

**METABOLIC ENGINEERING OF THE THERMOPHILE *GEOBACILLUS*
TO PRODUCE THE ADVANCED BIOFUEL *N*-BUTANOL**

**By
Jennifer Spencer, BSc (Hons)**

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

2018

DECLARATION

Unless otherwise acknowledged, the work presented in this thesis is my own.
No part has been submitted for another degree at the University of
Nottingham or any other institute of learning.

Jennifer Spencer
February 2018

ABSTRACT

Limited fossil fuel resources and the environmental impacts of climate change are motivating the development of sustainable processes for the production of fuels and chemicals from renewable resources. The development of alternative energy sources, such as biofuels, will strengthen energy security and reduce dependence on fossil fuels. *n*-butanol is a biofuel and platform chemical. *n*-butanol is an advanced fuel with high energy content, compatible with existing infrastructure. Here the ability of microorganisms to use renewable resources for biofuel synthesis is exploited. In this work use of the thermophilic bacterium *Geobacillus thermoglucosidasius* is explored for the production of *n*-butanol. *Geobacillus* is considered a promising industrial process organism due to its high optimum growth temperature and ability to assimilate various substrates including both hexose and pentose sugars. As a relatively novel process organism, first the development of molecular tools was required to enable subsequent engineering of the host metabolism. Here four reporter assays were developed, three of which can be used simultaneously, providing for extensive analysis of qualitative and quantitative gene expression within the cell. A range of promoters and RBS' were screened. Extension of the *Geobacillus* vector series, pMTL60000, with new component parts for each module enabled co-transformation of two plasmids into *G. thermoglucosidasius*. The molecular tools developed were then applied in *Geobacillus* metabolic engineering work, with the aim of producing the target molecule *n*-butanol. Initially a CoA dependent *n*-butanol pathway, based on naturally occurring production by ABE fermentation, was considered. Following introduction of the pathway further metabolic engineering was employed to improve pathway flux, creating a driving force through the pathway and increasing the substrate pool. Production of 0.137 mM (10.166 mg/l) *n*-butanol demonstrated proof of concept. Next, the use of genes native to *Geobacillus* were investigated for improved enzyme compatibility. This approach did not generate *n*-butanol here. Finally a CoA independent pathway utilising the host's native fatty acid biosynthesis pathway was considered. Using this

approach, butyric acid was produced. Butyric acid can subsequently be further converted to *n*-butanol however this was not demonstrated here. In addition to metabolite pathway introduction, host strain engineering was carried out with the aim of adaptation towards industrially desired properties. Directed evolution resulted in selection of a strain with an increased *n*-butanol tolerance of 2.5% (v/v). Such modifications resulted in an improved process organism for biotechnological application.

This work provides the first reported production of *n*-butanol in thermophilic and aerobic conditions. Multiple approaches to *n*-butanol production are evaluated here. Use of heterologous and native genes are considered. Both CoA and ACP dependent pathways were introduced. Each approach presented advantages and drawbacks. A system compatible for use in *Geobacillus* has demonstrated proof of concept *n*-butanol production. Further development is required to increase production to industrially feasible quantities.

ACKNOWLEDGEMENTS

I wish to thank the following organisations for funding and supporting this research:

Biotechnology and Biological Sciences Research Council (BBSRC)

The University of Nottingham

The Synthetic Biology Research Centre (SBRC)

Newton Bhabha Fund, The British Council

International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

TABLE OF CONTENTS

DECLARATION	1
ABSTRACT	3
ACKNOWLEDGEMENTS.....	5
TABLE OF CONTENTS	7
List of Figures	15
List of Tables	21
Abbreviations.....	23
CHAPTER 1. INTRODUCTION.....	29
1.1 Prologue.....	30
1.2 The need for fossil fuel alternatives	30
1.2.1 Climate change	30
1.2.2 Finite resources	32
1.2.3 Energy security	32
1.3 Energy demand.....	32
1.4 Energy policies	34
1.5 Biofuels	34
1.5.1 Biofuel production pathways	35
1.5.1.1 Keto acid pathway	36
1.5.1.2 Isoprenoid pathway.....	38
1.5.1.3 Fatty acid biosynthesis pathway.....	38
1.5.1.4 CoA dependent pathway	39
1.5.2 Biofuel substrates	41
1.5.2.1 Plant biomass.....	43
1.5.2.2 Alternative substrates	45
1.5.3 <i>n</i> -butanol	46
1.5.4 Microbial <i>n</i> -butanol production	47
1.5.4.1 Native <i>n</i> -butanol production	49
1.5.4.2 Heterologous <i>n</i> -butanol production.....	49
1.6 <i>Geobacillus</i> ; a host organism.....	51
1.6.1 Properties	51
1.6.2 Taxonomy	52
1.6.3 Habitat	55

1.6.4 Biotechnological application.....	56
1.6.5 High temperature fermentation	58
1.6.6 <i>Geobacillus thermoglucosidasius</i>	59
1.6.7 Biofuel production using <i>G. thermoglucosidasius</i>	60
1.7 Metabolic engineering for fuel and chemical production	61
1.7.1 Methods for engineering <i>Geobacillus</i>	61
1.7.2 Pathway control	64
1.8 Optimizing biofuel production	65
1.8.1 Balancing co-factors.....	66
1.8.2 Consolidated bioprocessing.....	66
1.8.3 Pathway flux.....	68
1.9 Challenges of biofuel production.....	69
1.9.1 Theoretical yields	69
1.9.2 Thermophilic production	71
1.9.3 Aerobic production	71
1.10 Aims of this study.....	72
CHAPTER 2. MATERIALS AND METHODS	73
2.1 Bacterial growth conditions.....	74
2.1.1 Bacterial strains.....	74
2.1.2 Growth media	75
2.1.3 Growth conditions	78
2.1.4 Cell preparation	79
2.2 Laboratory reagents.....	80
2.3 DNA purification.....	81
2.4 Bioinformatic tools.....	81
2.5 Synthetic genes	82
2.5.1 Tags	83
2.6 Cloning strategy	84
2.6.1 Vectors	84
2.6.2 Oligonucleotide primers	85
2.6.3 Biobrick cloning.....	85
2.6.4 NEBulider HiFi DNA assembly	86
2.6.5 Cloning protocol.....	88

2.7	Detection of protein expression	91
2.7.1	Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	91
2.7.2	Western Blot of tagged proteins	92
2.8	Enzyme Assays	93
2.8.1	Cell lysis.....	93
2.8.2	Reporter gene assays.....	93
2.8.3	<i>n</i> -butanol gene assays	95
2.9	Mutant strains	97
2.9.1	Genomic deletions.....	98
2.10	Culture conditions	98
2.10.1	Starter culture preparation	98
2.10.2	Batch culture.....	99
2.11	Analytical procedures	100
2.11.1	High Performance Liquid Chromatography (HPLC)	100
2.11.2	Gas Chromatography (GC).....	101
2.11.3	Gas Chromatography Mass Spectrometry (GC-MS).....	101
2.12	Plasmid segregational stability	102
2.13	Plasmid copy number	103
2.14	<i>n</i> -butanol toxicity.....	104
2.14.1	Determining tolerance.....	104
2.14.2	Directed evolution	104
2.15	Whole genome sequencing	105
2.15.1	Sequencing data analysis.....	105
2.16	Microscopy	106
2.16.1	Light microscopy.....	106
2.16.2	Confocal microscopy.....	107
2.17	Prophage Induction	108
	CHAPTER 3. DEVELOPMENT OF MOLECULAR TOOLS	111
3.1	Introduction	112
3.2	Reporter systems.....	112
3.2.1	PheB.....	113
3.2.2	Green fluorescent proteins.....	116

3.2.3 LicB	117
3.3 Promoters	118
3.4 RBS	121
3.5 Orthogonal expression system	127
3.6 Development of the modular vector system.....	129
3.6.1 Gram-positive replicons.....	131
3.6.1.1 Plasmid segregational stability	133
3.6.1.2 Plasmid copy number	138
3.6.1.3 Host range.....	143
3.6.1.4 Compatibility.....	145
3.6.2 Marker.....	146
3.6.3 Gram-negative replicons.....	150
3.6.4 Application specific module.....	151
3.7 Co-transformation	151
3.8 Mega plasmid free strain	155
3.9 Toxin/Antitoxin module	160
3.10 Discussion.....	163
CHAPTER 4. CoA DEPENDENT <i>n</i>-BUTANOL PRODUCTION	169
4.1 Introduction	170
4.2 A synthetic <i>n</i> -butanol metabolic pathway.....	170
4.2.1 A thermophilic gene source.....	172
4.2.2 Selection of heterologous genes	173
4.2.3 Codon optimisation.....	176
4.2.4 Pathway construction and expression.....	177
4.2.5 Utilisation of a modified host strain	182
4.2.6 Culture for <i>n</i> -butanol production	183
4.2.7 Discussion.....	189
4.3 Upregulation of <i>Geobacillus</i> native genes for <i>n</i> -butanol production.....	190
4.3.1 Overexpression of native genes.....	191
4.3.2 Enzyme activity assays	196
4.3.3 A fully native system	199
4.3.4 Culture for <i>n</i> -butanol production	201
4.3.5 Discussion.....	205

4.4 Modification of the synthetic pathway	210
4.4.1 Creating a driving force	210
4.4.2 AdhE alternatives.....	214
4.4.3 The final production strain	219
4.4.5 Discussion	229
4.5 Discussion	231
4.5.1 The CoA dependent <i>n</i> -butanol pathway	231
4.5.2 Problems encountered	232
4.5.3 Further work	235
CHAPTER 5. CoA INDEPENDENT <i>n</i>-BUTANOL PRODUCTION	241
5.1 Introduction	242
5.2 Fatty acid metabolism	242
5.3 Production of Short Chain Fatty Acids.....	244
5.3.1 Thioesterase selection	245
5.3.2 Expression in <i>Geobacillus</i>	248
5.3.3 Culture for butyric acid production	253
5.3.4 Fatty acid profiling	254
5.3.5 Supplemented culture	258
5.4 Conversion of butyric acid to <i>n</i> -butanol	258
5.4.1 Expression in <i>Geobacillus</i>	263
5.4.2 Culture for <i>n</i> -butanol production	265
5.4.3 Fatty acid profiling	268
5.5 Discussion	269
5.5.1 Thioesterase	270
5.5.2 Reduction of butyric acid to butyraldehyde	273
5.5.3 Host strain metabolic engineering	274
5.5.4 Conclusions	274
CHAPTER 6. HOST STRAIN CHARACTERISATION	277
6.1 Introduction	278
6.2 Prophage.....	278
6.3 Product toxicity.....	280
6.3.1 Toxicity determination	281
6.3.2 Directed evolution for increased <i>n</i> -butanol tolerance.....	282

6.3.3 Sequencing of the evolved strain	283
6.4 Sporulation.....	290
6.4.1 Generation of a sporulation mutant.....	290
6.5 Phenotypic observations.....	294
6.5.1 Cell aggregation	294
6.5.2 Floating biofilm	296
6.6 Discussion.....	303
6.6.1 n-butanol toxicity and directed evolution	303
6.6.2 Sporulation mutant.....	305
6.6.3 Phenotypic observations of clumping and biofilm formation	306
6.6.4 <i>Geobacillus</i> as an industrial host strain	306
CHAPTER 7. DISCUSSION.....	309
7.1 Background	310
7.2 Key findings of this work.....	310
7.2.1 Development of new tools for Gram-positive thermophiles	310
7.2.2 <i>n</i> -butanol pathways investigated	312
7.2.3 Host strain characterisation.....	314
7.3 Limitations.....	314
7.4 Future perspectives	317
7.4.1 Alternative potential pathways	318
7.4.2 Use of weak promoters.....	319
7.4.3 Balancing redox, cofactor generation and ATP synthesis.....	320
7.5 Concluding remarks	321
BIBLIOGRAPHY	323
APPENDIX.....	339
A1. Primers.....	340
A2. Reporter gene sequences.....	343
A3. Promoter sequences.....	345
A4. TcdR/TcdB sequences.....	346
A5. Sequences of Gram-positive replicons	346
A6. Sequences of marker genes.....	352
A7. Sequences of Toxin and Antitoxin	353
A8. Codon optimisation table	354

