_F	GCACTTCCTTGGCCTAATGGG	Primers for internal screening of *cox*
Cox_Inter_R	GGATTTTCAAAATTTGTGTGCTGCGC	Primers for internal screening of *cox*
LDH_gRNA1_F	ctagggatccatatacgcatgtacattggtGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting *ldh*
LDH_gRNA2_F	ctagggatccgcaggggctaaccaaaagccGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting *ldh*
LDH_gRNA3_F	ctagggatccgtaaataagaataaggcagaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA3 targeting *ldh*
LDH_LF	gctagctagcGCGCTATCCGATTGACATTTATCATCG	Amplification of *ldh* left homology arm
LDH_LR	GTTGTATTCAAAGTCAGTACGCTGTCTGTCATCCTTTCCAAAAAAT	Amplification of *ldh* left homology arm

LDH_RF	GGATGACAGACAGCGTACTGACTTTGAATACAACAAGGTGAACATCG	Amplification of *ldh* right homology arm
LDH_RR	gctaggcgcgccGCAGCATTGGCAATTAAGCAAAGC	Amplification of *ldh* right homology arm
CP_LDH_F	CGACGATCAGCTTCTGGTCC	Primers for screening of the *ldh* locus
CP_LDH_R	CCGACAATAACCGGAATTCCTATCG	Primers for screening of the *ldh* locus
LDH_Inter_F	ATGAAACAACAAGGCATGAATCGAGTAG	Primers for internal screening of *ldh*
LDH_Inter_R	CGGGAAAGAATGTCTTTTAATACAGTTACACTATG	Primers for internal screening of *ldh*
SDH_gRNA1_F	ctaggatccatttatccaaattcacccgaGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA1 targeting *sdh*
SDH_gRNA2_F	ctaggatccggtaaaagcaatgtgcgaggGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA2 targeting *sdh*
SDH_gRNA3_F	ctaggatcctcaagcatataaatgatgccGTTTTAGAGCTGTGTTGTTTGTTAAAAC	Amplification of sgRNA3 targeting *sdh*
SDH_LF	gctagctagcGCTGTTGACGTTTGGACATGC	Amplification of *sdh* left homology arm
SDH_LR	GGAGTGAGTGGCAACGATGAGCGAGAAAAAAACGG	Amplification of *sdh* left homology arm

SDH_RF	CTCGCTCATCGTTGCCACTCACTCCCTAAAAGATTGC	Amplification of *sdh* right homology arm
SDH_RR	gctaggcgcgccGCTTCTTTCATGGGGAATTTGCC	Amplification of *sdh* right homology arm
CP_SDH_F	GCGTATTAAAAAGGCGCACTTGG	Primers for screening of the *sdh* locus
CP_SDH_R	CGTAAACCATTTTGCGACAAAAGGG	Primers for screening of the *sdh* locus
SDH_Inter_F	CGGTTTTCAGCAACTTGTCGTTATAGC	Primers for internal screening of *sdh*
SDH_Inter_R	CGACTTTTTAGCAAACCAGCCTCC	Primers for internal screening of *sdh*