A9. <i>n</i> -butanol gene sequences	354
A9.1 CoA dependent synthetic genes.....	354
A9.2 CoA dependent <i>Geobacillus</i> native genes.....	362
A9.3 CoA Independent genes	365

List of Figures

Chapter 1. Introduction

Figure 1.1. World energy consumption by source (BP Statistical Review of World Energy 2017).	33
Figure 1.2. Biosynthetic pathways of biofuels (Liao et al., 2016).	35
Figure 1.3. Schematic representation of <i>n</i> -butanol production via the norvaline biosynthetic pathway in engineered <i>E. coli</i> (Zheng et al., 2009).	37
Figure 1.4. Schematic diagram representing the ABE fermentation pathway (Zheng et al., 2009).	39
Figure 1.5. <i>n</i> -Butanol production by enabling reversal β -fatty acid oxidation.	40
Figure 1.6. Schematic diagram representing alcohol production process from LCM.	44
Figure 1.7. The chemical structure of <i>n</i> -butanol.	47
Figure 1.8. Whole genome phylogeny of the genus <i>Geobacillus</i> (Aliyu et al., 2016).	54
Figure 1.9. Geographical and environmental diversity of <i>Geobacillus</i> isolation (Zeigler, 2014).	55
Figure 1.10 Systematic approach of strain development for the production of chemicals from engineered organisms (Lan & Liao, 2013).	65
Figure 1.11 Consolidated bioprocessing (Liao et al., 2016).	67
Figure 1.12. Carbon conservation pathways (Liao et al., 2016).	70

Chapter 2. Materials and methods

Figure 2.1. The structure of synthetic genes used in this work.	83
Figure 2.2. The standardised format of a BioBrick part.	85
Figure 2.3. Diagram of the Biobrick cloning method.	86
Figure 2.4. Diagram of the NEBuilder HiFi DNA Assembly Method.	87

Chapter 3. Tool development

Figure 3.1. Qualitative visualisation of <i>pheB</i> positive cultures.	114
---	-----

Figure 3.2. Functional quantitative use of the PheB reporter system across a temperature range.	115
Figure 3.3. Excitation and emission spectra of sGFP and eCGP123.	117
Figure 3.4. Promoter strength in <i>E.coli</i> as measured by <i>pheB</i> expression.	119
Figure 3.5. Promoter strength in <i>G. thermoglucosidasius</i> as measured by <i>pheB</i> expression.	120
Figure 3.6. RBS screening using sGFP assay.	123
Figure 3.7. Comparative sGFP expression from a strong <i>C. sporogenes</i> RBS and the <i>Geobacillus</i> synthetic RBS.	126
Figure 3.8. Use of the <i>tcdR/tcdB</i> orthogonal expression system in <i>G. thermoglucosidasius</i> .	128
Figure 3.9. The pMTL60000 vector series format.	130
Figure 3.10. Plasmid retention per generation at 52°C.	133
Figure 3.11. Plasmid retention per generation at 60°C.	134
Figure 3.12. Visualisation of colony PCR confirming plasmid presence at the end of the stability experiment.	135
Figure 3.13. Plasmid retention per generation of plasmids containing UB110 replicons following modification (Data from Sheng et al., 2017).	137
Figure 3.14. Visualisation of plasmid and chromosomal target amplification following a standard PCR protocol.	140
Figure 3.15. Estimation of plasmid copy number per chromosome.	141
Figure 3.16. Colony PCR confirmation of plasmid presence in <i>B. subtilis</i> .	144
Figure 3.17. Detection of plasmid based gene expression in <i>B. subtilis</i> .	145
Figure 3.18. Colony PCR confirming <i>aad</i> derived spectinomycin resistance.	149
Figure 3.19. Maps of the plasmids co-transformed in <i>E. coli</i> and <i>G. thermoglucosidasius</i> .	152
Figure 3.20. LicB assay of co-transformed <i>E. coli</i> and <i>G. thermoglucosidasius</i> culture.	153
Figure 3.21. sGFP assay of co-transformed <i>E. coli</i> culture.	154
Figure 3.22. sGFP assay of co-transformed <i>G. thermoglucosidasius</i> culture.	155

Figure 3.23. Visualisation of PCR showing both plasmid and megaplasmid presence in <i>G. thermoglucosidasius</i> .	156
Figure 3.24. PCR confirmation showing transformation of pMTL64330 did not result in loss of pNCI001.	157
Figure 3.25. Recombination analysis.	158
Figure 3.26. Assay showing plasmid based GFP expression in megaplasmid knock out strains.	159
Figure 3.27. Depiction of Phd-Doc function resulting in postsegregational killing (Liu et al., 2008).	161
Figure 3.28. Application of TA systems for DNA cloning (Unterholzner et al., 2013).	162

Chapter 4. CoA dependent *n*-butanol production

Figure 4.1. Diagram showing the ABE fermentation pathway reactions as present in clostridia.	172
Figure 4.2. The metabolic ABE pathway for the production of <i>n</i> -butanol.	173
Figure 4.3. The desired operon construction for the <i>n</i> -butanol pathway.	178
Figure 4.4. Western blot of <i>Geobacillus</i> expressing synthetic <i>n</i> -butanol pathway genes.	179
Figure 4.5. Diagram of a four gene operon, a partial <i>n</i> -butanol pathway.	180
Figure 4.6. Western blot of <i>Geobacillus</i> expressing an operon of four synthetic <i>n</i> -butanol pathway genes.	181
Figure 4.7. Acetate and ethanol production of strains cultured in aerobic and micro aerobic conditions.	185
Figure 4.8. Visualisation of PCR amplification to confirm <i>ter</i> genomic integration.	186
Figure 4.9. Visualisation of PCR amplification to confirm <i>ter</i> genomic integration stability at 60°C.	187
Figure 4.10. Butanol and butyrate evaporation.	188
Figure 4.11. Pathway diagram showing butanoate metabolism in <i>G. thermoglucosidasius</i> .	190

Figure 4.12. Visualisation of native gene amplification from genomic DNA.	194
Figure 4.13. Plasmid maps of the pMTL61222 vector containing native genes.	195
Figure 4.14. Enzyme activities of synthetic and native <i>n</i> -butanol pathway genes.	197
Figure 4.15. Pathway diagram showing fatty acid biosynthesis in <i>G. thermoglucosidasius</i> .	200
Figure 4.16. Maps of the native gene plasmids co-transformed in <i>G. thermoglucosidasius</i> .	202
Figure 4.17. The final native gene pathway.	203
Figure 4.18. Glucose consumption, metabolite production, growth and pH of three <i>G. thermoglucosidasius</i> strains expressing the native gene <i>n</i> -butanol pathway.	204
Figure 4.19. Diagram depicting the reaction catalysed by both the thiolase and acetoacetyl-CoA synthase genes (Okamura et al., 2010).	211
Figure 4.20. Variations in the CoA-dependent <i>n</i> -butanol pathway (Lan & Liao, 2012).	212
Figure 4.21. The reaction mechanism of acetyl-CoA carboxylase (Davis et al., 2000).	213
Figure 4.22. Enzyme activities of potential AdhE replacements.	217
Figure 4.23. A modified CoA dependent <i>n</i> -butanol pathway.	219
Figure 4.24. Maps of the two plasmids containing the modified CoA dependent <i>n</i> -butanol pathway.	220
Figure 4.25. Co-transformation colony PCR.	221
Figure 4.26. <i>n</i> -butanol production results under aerobic growth conditions.	222
Figure 4.27. <i>n</i> -butanol production results under micro aerobic growth conditions.	223
Figure 4.28. Growth and relative yield of the modified <i>G. thermoglucosidasius</i> <i>n</i> -butanol production strain.	224
Figure 4.29. Metabolite production of the modified <i>G. thermoglucosidasius</i> <i>n</i> -butanol production strain.	225

Figure 4.30. Scaled up culture of the <i>G. thermoglucosidasius</i> <i>n</i> -butanol production strain.	228
Figure 4.31. Sequencing of recovered plasmid containing the four sub-unit <i>acc</i> gene.	233
Figure 4.32. Sequencing of recovered plasmid containing the synthetic <i>adhE</i> gene.	234
Figure 4.33. The enoyl-CoA hydratase and (S)-3-hydroxyacyl-CoA dehydrogenase reactions of the fatty acid degradation pathway, catalyzed by FadB (Volodina & Steinbüchel, 2014).	236
Figure 4.34. Future engineering the CoA-dependent <i>n</i> -butanol production pathway for possible higher efficiency (Chen & Liao, 2016).	238

Chapter 5. CoA independent *n*-butanol production

Figure 5.1. Overview of metabolic pathways that lead to the production of fatty acids and fatty acid-derived chemicals (Yu et al., 2014).	244
Figure 5.2. Cloning for construction of the pMTL61222+ <i>tesBT</i> vector.	249
Figure 5.3. Western blot of <i>Geobacillus</i> expressing the TesBT protein.	251
Figure 5.4. Thioesterase enzyme activity assay.	252
Figure 5.5. Free fatty acid profiles of <i>G. thermoglucosidasius</i> wild type and <i>G. thermoglucosidasius</i> pMTL61222+ <i>tesBT</i> cells.	255
Figure 5.6. Oxygen tolerant <i>n</i> -butanol synthesis (Pasztor et al., 2014).	259
Figure 5.7. The catalytic cycle of Car (Akhtar et al., 2013).	260
Figure 5.8. A synthetic TSCA operon in the pMTL <i>Geobacillus</i> shuttle vector.	263
Figure 5.9. Western blot of <i>Geobacillus</i> expressing the TSCA operon.	264
Figure 5.10. Growth, pH and product formation of <i>G. thermoglucosidasius</i> expressing the TSCA operon.	266
Figure 5.11. Free fatty acid profiles of <i>G. thermoglucosidasius</i> wild type and <i>G. thermoglucosidasius</i> pMTL61222+TSCA cells.	269
Figure 5.12. The kinetic characterization of Car (Akhtar et al., 2013).	273

Chapter 6. Host strain characterisation

Figure 6.1. Locations of six incomplete phage regions in the <i>G. thermoglucosidasius</i> genome.	279
Figure 6.2. Growth of <i>G. thermoglucosidasius</i> when exposed to a range of <i>n</i> -butanol concentrations.	281
Figure 6.3. Variants identified in gene BCV53_13285.	287
Figure 6.4. Variants identified in gene BCV53_14485.	288
Figure 6.5. Screening of $\Delta sigF$ <i>G. thermoglucosidasius</i> mutant strains.	293
Figure 6.6. Light microscopy image of <i>G. thermoglucosidasius</i> $\Delta sigF$ cells.	294
Figure 6.7. <i>G. thermoglucosidasius</i> cell aggregation.	295
Figure 6.8. Floating biofilm formation.	296
Figure 6.9. Cell morphology during biofilm formation; cells form long chains.	298
Figure 6.10. Cell morphology during biofilm formation; chains of cells aggregate.	299
Figure 6.11. Cell morphology during biofilm formation; the clusters of cell chains form a woven string-like structure.	299
Figure 6.12. Live/dead cell staining of <i>G. thermoglucosidasius</i> biofilm with visualisation using laser confocal microscopy.	301
Figure 6.13. Observation of <i>G. thermoglucosidasius</i> biofilm's 3D structure visualised by Z-stacked laser confocal microscopy images.	302
Figure 6.14. Observation of <i>G. thermoglucosidasius</i> biofilm, visualised by Z-stacked laser confocal microscopy images.	303

Chapter 7. Discussion

Figure 7.1. Schematic representation of the heterologous pathways for <i>n</i> -butanol biosynthesis investigated.	312
Figure 7.2. Overview of potential applications of synthetic biology tools for construction and optimization of metabolic pathways (Yu et al., 2014).	317

Appendix

Figure A8.1. Codon usage table for <i>G. thermoglucosidasius</i> .	354
--	-----

List of Tables

Chapter 1. Introduction

Table 1.1. Summary of microbial <i>n</i> -butanol production (Chen and Liao, 2016)	48
Table 1.2. Description of the <i>Geobacillus</i> genus.	52
Table 1.3. A selection of U.S patents for <i>Geobacillus</i> products or processes.	57
Table 1.4. Essential and desirable traits for efficient fermentation of LCM.	57

Chapter 2. Materials and methods

Table 2.1. <i>E. coli</i> bacterial strains used in this study.	74
Table 2.2. Plasmids used in this work.	84
Table 2.3. HiFi DNA assembly protocol (NEB).	88

Chapter 3. Tool development

Table 3.1. Sequences of the screened RBS.	125
Table 3.2. Component modules of the pMTL60000 vector series.	131
Table 3.3. Gram positive replicon source plasmids.	132
Table 3.4. Plasmid presence in culture after 72 hours at 52°C.	134
Table 3.5. Plasmid presence in culture after 72 hours at 60°C.	135
Table 3.6. qPCR primers.	139
Table 3.7. Size of qPCR amplification products.	140
Table 3.8. Compatibility of Gram positive replicons.	146
Table 3.9. Minimum inhibitory concentrations (MIC) of spectinomycin.	148
Table 3.10. Minimum inhibitory concentrations (MIC) of gentamicin.	150

Chapter 4. CoA dependent *n*-butanol production

Table 4.1. The five enzymes selected for conversion of acetyl-CoA to <i>n</i> -butanol.	174
Table 4.2 Strains screened for <i>n</i> -butanol production.	184

Table 4.3. Blastp (protein-protein BLAST) comparison of protein to protein sequences of the native genes compared to the genome of two <i>n</i> -butanol producing organisms.	192
Table 4.4. Aldehyde and alcohol dehydrogenase genes screened.	215
Table 4.5. Concentration of product formation from supplemented culture.	218

Chapter 5. CoA independent *n*-butanol production

Table 5.1. Molar percentages and total concentrations of fatty acids produced by different thioesterases in <i>E. coli</i> (Jing et al., 2011).	247
Table 5.2 Chain length specificity of thioesterases (Jawed et al., 2016).	272

Chapter 6. Host strain characterisation

Table 6.1. Single nucleotide variations in the <i>n</i> -butanol resistant strain in comparison to the <i>G. thermoglucosidasius</i> NCIMB 11955 reference strain.	285
Table 6.2. Multi nucleotide variants, insertions and deletions in the <i>n</i> -butanol resistant strain in comparison to the <i>G. thermoglucosidasius</i> NCIMB 11955 reference strain.	286

Abbreviations

%, percent

°C, degrees Celsius

μM, micro molar

5-FOA, 5-fluoroorotic acid

aac, gentamicin resistance gene

aad, spectinomycin resistance gene

ABE, acetone, butanol and ethanol

Acc, acetyl-CoA carboxylase

ACE, allelic coupled exchange

ACP, acyl carrier protein

adhE/adhE2, aldehyde/alcohol dehydrogenase

ADP, adenosine diphosphate

Ahr, aldehyde reductase

ANOVA, analysis of variance

ATP, adenosine triphosphate

Bcd/Etf, butyryl-CoA dehydrogenase electron transferring flavoprotein complex

BLAST, basic local alignment tool

Bp, base pairs

C23O, catechol 2,3-dioxygenase

C4, 4 carbon chain length

Car, carboxylic acid reductase

Cas9, CRISPR-associated protein-9 nuclease

CBM, Clostridia base medium

CBM1X, Clostridia base medium 1% xylose

CFU, colony forming units

CH₄, methane

CO, carbon monoxide

CO₂, carbon dioxide

CoA, Coenzyme A

CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats

Crt/*crt*, crotonase

Ct, Cycle threshold

DNA, deoxyribonucleic acid

eCGP123, enhanced consensus green protein variant 123

EPB, electroporation buffer

EPS, extracellular polymeric substance

EU, European Union

FAEE, fatty acid ethyl esters

FAME, fatty acid methyl esters

FFA, free fatty acids

FFA, free fatty acids

g/l, gram per litre

g/mol, gram per mole

G3P, glyceraldehyde 3 phosphate

GC, Gas Chromatography

GC, guanine-cytosine

GC-MS, Gas Chromatography Mass Spectrometry

GDP, Gross domestic product

GFP, green fluorescent protein

GHG, Greenhouse gas

GRAS, generally regarded as safe

Gt, gigatonne

h, hours

H₂, hydrogen

Hbd/*hbd*, 3-hydroxybutyryl-CoA dehydrogenase

HMSA, 2-hydroxymuconic semi aldehyde

HPLC, High Performance Liquid Chromatography

ICGEB, International Centre for Genetic Engineering and Biotechnology

IU, international unit

kan, kanamycin adenyltransferase gene

Kbp, kilobase pair/ 1000 base pairs

kcat/K_m, catalytic activity

KEGG, Kyoto Encyclopedia of Genes and Genomes

kg/m³, kilogram per cubic metre

LB, Luria-Bertani/Lysogeny broth

LCM, lignocellulosic material

LHA, left homology arm

MCS, multiple cloning site

mg, milligram

MIC, minimum inhibitory concentration

MJ/kg, megajoules per kilogram

ml, millilitre

mM, milli molar

mm, millimetre

MNV, multi-nucleotide variants

Mt, megaton

N₂O, nitrous oxide

NAD⁺/NADH, nicotinamide adenine dinucleotide oxidised/reduced

NADP⁺/NADPH, nicotinamide adenine dinucleotide phosphate oxidised/reduced

NEB, New England Biolabs

nm, nanometre

NOG, non-oxidative glycolysis pathway

O₂, oxygen

OD, optical density
ORI, origin of replication
OriT, origin of transfer
P value, probability value
p<, p value less than
PCR, polymerase chain reaction
pH, potential of hydrogen
Phage, bacteriophage
PI, propidium iodide
pKa, acid dissociation constant at logarithmic scale
PLP, pyridoxal phosphate dependent aminotransferase
pyrE, orotate phosphoribosyltransferase
qPCR, real-time quantitative PCR
R/P, reserves-to-production ratio
RBS, ribosome binding site
RC, rolling circle
Rep, replication-initiation protein
RFU, relative fluorescent units
rGS, synthetic reverse glyoxylate shunt
RHA, right homology arm
RNA, ribonucleic acid
RPM, revolutions per minute
rRNA, ribosomal RNA
SCFA, short-chain fatty acid
SD, Shine-Dalgarno
SNV, single nucleotide variants
sp./spp., species singular/multiple
Syngas, synthetic gas
TA, toxin/antitoxin

Ter, *ter*, trans-2-enoyl-reductase
Thl/*thl*, thiolase
T_M, melting temperature
tRNA^{fMet}, formyl-methionyl-transfer RNA
TSA, tryptone soya agar
UK, United Kingdom
UN, United Nations
US, United States
UV, ultraviolet
v/v, volume/volume concentration
w/v, weight/volume concentration
WHO, World Health Organisation
WT, wild type
ΔG (kcal/mole), free energy kilocalories per mole

CHAPTER 1

INTRODUCTION

1.1 Prologue

This work forms part of the Ricefuel project, an international collaboration, funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Indian Department of Biotechnology (DBT), with the overall aim to develop a microbial based process able to convert waste rice straw into advanced butanol and alkane biofuels.

Streams of this research programme aim to; 1. Develop a novel enzyme cocktail optimised for rice straw deconstruction, 2. Develop engineered strains of *Geobacillus* to produce advanced biofuels, 3. Incorporate the two into a single process capable of producing advanced biofuels from rice straw.

The focus of this work sits within the second stream, aiming to develop a *Geobacillus thermoglucosidasius* strain capable of efficiently generating *n*-butanol.

1.2 The need for fossil fuel alternatives

Limited fossil fuel resources and the environmental impacts of climate change are motivating the development of sustainable processes for the production of fuels and chemicals from renewable resources. The development of alternative energy sources, such as biofuels, could strengthen energy security and reduce dependence on fossil fuels.

1.2.1 Climate change

Current global energy demands are met by the burning of fossil fuels; coal, oil and gas. Burning fossil fuel releases greenhouse gasses, carbon dioxide (CO₂), methane (CH₄) and nitrous oxide (N₂O), into the atmosphere. Approximately 7 Gt of carbon is released into the atmosphere each year from fossil fuel usage. If current trends continue, the CO₂ release rate is predicted to double by 2050 (Liao et al., 2016). As the amount of CO₂ in the atmosphere increases, so does the Earth's average temperature (Cocks, 2009). An average world temperature rise of more than 2°C above the preindustrial level, equivalent to around 1.2°C

above current temperature, will greatly increase the risk of large scale, irreversible changes in the global environment (European Commission, 2012).

The potential impacts of climate change include; melting of polar ice caps causing sea levels to rise, extreme weather events such as floods, droughts and heatwaves, the spread of tropical diseases, plant and animal species at increased risk of extinction. Through its impact on water resources and food production, climate change could threaten regional and international security by triggering or exacerbating conflicts, famines and refugee movements (European Commission, 2012). The UN and the WHO recognise climate change as a priority, due to the significant threat posed to public health. The UN estimates climate change could cause approximately 250,000 additional deaths per year, from heat stress, malnutrition, diarrhoea and malaria, between 2030 and 2050 (UNFCCC, 2017).

The need to keep global warming below 2°C is now recognised by the international community. Preventing global warming from exceeding this threshold is both technologically feasible and economically affordable; estimated to cost around 1% of global GDP. With the cost of taking no action expected to be 5-20% of global GDP (European Commission, 2012). To avoid reaching the 2°C threshold, a substantial decrease in carbon emissions from fossil fuel usage, the scale-up of renewable energy supplies and a decrease in energy demand, are required. This is currently a global priority. The Paris agreement sits within the United Nations Framework Convention on Climate Change (UNFCCC). Starting in the year 2020, the agreement deals with greenhouse gas emissions mitigation, adaptation and finance. As of November 2017, 195 parties have signed the agreement. Committing governments to move their economies away from fossil fuels.

1.2.2 Finite resources

By their nature, fossil fuel resources are finite. The reserves-to-production ratio (R/P) calculates the remaining reserves divided by the annual production, resulting in the length of time the remaining reserves will last, if production were to continue at the current rate. At the end of 2016 the R/P for coal was 153 years, oil was 50.6 years and natural gas was 52.5 years (BP Statistical Review of World Energy, 2017). Therefore, if usage was to remain constant, current fossil fuel reserves will be depleted by 2169. To maintain energy supply in the future alternatives are required.

1.2.3 Energy security

Many nations, including the UK, are dependent on importing energy. Many countries are also heavily reliant on a single supplier. For example relying entirely on Russia for natural gas and the Middle East for oil. This dependence leaves them vulnerable to supply disruptions, caused by political or commercial disputes, or infrastructure failure. An unstable geopolitical situation means energy prices and supply are insecure. One measure to reduce this reliance and mitigate the impacts of supply disruptions, is to increase internal energy production, such as biofuels.

1.3 Energy demand

There is a desire for cheaper, environmentally friendly and secure energy. Consumers want energy to be affordable, societies want supplies to be secure and reliable and responsible global citizens want power to be clean, green and low carbon. This presents a three way battle between competitive pricing, security of supply and carbon saving commitments. One solution would be to consume less energy. This would result less pronounced peaks and troughs in demand, cheaper energy and a reduction in carbon emissions. However as world consumption is ever increasing, as depicted in Figure 1.1, alternative solutions must be developed.

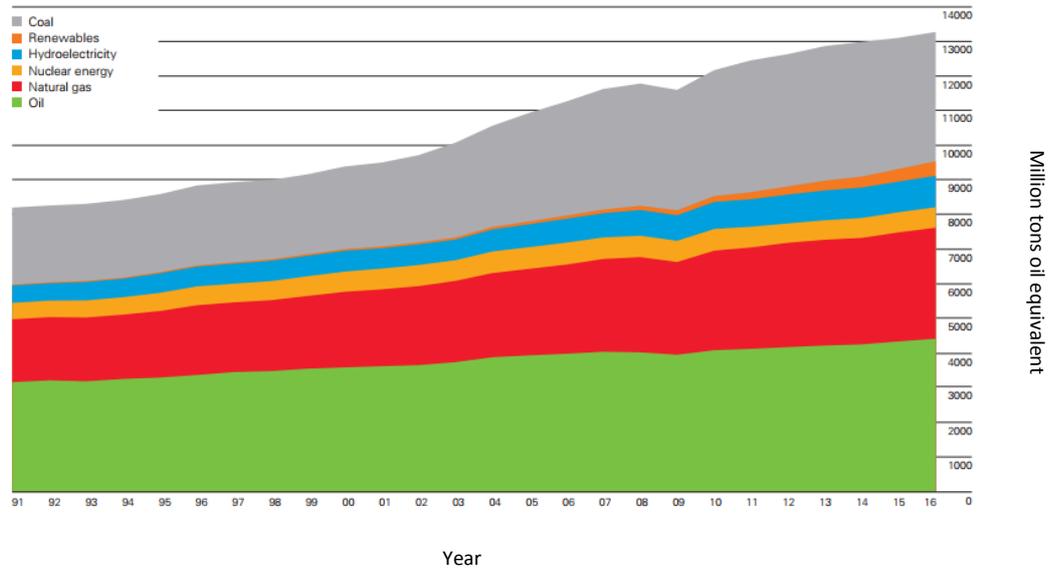


Figure 1.1. World energy consumption by source (BP Statistical Review of World Energy 2017).

World energy consumption is ever increasing. In 2016 consumption grew by 1%. Oil is the world's dominant fuel, accounting for a third of all energy consumed. In 2016 oil gained global market share for the second consecutive year, following 15 years of decline from 1999 to 2014. Coal's market share fell to 28.1% in 2016, the lowest level since 2004. Renewable power accounted for a record 3.2% of global energy consumption. Although at a low level, renewable energy usage is increasing. Renewable energy led by wind and solar power (excluding hydro-power) grew by 14.1% in 2016. This strong growth is a result of continuing technological advances (BP Statistical Review of World Energy 2017).

Rapid growth and improving prosperity mean growth in energy demand is increasingly coming from developing economies, particularly within Asia. However, improving energy efficiency is causing global energy consumption overall to decelerate. The energy mix is shifting towards cleaner, lower carbon fuels, driven by the environmental need and technological advances. Global biofuels production rose by 2.6% in 2016, with the largest increase from US

production. Global ethanol production increased by 0.7% and biodiesel production rose by 6.5%. An increase in biofuel use will help the transport sector in decreasing its dependency on oil.

1.4 Energy policies

Government policies play an important role in shaping national energy use and production. Governments across the world have targets to increase market share of renewable energy. In order to meet these targets, policy instruments creating favourable economic and legal conditions for biofuels are needed (Wiesenthal et al., 2009). Policy tools such as; tax exemptions, obligations, subsidies and price support can be applied in a number of ways in order to create a market niche for biofuels. Policy support is required to overcome the initial costs of technological innovation and market development required to make the sector competitive (Darmani et al., 2014). Various policies are in place across the world mandating or targeting minimum volumes of biofuel and minimum percentage of renewable energy use. For example, in the EU 20% of the overall energy mix must be renewable by 2020 with a target of 5% advanced (non-food) biofuels (Araújo et al., 2017).

1.5 Biofuels

Biofuels are defined as fuels produced from biomass by either biological or non-biological processes and fuels produced from other renewable resources, such as landfill gas and CO₂, by biological processes (Liao et al., 2016). Microbial production of fuel and utilisation of biomass, offers an approach to providing sustainable energy with reduced environmental impact. Advanced biofuels have high energy content, are compatible with existing infrastructure and can be produced on an industrial scale. In 2015 the global biofuel supply was approximately 35 billion gallons (Araújo et al., 2017). This constituted roughly 3:1 bioethanol to biodiesel.

1.5.1 Biofuel production pathways

Following the saccharification of biomass or the assimilation of carbon from CO₂ or methane, discussed in section 1.5.2, raw materials are metabolized to either pyruvate or acetyl-CoA. These are common starting metabolites from which biofuels can be produced. Although these metabolic pathways exist in some microorganisms, few have a complete pathway or can synthesize the desired molecule efficiently. The challenge is to increase the titre, yield and productivity associated with biofuel production. Approaches to achieve this include; engineering modifications to the pathway in its native host or introducing the pathway into an alternative organism.

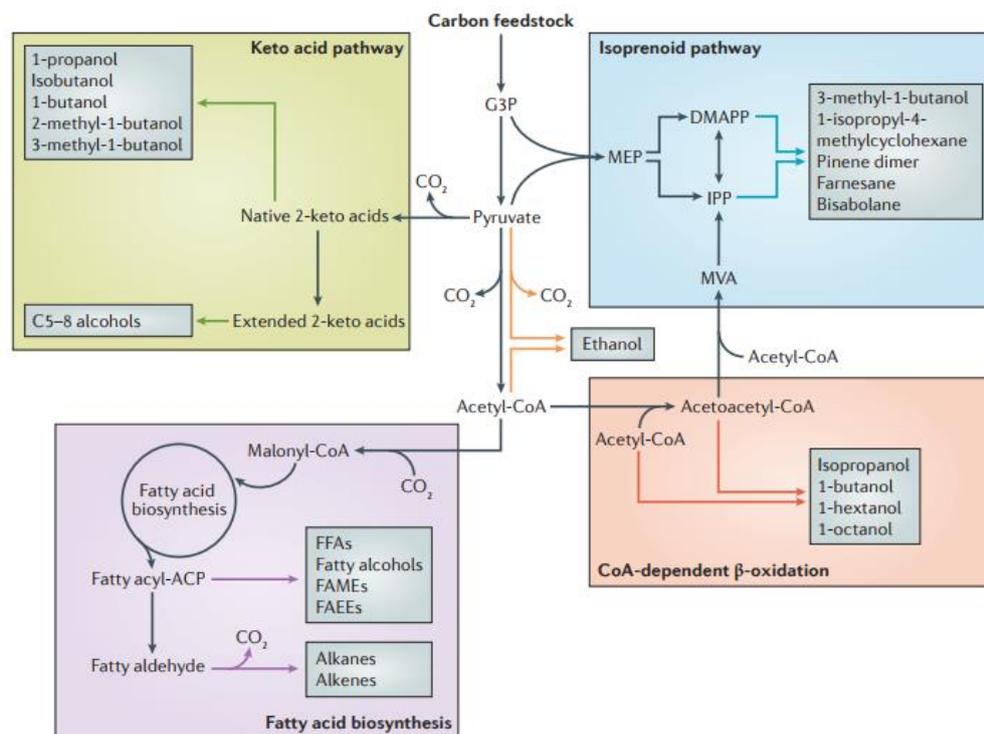


Figure 1.2. Biosynthetic pathways of biofuels (Liao et al., 2016). Ethanol is produced from either pyruvate or acetyl-CoA (orange arrows), with acetaldehyde as a common intermediate. The keto acid pathway (green arrows) can be used to produce both branched and straight-chain alcohols. It uses parts of amino acid biosynthesis pathways for keto acid chain elongation. This is followed by decarboxylation and reduction of the keto acid, analogous to the conversion of pyruvate to ethanol. Fatty acid synthesis (purple arrows)

extends acyl-acyl carrier proteins (ACPs) in a cyclical manner, using malonyl-CoA as a precursor. Fatty acyl-ACPs may be converted into free fatty acids (FFAs) with acyl-ACP thioesterase. FFAs can be esterified to esters, such as fatty acid methyl esters (FAMEs) or fatty acid ethyl esters (FAEEs), reduced to fatty alcohols, or reduced to fatty aldehydes followed by decarbonylation to alkanes and alkenes. The CoA-dependent pathway (red arrows) uses reverse β oxidation chemistry for the production of higher alcohols or decarboxylation of the precursor acetoacetyl-CoA for the production of isopropanol. Isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), the universal precursors of isoprenoid biofuel biosynthesis (blue arrows), may be produced either through the mevalonate (MVA) or methylerythritol 4 phosphate (MEP) pathway. G3P, glyceraldehyde 3 phosphate.

1.5.1.1 Keto acid pathway

A route for the production of alcohols including *n*-butanol is the Ehrlich, or 2-keto-acid, amino acid synthesis pathway. This pathway decarboxylates amino acid precursors, keto acids, into aldehydes and reduces them to alcohols. Amino acid based alcohols, such as *n*-propanol from isoleucine, isobutanol from valine and *n*-butanol from norvaline, can be produced using this pathway. Atsumi et al. (2008) engineered an *Escherichia coli* strain to produce isobutanol at a high titre (22 g/l which represents 86% of the theoretical yield) from glucose, using the keto acid pathway. Isobutanol is currently being commercialized as a biofuel using this pathway.

The pathway for *n*-butanol production relies on 2-ketovalerate. 2-ketovalerate is a rare metabolite and the precursor of the amino acid norvaline. 2-ketovalerate can be synthesized from 2-ketobutyrate by keto acid chain elongation catalyzed by enzymes LeuABCD (Shen and Liao, 2008), as shown in Figure 1.3. 2-ketobutyrate has been synthesized from two routes: threonine deamination and the citramalate pathway from *Methanococcus jannaschii*. In both cases, 0.5 to 1 g/l of *n*-butanol were produced from glucose in *E. coli*

(Atsumi and Liao, 2008; Shen and Liao, 2008). One difficulty in producing *n*-butanol using this pathway is the non-specificity of KivD, which also decarboxylates 2-ketobutyrate to form 1-propanol. Therefore, efficient production of *n*-butanol using this pathway requires significant protein engineering and metabolic engineering efforts.

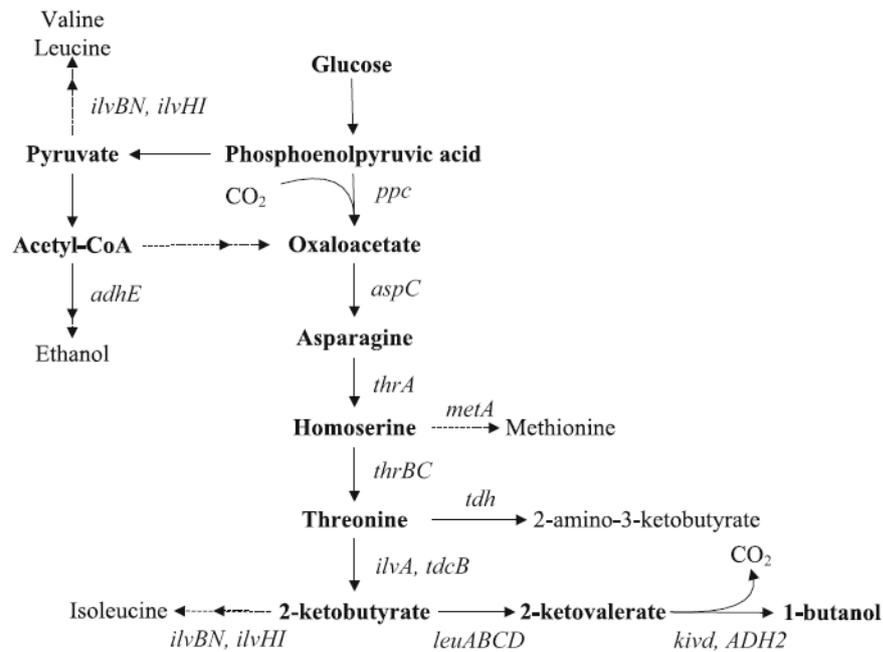


Figure 1.3. Schematic representation of *n*-butanol production via the norvaline biosynthetic pathway in engineered *E. coli* (Zheng et al., 2009).

The advantage of biofuel production using keto acid pathways is their compatibility with many organisms, as they build on the amino acid biosynthesis pathways already present. A large number of pathway enzymes are available from a variety of sources. Drawbacks of using amino acid pathways include; lack of enzyme specificity resulting in alcohol mixtures rather than sole production of a desired alcohol (Peralta-Yahya and Keasling, 2010). Amino acid biosynthesis is tightly regulated by feedback inhibition of intermediates. To increase alcohol production novel feedback resistant

enzymes are required to deregulate feedback (Peralta-Yahya and Keasling, 2010).

1.5.1.2 Isoprenoid pathway

A class of compounds widely used as flavours and pharmaceuticals, isoprenoids have potential as advanced biofuels because of the branches and rings found in their hydrocarbon chain. The isoprenoid pathway can generate a range of terpene structures. To date most work developing the isoprenoids has been done for therapeutics and nutraceutical production. However terpenes are also potential biosynthetic alternatives to petrol, diesel, and jet fuel. Branched chain, short terpenes, such as isopentanol, may be suitable petrol substitutes and longer cyclic or branched chain terpenes may be suitable diesel and jet fuel substitutes. Terpene compounds currently being explored as fuels include pinene, sabinene and terpinene (Peralta-Yahya and Keasling, 2010).

Plants are natural sources of isoprenoids, but they do not produce sufficient quantities for industrial production. Algae, such as *Botryococcus braunii*, produce large quantities, but they grow slowly and produce large amounts of fatty acids. *E. coli* and *Saccharomyces cerevisiae* have been used as platforms for isoprenoid production. Of the isoprenoid based biofuels, farnesane is the closest to commercialisation (Peralta-Yahya, 2012).

1.5.1.3 Fatty acid biosynthesis pathway

Fatty acids are an integral part of all living organisms. Fatty acids take intracellular forms of fatty acyl–acyl carrier protein, fatty acyl-coenzyme A ester, storage lipids, eicosanoids and unesterified free fatty acids (Yu et al., 2014). Fatty acid derivatives have wider applications such as biofuels, biomaterials and other chemicals. Alcohols can be produced by sequential reduction of fatty acids to fatty alcohols. Alkanes can be produced by reduction of fatty acid to aldehyde followed by decarbonylation. Biodiesel can be

produced by conversion of fatty acids to esters via esterification with small alcohols (Peralta-Yahya and Keasling, 2010).

1.5.1.4 CoA dependent pathway

Butanol producing clostridia such as *Clostridium acetobutylicum*, *Clostridium beijerinckii* and *Clostridium pasteurianum* use ABE fermentation to produce; solvents (acetone, butanol and ethanol), organic acids (acetic acid, lactic acid and butyric acid) and gases (carbon dioxide and hydrogen). The biosynthesis of acetone, butanol and ethanol share the same metabolic pathway from glucose to acetyl-CoA, branching into different pathways thereafter, as shown in Figure 1.4. A typical ABE fermentation using *C. acetobutylicum* yields acetone, butanol and ethanol in the ratio of 3:6:1.

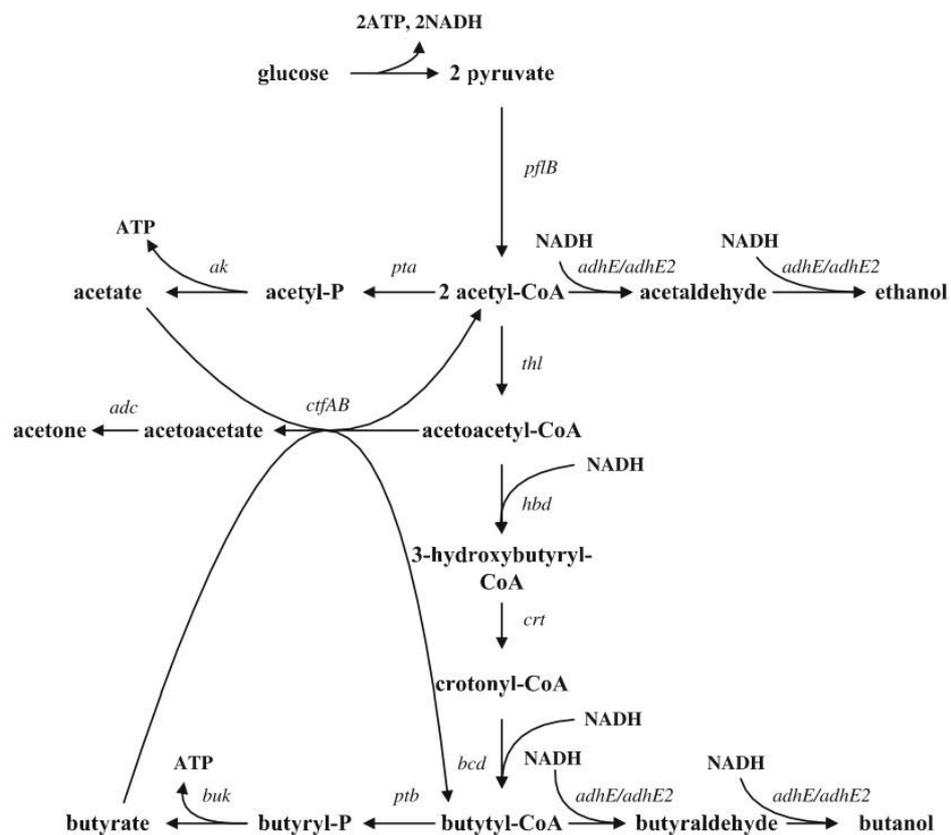


Figure 1.4. Schematic diagram representing the ABE fermentation pathway (Zheng et al., 2009). The genes are shown in italics, their corresponding enzymes are as follows: *pflB*, pyruvate ferredoxin oxidoreductase; *thl*, thiolase;

hbd, 3-hydroxybutyryl-CoA dehydrogenase; *crt*, crotonase; *bcd*, butyryl-CoA dehydrogenase; *pta*, phosphotransacetylase; *ak*, acetate kinase; *ptb*, phosphotransbutyrylase; *buk*, butyrate kinase; *ctfAB*, acetoacetyl-CoA:acetate/butyrate:CoA transferase; *adc*, acetoacetate decarboxylase; *adhE/adhE2*, aldehyde/alcohol dehydrogenase.

In a similar process the β -oxidation cycle, run in reverse, can be used to produce primary alcohols, such as *n*-butanol in a CoA dependent manner. The reverse β -oxidation pathway uses the same chemistry as β -oxidation for fatty acid degradation, but in the reverse direction. In contrast to fatty acid biosynthesis, this pathway uses CoA as opposed to acyl carrier protein (ACP) to activate the acyl group. In this pathway, two acetyl-CoA molecules are condensed to produce acetoacetyl-CoA, which then reverse β -oxidizes to produce 1-butyryl-CoA. This acyl-CoA can then be reduced to produce *n*-butanol, or go through the cycle again to increase the chain length by two carbon atoms in each cycle before reduction to longer chain alcohols.

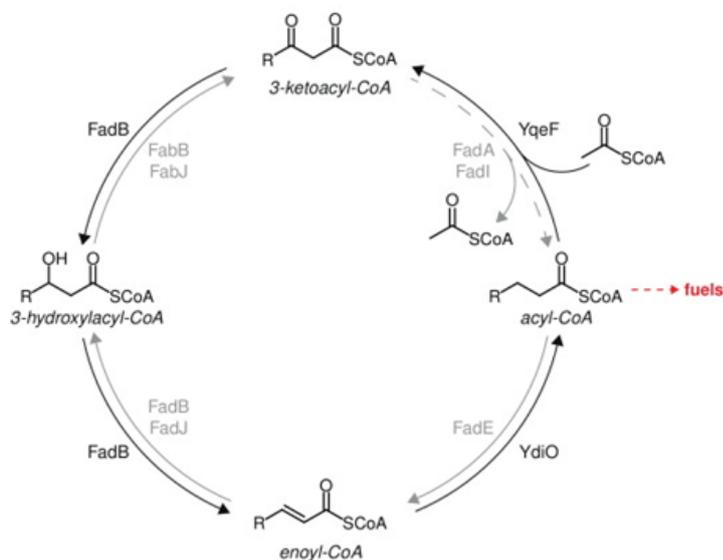


Figure 1.5. *n*-Butanol production by enabling reversal β -fatty acid oxidation.

The outside cycle (black) depicts enzymes possibly involved in *n*-butanol synthesis. The inner cycle (grey) depicts enzymes in the β -fatty acid oxidation pathway.

CoA dependent reverse β -oxidation has been engineered into several organisms that do not possess this pathway, such as *E. coli* (yielding 15g/l butanol), *S. cerevisiae* (yielding 0.0025g/l butanol), *Bacillus subtilis* (yielding 0.024g/l butanol), *Pseudomonas putida* (yielding 0.05g/l butanol) and *Synechococcus elongatus* (yielding 0.3g/l butanol).

Reversal of the β -oxidation cycle has the potential for high efficiency as it directly uses acetyl-CoA as the donor of two-carbon units during chain elongation and it functions with acyl-CoA intermediates, which are the precursors of *n*-alcohols and other important products (Dellomonaco et al., 2011).

1.5.2 Biofuel substrates

The feedstock used for *n*-butanol production is a critical factor in determining economic feasibility and the environmental and social benefits. An advantage of using microorganisms for the production of next-generation biofuels is the metabolic diversity of bacteria, fungi and algae, enabling the use of different substrates. Inexpensive and abundant raw materials are desirable substrates. For example agricultural wastes such as; maize stover, rice straw, barley straw and switchgrass.

First generation biofuels

First generation biofuels are produced from food crops grown explicitly for fuel production. Conventional production of first generation biofuels is based on fermentation of sugars and starch from sugarcane, corn and cassava. Replacing fossil fuels with first generation biofuels diverts farmland and crops to biofuel production, competing with food supply. This reduces food security and increases food prices. Cultivating food crops for biofuel production consumes large amounts of water, fertilizers and pesticides (Zhang et al., 2011). Raising concerns regarding their environmental, social and economic sustainability

(Daniell et al., 2012). Although first generation biofuels offer CO₂ benefits and can help to improve domestic energy security, they are unsustainable because of the stress their production places on food commodities (Naik et al., 2010). Technologies that do not use extensive monocultures pose less threat to food security (Daniell et al., 2012).

Second generation biofuels

Second generation biofuels have been developed to overcome the limitations of the first generation, using inedible waste product as a substrate. The use of waste feedstocks does not impact the food chain, allowing the production of biofuel without using arable lands (Sanchez et al., 2008). With an ever increasing population the need for second generation biofuels, which do not compete with food supply is vital. The production of biofuels should not threaten the availability or affordability of critical resources such as food (Daniell et al., 2012). In 2015 the global production capacity for second generation biofuels was estimated to be 225 million gallons per year (Araújo et al., 2017). The vast majority of this capacity is in bioethanol production.

Third generation biofuels

Third generation biofuels are derived from algal biomass. The algae are cultured to act as a low cost, high energy renewable feedstock. It is predicted that algae will have the potential to produce more energy per acre than conventional crops. Algae can also be grown using land and water unsuitable for food production.

Fourth generation biofuels

The fourth generation of biofuels includes carbon capture and storage. Not only producing sustainable energy but also capturing and storing CO₂ in the process. During production carbon dioxide is captured using processes such as oxy-fuel combustion. The carbon dioxide can then be geosequestered by storing it in used oil and gas fields or saline aquifers. This carbon capture makes

fourth generation biofuel production carbon negative, as it is 'locks' away more carbon than it produces. This system captures and stores carbon dioxide from the atmosphere and reduces CO₂ emissions by replacing fossil fuels.

1.5.2.1 Plant biomass

Plant biomass is the most abundant renewable resource on Earth. Accounting for approximately 50% of the biomass in the world (10–50 billion tons) (Claasen et al., 1999). Plant biomass forms the majority of the agricultural waste which can be utilised as a feedstock for second and fourth generation biofuels. Plant biomass or lignocellulosic material (LCM) is a promising substrate for biofuel production as it is a highly available, low cost feedstock. LCM is not a food source, is more abundant than starch and its use is carbon dioxide neutral. Advances in the development of fast and efficient alcohol production from LCM, could make second generation biofuels an economically viable option for industrial application.

LCM is made up of a mixture of carbohydrate polymers from the plant cell walls. Composed of; cellulose (40–50%), hemicelluloses (25–35%) and lignin (15–20%) (Gray et al., 2006). LCM has to be deconstructed to simpler chemical forms before further processing. These intermediate forms include sugars, synthetic gas (syngas), organic acids and methane.

Lignocellulose deconstruction to sugars is a multi-step process where biomass is pre-treated followed by enzymatic hydrolysis. The pre-treatment process can be physical, chemical, biological or a combination. The pre-treated biomass is then hydrolysed by either cellulase enzyme cocktails or by cellulolytic microorganisms. Cellulose degrades to monomeric hexose sugars, such as; glucose, mannose and rhamnose, which can be fermented to alcohols (Gray et al., 2006). Hemicelluloses degrade to monomeric pentose sugars, such as; xylose and arabinose. The microbial conversion of pentoses has been identified as a major research challenge to making lignocellulosic technology mature for

industry (Fromanger et al., 2010). It is crucial to use all sugar polymers to improve the cost competitiveness of the production process.

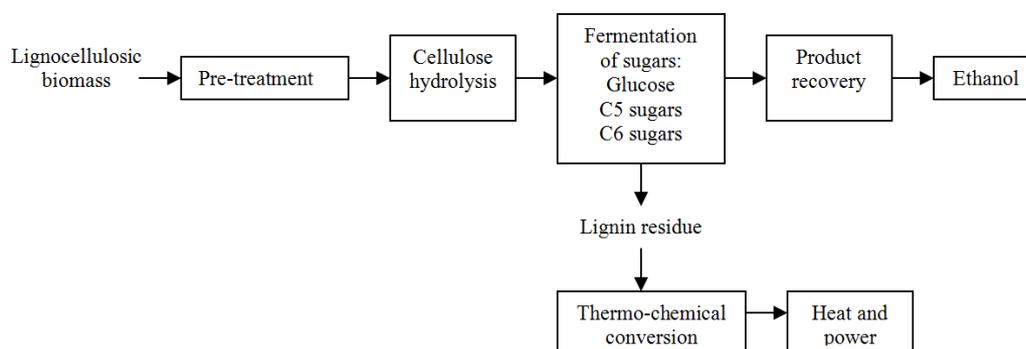


Figure 1.6. Schematic diagram representing the basic steps in an alcohol production process from LCM.

Currently the major economic limitation to the use of lignocellulosic biomass in biofuel production is the cost of the enzymatic treatments used to generate monosaccharides and the lack of microbial strains capable of fermenting lignocellulosic-derived sugars. The main industrial ethanol producers, *S. cerevisiae* and *Zymomonas mobilis*, cannot utilize xylose, the major pentose in plant biomass, as a carbon source for biofuel production. Attempts to overcome this problem by engineering the strains to utilize xylose have resulted in varying degrees of success. However, simultaneous, efficient and complete utilization of hexose and pentose sugars remains a challenge. *Geobacillus* is a promising organism for LCM utilisation as it is able to utilise hexose and pentose sugars.

Rice straw

As a waste product of the Indian agricultural industry, rice straw offers an abundant lignocellulose substrate for microbial fermentation. Rice is the major staple crop for most tropical nations and represents the third biggest crop grown globally. 2-4 crops per year mean rice straw is produced in large quantities. It has been estimated that 97.19 Mt of rice straw residue is produced in India annually. With 23% burned on the field for disposal purposes

(Gadde et al., 2009). This causes large scale emissions of black carbon, CO₂, methane and leads to large scale generation of tropospheric ozone. The UN Environment Programme recently identified reducing the burning of agricultural residues as a priority to help slow global warming. In November 2017 a public health emergency was declared in India's capital Delhi, as pollutant levels reached 710 micrograms per cubic metre, more than 11 times the World Health Organisation's safe limit. This pollution was linked to the burning of crop residues in neighbouring states.

Utilization of rice straw as a substrate for biofuel production will lead to a reduction in emissions caused by burning the straw and also generate a valuable product from the waste. This combination has the potential to provide great benefit, both environmental and economic. A commercial scale cellulosic biofuel plant operated by Novozymes in Crescentino, Italy, is currently producing 50 million litres of ethanol every year from rice straw.

1.5.2.2 Alternative substrates

Direct CO₂ utilization

An alternative to using biomass for biofuel production is the direct conversion of CO₂ into liquid fuels. Phototrophic organisms, such as eukaryotic microalgae and prokaryotic cyanobacteria, can fix CO₂. Atsumi et al. (2009) demonstrated photosynthesis based production of biofuel from CO₂ using recombinant cyanobacterium *S. elongatus*. Isobutanol was synthesized using the valine biosynthetic pathway. This approach has been used to convert electrical energy into chemical energy in the form of biofuel (Hawkins et al., 2011; Li et al., 2012), providing a way to store electrical energy.

Methane utilization

Methane is a greenhouse gas more potent than CO₂. There is, therefore, incentive to utilise methane as a carbon source. Methane can be produced from landfill, by anaerobic digestion of organic wastes and is the major

component of natural gas. The conversion of methane into liquid biofuels, is thermodynamically favourable and can be accomplished through chemical or biological methods. Microbial oxidation by methanotrophs is the only known biological sink of methane. Methanotrophs can be used as host organisms for biofuel production.

Waste protein

Protein from animal waste such as manure is a potential resource for biofuel production. With volumes increasing and a limit to the amount of manure which can be used as fertiliser, excess manure accumulates and disposal becomes more challenging. Recycling of animal waste into biofuel also re-uses reduced nitrogen (Huo et al., 2012). An *E. coli* strain capable of metabolising protein and its amino acid residues was engineered by Huo et al. (2012). This engineered strain was able to grow on 13 individual amino acids, as opposed to the wild type which could only use 4 amino acids as a sole carbon source.

Syngas

Syngas is a mixture of CO and H₂ and can be produced from renewable resources such as biomass and municipal waste by gasification. Syngas has been used as a substrate for production of butanol and acetone by *Clostridium ljungdahlii* and *Clostridium autoethanogenum* (LanzaTech and Daniell et al., 2012).

1.5.3 *n*-butanol

n-butanol, 1-butanol, butyl alcohol or BuOH is a naturally occurring four carbon alcohol with the formula C₄H₉OH. A colourless, flammable liquid mainly used as a solvent. *n*-butanol is an important platform chemical with broad applications including use as a biofuel.



Figure 1.7. The chemical structure of *n*-butanol. *n*-butanol is a straight chain isomer with the alcohol functional group at the terminal carbon.

n-butanol has a molar mass of 74.12 g/mol, a boiling point of 117°C, solubility in water of 68g/l at 25°C, density of 810 kg/m³ and a dissociation constant of 16.1pKa at 25°C (Ndaba and Marx, 2015). *n*-butanol combustion; $2C_4H_9OH + 5O_2 \rightarrow 8CO_2 + 10H_2O$. Combustion of *n*-butanol results in 29 MJ/kg. Slightly less than petrol (33MJ/KG) but more than ethanol (20MJ/kg). Combustion of *n*-butanol is preferable to both petrol and ethanol, producing less carbon dioxide per MJ. *n*-butanol produces 2.03kg of carbon dioxide per kilogram while petrol produces 3.3kg of carbon dioxide per kilogram.

n-butanol is considered a next generation biofuel due to many advantages over ethanol. *n*-butanol has higher energy density, lower hygroscopicity and is less volatile (Atsumi et al., 2008). *n*-butanol can be used as a substitute for petrol in the current petroleum infrastructure. It can be blended with petrol in any proportion and can be used in existing car engines without any modifications (Baral et al., 2014). Not only is *n*-butanol a superior transportation fuel, it can also easily be converted to acrylates, ethers, and butyl acetate which are utilized in paints, resin formulations and lacquers (Harvey and Meylemans, 2011). This makes *n*-butanol a valuable product.

1.5.4 Microbial *n*-butanol production

The development of efficient, low cost and environmentally friendly processes for biofuel production requires the selection of a suitable production organism. Microbial production strains must be robust against several challenges in industrial settings, such as high product concentration, high temperature and high acid levels.

Table 1.1. Summary of microbial *n*-butanol production (Chen and Liao, 2016).

Organism	Substrate	Pathway	Gene overexpressed	Strain engineering	Titer	Yield	Fermentation reactor	Reference
<i>E. coli</i>	TB + glycerol	CoA-dependent pathway	<i>atoB, hbd, crt, bcd, etfAB, adhE2</i>	$\Delta adhE, \Delta ldhA, \Delta frdBC, \Delta fnr, \Delta pta$	0.552 g L ⁻¹	–	Flask	Atsumi et al. (2008)
	Glucose	Reverse β -oxidation	<i>yqeF, fucO</i>	$\Delta arcA, \Delta adhE, \Delta pta, \Delta frdA, \Delta yqhD, \Delta eutE, fadR^*, atoC(Con), crp^*$	2.2 g L ⁻¹	0.28 g g ⁻¹ glucose	Flask	Dellomonaco et al. (2011)
	Glucose	Modified CoA-dependent pathway	<i>atoB, hbd, crt, ter, adhE2, fdh</i>	$\Delta adhE, \Delta ldhA, \Delta frdBC, \Delta pta$	14 g L ⁻¹	0.33 g g ⁻¹ glucose	Bioreactor, batch	Shen et al. (2011)
					15 g L ⁻¹	0.36 g g ⁻¹ glucose	Tube	
					30 g L ⁻¹	0.29 g g ⁻¹ glucose	Bioreactor, fed-batch, in situ product removal	
	Glucose	Keto acid pathway via citramalate	<i>cimA3.7, leuABCD, kivD, ADH2</i>	$\Delta ilvI, \Delta ilvB$	0.524 g L ⁻¹	–	Flask	Atsumi and Liao (2008)
	Glucose	Keto acid pathway via threonine	<i>thrA^{Br}BC, ilvA, leuABCD, kivD, ADH2</i>	$\Delta ilvI, \Delta ilvB, \Delta metA, \Delta tdh, \Delta adhE$	1 g L ⁻¹	–	Flask	Shen and Liao (2008)
<i>S. cerevisiae</i>	Galactose	CoA-dependent pathway	<i>thl, hbd, crt, bcd, etfAB, adhE2</i>	–	2.5 m g L ⁻¹	–	Vial	Steen et al. (2008)
<i>C. cellulovorans</i>	Cellulose	CoA-dependent pathway	<i>adhE2</i>	–	1.42 g L ⁻¹	0.39 g g ⁻¹ cellulose	Bottle	Yang, Xu and Yang (2015)
<i>T. saccharolyticum</i>	Xylose	CoA-dependent pathway	<i>thl, hbd, crt, bcd, etfAB, adhE2</i>	$\Delta ldh, ermR$	1.05 g L ⁻¹	0.10 g g ⁻¹ xylose	Tube	Bhandiwad et al. (2014)
<i>P. putida</i>	Glycerol	CoA-dependent pathway	<i>thl, hbd, crt, bcd, etfAB, adhE1</i>	–	0.122 g L ⁻¹	–	Flask	Nielsen et al. (2009)
<i>L. brevis</i>	Glucose	CoA-dependent pathway	<i>thl, hbd, crt, bcd, etfAB</i>	–	0.300 g L ⁻¹	–	Vial	Berezina et al. (2010)
<i>B. subtilis</i>	Glycerol	CoA-dependent pathway	<i>thl, hbd, crt, bcd, etfAB, adhE2</i>	–	24 m g L ⁻¹	–	Flask	Nielsen et al. (2009)
<i>S. elongatus</i> PCC7942	CO ₂	CoA-dependent pathway	<i>atoB, hbd, crt, ter, adhE2</i>	–	15 m g L ⁻¹	–	Tube	Lan and Liao (2011)
	CO ₂	Malonyl-CoA dependent modified CoA pathway	<i>nphT7, phaB, phaJ, ter, bldh, yqhD</i>	–	30 m g L ⁻¹	–	Flask	Lan and Liao (2012)
	CO ₂	Malonyl-CoA dependent modified CoA pathway	<i>nphT7, crt, hbd, ter, pdup, yqhD</i>	–	0.404 g L ⁻¹	–	Flask	Lan, Ro and Liao (2013)
<i>C. acetobutylicum</i> ATCC824	Glucose	CoA-dependent pathway	<i>adhE1^{D485G}</i>	$\Delta buk, \Delta pta$	18.9 g L ⁻¹	0.29 g g ⁻¹ glucose	Bioreactor, batch	Jang et al. (2012a)
	Glucose	CoA-dependent pathway	<i>groESL operon</i>	–	17.1 g L ⁻¹	–	Bioreactor, fed-batch	Tomas, Welker and Papoutsakis (2003)
<i>C. beijerinckii</i> BA101	Glucose	CoA-dependent pathway	–	–	18.6 g L ⁻¹	0.32 g g ⁻¹ glucose	Bioreactor, batch	Formanek, Mackie and Blaschek (1997)

Table 1.1 shows some of the recombinant organisms investigated for butanol production. A range of genes have been targeted from different pathways. This demonstrates the diversity of engineering approaches available, all with the same goal of high butanol titre.

1.5.4.1 Native *n*-butanol production

n-butanol is naturally produced by certain species of *Clostridium*. Discovered by French microbiologist Louis Pasteur in 1861. *n*-butanol was first commercially produced by solventogenic *C. acetobutylicum* from sugar a century ago (Jones and Woods, 1986). The fermentation co-produced acetone, butanol and ethanol. Despite success, biological production of *n*-butanol was outcompeted in the 1950's due to cheap oil prices and development of a petrochemical production route (Green, 2011). The interest in microbial production of *n*-butanol has re-emerged in recent years, with many manufacturing plants producing *n*-butanol through fermentation processes across the world (Lan and Liao, 2013).

The highest *n*-butanol titre reported in *C. acetobutylicum* is 18.9g/l (Jang et al., 2012). Butanol yields in clostridia are limited by slow growth rate and intolerance to butanol above 13 g/l (Peralta-Yahya et al., 2009). Improvements have been made to the ABE fermentation process including; utilization of low-cost feedstocks, butanol tolerance, oxygen tolerance, product selectivity and improving cell density and sustainable viability (Chen and Liao, 2016).

1.5.4.2 Heterologous *n*-butanol production

Some clostridial limitations can be overcome by the introduction of its butanol pathway into organisms that grow faster, can tolerate high concentrations of butanol or can metabolize alternative feedstocks (Peralta-Yahya et al., 2009). The *n*-butanol production pathway has been introduced to various non-native host organisms with an aim to explore a wider range of metabolic capability.

The *Clostridium n*-butanol pathway has been reconstructed and expressed in *E. coli* with the aim of taking advantage of high growth rates, high protein expression levels and the efficiency of genetic tools. An engineered *E. coli* strain successfully produced 15g/l of *n*-butanol, a titre comparable to native *Clostridium* production (Shen et al., 2011).

Saccharomyces cerevisiae is commonly used in industrial processing due to high growth rates and the efficiency of genetic tools. Yeast's robustness, extensive fermentation knowledge, the availability of genetic tools, high tolerance to industrial conditions and solvents (butanol tolerance of more than 20 g/l), low media pH and a lack of susceptibility to bacteriophage make it an attractive organism for the production of butanol. However, it is unable to digest the pentose sugars present in lignocellulosic biomass. Steen et al. (2008) constructed a butanol biosynthetic pathway in *S. cerevisiae*. The butanol yield obtained from the engineered *S. cerevisiae* (2.5 mg/l) was two orders of magnitude lower than that obtained from *E. coli* (550 mg/l). The authors suggest a potential bottleneck at the final reduction steps in the pathway due to the insolubility of AdhE2. Additionally, this may be explained by the fact that acetyl-CoA is the precursor for butanol production. In yeast, some of the acetyl-CoA pool is trapped in the mitochondrion, preventing it from being utilized in the cytosolic butanol pathway (Peralta-Yahya and Keasling, 2010).

Pseudomonas putida, *Bacillus subtilis* and *Lactobacillus brevis* were used for their potentially higher solvent tolerance (Nielsen et al., 2009; Berezina et al., 2010). *Pseudomonas putida* can overcome toxicity using efflux pumps; *Bacillus subtilis* can change its cell wall composition in response to solvent toxicity; *L. brevis*, in addition to *n*-butanol tolerance, digests C5 and C6 substrates. The cyanobacterium *S. elongatus* was used for direct conversion of CO₂ to *n*-butanol by photosynthesis (Lan and Liao, 2011). Algal biomass is a compelling alternative because it can be produced in salt water rather than on arable land, however the collection and dewatering is challenging (Peralta-Yahya et al.,

2012). Despite advantageous properties, the *n*-butanol titres produced from these recombinant organisms are much lower than that of the native clostridia producers. To date none of these alternative organisms have reported a titre above 1.5 g/l. Limited success in alternative hosts could be due to a number of factors including; enzyme oxygen sensitivities, less well-defined redox reactions, the butyryl-CoA dehydrogenase electron transferring flavoprotein (Bcd/Etf) complex being poorly expressed in the recombinant organisms, the requirement of ferredoxin as an additional redox partner (Li et al., 2008). These results demonstrate the significant further work required in order develop a successful *n*-butanol production strain.

1.6 *Geobacillus*; a host organism

The *Geobacillus* genus contains sixteen species of thermophilic bacteria. *Geobacillus* bacteria have attracted recent attention for exploitation in a range of biotechnological applications. Applications include; sources of thermostable enzymes, as platforms for biofuel production and as potential components of bioremediation (Zeigler, 2014). The organism is currently commonly used as a biological indicator for steam sterilisation.

1.6.1 Properties

Geobacillus are aerobic, rod shaped bacilli. Gram-positive spore formers widely distributed in nature. *Geobacillus* are thermophilic; capable of growth between 37°C and 75°C. The bacteria have catabolic versatility, particularly in the degradation of hemicellulose and starch. They are able to ferment both hexose and pentose sugars to generate lactate, formate, acetate and ethanol products (Nazina et al., 2001). All *Geobacillus* spp. use a combination of the Emden-Meyerhof-Parnas glycolysis pathway and the oxidative pentose phosphate pathway for carbohydrate metabolism. They have a classical TCA cycle and glyoxylate cycle. Further characteristics of the genera are summarised in Table 1.2.

Table 1.2. Description of the *Geobacillus* genus.

Morphology	
Vegetative cell	Rod-shaped cells, occurring either singly or in short chains and motile by means of peritrichous flagella. The cell wall structure is Gram-positive, but the Gram-stain reaction may vary between positive and negative.
Spore	One ellipsoidal or cylindrical endospore per cell, located terminally or subterminally in slightly swollen or non-swollen sporangia
Colony	Variable shape and size; pigments may be produced on certain media.
Metabolism	
Energy	Chemo-organotrophic
Oxygen	Aerobic or facultatively anaerobic. O ₂ is the electron acceptor, replaceable in some species by nitrate.
Temperature	Obligately thermophilic. The growth-temperature range is 37-75°C, with an optimum at 55-65°C.
pH	Growth occurs in a pH range of 6.0 to 8.5, with an optimum at pH 6.2-7.5.
Requirements	Growth factors, vitamins, NaCl and KCl are not required by most species.
Identification tests	
Carbohydrates	Acid but no gas is produced from glucose, fructose, maltose, mannose and sucrose. Most species do not produce acid from lactose.
Enzymes	Most species form catalase. Phenylalanine is not deaminated, tyrosine is not degraded, indole is not produced, the Voges-Proskauer reaction is negative. Oxidase-positive or negative.
Biochemicals	The major cellular fatty acids are iso-15:0, iso-16:0 and iso-17:0, which make up more than 60% of the total. The main menaquinone type is MK-7.
DNA characterization	
G-C content	48.2-58 mol% (thermal denaturation method)
16S rRNA	Sequence identities higher than 96.5% among the members of this genus
Ecology	
Prevalence	Most species are widely distributed in nature.

(*Bacillus* Genetic Stock Centre, Catalogue of Strains, Seventh Edition, Volume 3: The Genus *Geobacillus* 2001)

As thermophilic bacteria, *Geobacillus* have genes for the production of protamine and spermine, small cationic histone-like proteins, which allow chromosome condensation and increase the T_m of DNA (Takami et al., 2004), potentially enabling growth at high temperatures.

1.6.2 Taxonomy

Geobacillus belong to the family *Bacillaceae*. The family contains mostly saprophytic bacteria and are therefore commonly found in soil. As of May 2018, there were sixteen *Geobacillus* species (*G. caldxylosilyticus*, *G. galactosidasius*, *G. icigianus*, *G. jurassicus*, *G. kaustophilus*, *G. lituanicus*, *G. stearothermophilus*, *G. subterraneus*, *G. thermantarcticus*, *G. thermocatenulatus*, *G. thermodenitrificans*, *G. thermoglucosidasius*, *G. thermoleovorans*, *G. toebii*, *G. uzenensis* and *G. vulcani*) described with validly

published names (Parte, 2018). There are genome sequences publicly available for sixty-three *Geobacillus* strains.

INTRODUCTION

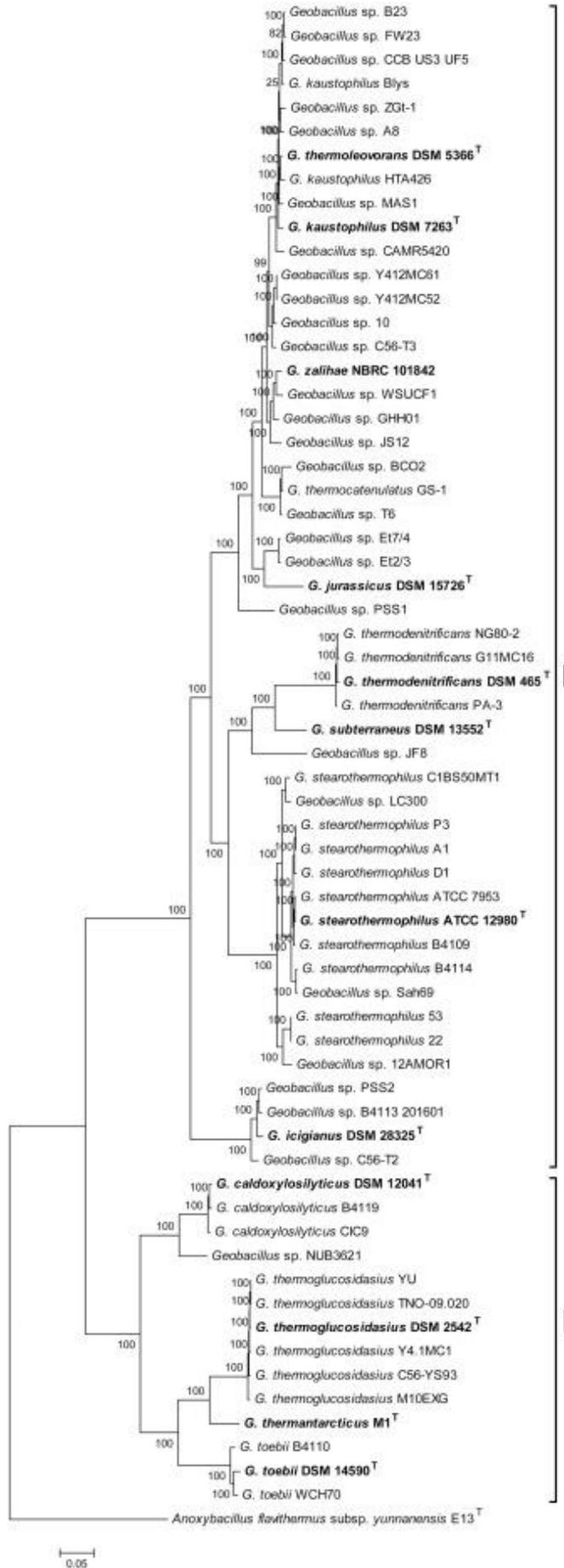


Figure 1.8. Whole genome phylogeny of the genus *Geobacillus* (Aliyu et al., 2016). The maximum likelihood tree was constructed based on the alignment of 1048 concatenated core genes (total alignment length: 584,424 nucleotides) of sixty-three *Geobacillus* strains and *Anoxybacillus flavithermus* E13T (used as outgroup). The values at the nodes indicate bootstrap values expressed as percentages of 1000 replications while the bar length indicates 0.05 substitutions per site.

There is a distinct clustering of *Geobacillus* species into two clades, showing low genomic similarity and distinct nucleotide based compositions. This suggests *Geobacillus* may consist of two distinct genera.

1.6.3 Habitat

Despite thermophilic requirements, *Geobacillus* have been found widely distributed in nature and have been isolated from a diverse range of environments.

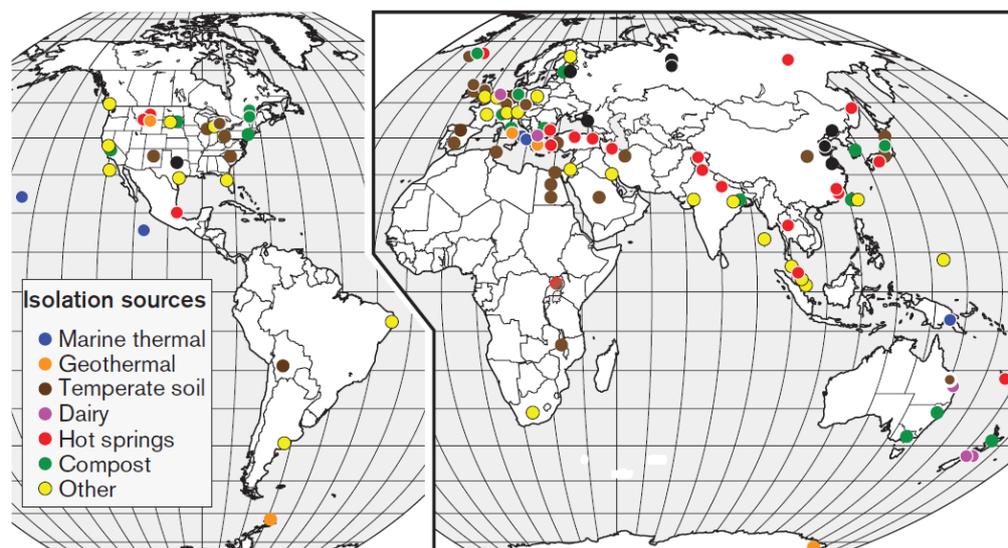


Figure 1.9. Geographical and environmental diversity of *Geobacillus* isolation (Zeigler, 2014). Circles represent the geographical location of *Geobacillus* isolate sources. Circle colour denotes the type of source material.

The most frequently reported natural sources for *Geobacillus* isolates have been hot springs, geothermal soils, hot subterranean oil fields and natural gas wells and hydrothermal vents (Zeigler, 2014). However, *Geobacillus* have achieved worldwide distribution with significant population numbers. The bacteria have been isolated from diverse locations across the world at a range of temperatures including temperate and cold environments. Despite temperatures being significantly below the minimum growth requirement. For example; 33 *Geobacillus* isolates were identified in soil samples collected in Iceland, where temperatures averaged 14°C (Fields & Chen Lee, 1974). More recently, soils in Northern Ireland were found to contain 1.5–8.8 x 10⁴ /g aerobic thermophiles, including five *Geobacillus* species, despite soil temperatures never reaching the minimum growth requirement (Marchant et al., 2002).

Zeigler (2014) suggests the natural role of *Geobacillus* is an opportunistic decomposer of plant derived organic matter. Capable of rapid growth under transient thermophilic conditions. With highly resistant endospores ensuring long term survival when conditions are no longer favourable for growth. The spore morphology is slightly ovoid, with an average diameter 1mm. Particles of this size have the longest residence in the atmosphere (Burrows et al., 2009). Allowing the spores to be transported great distances, to more favourable locations for growth. Extreme resistance to UV light, temperature changes and desiccation, means spore longevity could be extreme. Based on extrapolation the longevity of *Geobacillus* spores at 40°C has been estimated to be 1.9 billion years (Zeigler, 2014). This would enable viable spores to accumulate gradually over time to achieve high population densities.

1.6.4 Biotechnological application

Microorganisms are utilised for production of a wide range of products. *Geobacillus* bacteria are currently utilised for a range of biotechnological applications. Table 1.3 lists a selection of the applications patented in the U.S,

with products including; thermostable enzymes, biofuels and biological indicators. As thermophilic bacteria, many of the patented applications utilise the thermostable properties of *Geobacillus* and its proteins.

Table 1.3. A selection of U.S patents for *Geobacillus* products or processes.

Product or Process	Patent No.
α -arabinofuranosidase	US05434071
acetate kinase	US05610045
alpha-amylase	US05824532, US05849549
arabino furanoside	US05491087
biological indicator for sterilization	US05073488, US05223401, US05252484, US05418167
BsrFI restriction endonuclease	US06066487
catalase	US06022721
cellobiose fermentation	US06102690
DNA polymerase	US05747298, US05830714, US05834253, US05874282, US06013451, US06066483, US06100078, US06238905
ethanol production	US05182199
glucose-6-phosphate dehydrogenase	US04331762
liquefying starch	US05756714
maleate dehydrogenase	US04331762
neutral proteases	US06103512
perillyl compounds	US05487988
polynucleotide phosphorylase	US04331762
prenyl diphosphate synthase	US06225096
pyruvate kinase	US04331762
riboflavin glucoside	US06190888
superoxide dismutase	US05772996
xylanase	US05434071
xylosidase	US05489526

(*Bacillus* Genetic Stock Centre, Catalogue of Strains, Seventh Edition, Volume 3: The Genus *Geobacillus* 2001)

Table 1.4. Essential and desirable traits for efficient fermentation of lignocellulosic material. Traits as identified by Picataggio and Zhang (1996). GRAS (generally regarded as safe) defined by the US Food and Drug Administration agency.

Essential traits	Desirable traits
Broad substrate utilization range	Simultaneous sugar utilization
High yield and productivity	Hemicellulose and cellulose hydrolysis
Minimal by product formation	GRAS status
High alcohol tolerance	Recyclable
Increased tolerance to inhibitors	Minimal nutrient supplementation
Tolerance to process hardness	Tolerance to low pH and high temperature

As a potential host organism for biofuel production, *Geobacillus* meets both the essential and desirable traits listed in Table 1.4. These properties make *Geobacillus* a promising choice organism for development.

1.6.5 High temperature fermentation

Advantages

Microbial production of fuels and chemicals from plant biomass at elevated temperatures is desirable. The use of high temperature fermentation has advantages over mesophilic processes including; reduced risk of contamination by mesophilic organisms, reduced cost and energy requirements for fermenter cooling and product separation, fermentation and alcohol recovery can be coupled as high temperature facilitates the removal of volatile products. Thermophiles may have lower sensitivity to organic solvents. The chemistry of processes may be accelerated at high temperatures, resulting in higher rates of feedstock conversion. Thermophilic bacteria with increased metabolic activity result in more efficient product formation (Taylor et al., 2011). As most cellulases have optimum temperatures of 50-55°C, less enzyme is needed and simultaneous cellulose hydrolysis and fermentation can take place (Lin et al., 2014). In addition to biofuel production, high temperature bioprocessing allows functional production of thermostable enzymes and efficient expression of enzymatic activities *in vivo* (Suzuki et al., 2013).

Disadvantages

High temperature fuel and chemical production faces challenges, including; enzyme stability, volatility of pathway intermediates and increased product toxicity (Lin et al., 2014). Additional heating costs may be encountered.

Examples

Xiao et al. (2012) demonstrated fermentation of two important biorefinery platform chemicals; Acetoin and 2,3-butanediol using a novel *Geobacillus* strain at temperatures of 45-55°C. Overcoming the problem of bacterial contamination in a previously mesophilic process.

To date, thermophilic biofuel production has been reported for; ethanol production using *G. thermoglucosidasius* (Cripps et al., 2009),

Thermoanaerobacterium saccharolyticum (Shaw et al., 2008) and *Clostridium thermocellum* (Argyros et al., 2011). Isobutanol production using *G. thermoglucosidasius* (Lin et al., 2014). *n*-butanol production using *T. saccharolyticum* (Bhandiwad et al., 2013). These examples demonstrate the feasibility of producing higher density liquid fuels at a high temperature, essential for producing next generation cellulosic biofuels using simultaneous saccharification and fermentation.

1.6.6 *Geobacillus thermoglucosidasius*

Geobacillus thermoglucosidasius may be referred to in literature as *G. thermoglucosidans*, *Parageobacillus thermoglucosidasius* or *Parageobacillus thermoglucosidans*.

G. thermoglucosidasius is a species of the *Geobacillus* family isolated from soil in Japan. *G. thermoglucosidasius* is in clade II with G + C content (42.1–44.4%). With starch hydrolyzing glucosidase activity. Most strains hydrolyze gelatin, pullulan, and starch and produce acid from adonitol, cellobiose, inositol, and D-xylitol. This species has been selected by various groups as the organism of choice for process development.

The *G. thermoglucosidasius* genome encodes 41 transposases, with 39 associated with insertion sequence elements (Hussein et al., 2015). CRISPR motifs are present in the genome with over 112 CRISPR-associated proteins. Complete prophage sequences can be found within the genomes of *G. kaustophilus*, *G. thermoleovorans* and *G. thermodenitrificans* but not in the genome of *G. thermoglucosidasius* (Hussein et al., 2015). The abundance of CRISPR sequences and Cas genes found in the genome *G. thermoglucosidasius* might explain the lack of prophage sequence within its genome. CRISPR regions as well as genes of unassigned function found between transposable elements means *G. thermoglucosidasius* has the largest genome of any *Geobacillus*.

Methods for genetic modification of most *Geobacillus* species are under developed. An exception is *G. thermoglucosidasius*. This species has an established transformation protocol and an integration vector system for gene replacements (Cripps et al., 2009). In addition Lin et al., (2014) identified a robust promoter to drive gene expression from a plasmid. *G. thermoglucosidasius* is therefore a choice organism to build on the metabolic manipulation toolkit.

1.6.7 Biofuel production using *G. thermoglucosidasius*

As previously discussed, *Geobacillus* offers many advantages for industrial application. Thermostability is a phenotype almost impossible to engineer into other cells (Martinez-Klimova, 2014). To gain thermophilic advantages, existing biofuel production pathways can be improved or diversified in *Geobacillus* species to generate high yields and new products. Production of *n*-butanol has yet to be reported in the literature.

Cripps et al. (2009) describe the metabolic engineering of *G. thermoglucosidasius* for ethanol production. They eliminated the lactate dehydrogenase and pyruvate formate lyase pathways by disruption of the *ldh* and *pf1B* genes, respectively, together with up regulation of expression of pyruvate dehydrogenase using knock out plasmids and an integration vector. The strain containing the modifications was able to ferment cellobiose and a mixed hexose and pentose feed at temperatures in excess of 60°C.

Lin et al. (2014) report the engineering of *G. thermoglucosidasius* to produce isobutanol at 50°C. They characterized thermostabilities of enzymes in the isobutanol synthesis pathway. Using an expression system based on the lactate dehydrogenase promoter from *G. thermodenitrificans*, they were able to produce isobutanol from both glucose and cellobiose using the valine biosynthesis pathway. Enzymes Kivd from *L. lactis*, KARI from *G. thermoglucosidasius* and AlsS from *B. subtilis* were used. Lin et al. (2014)

developed an expression system using the strong *ldh* promoter (P_{ldh}) from *G. thermodenitrificans*.

1.7 Metabolic engineering for fuel and chemical production

Metabolic engineering is a key technology for the alteration of microorganisms to produce desired compounds more efficiently (Lee et al., 2012). Microorganisms can be used for the production of a diverse range of chemicals. However, natural microbial metabolism will produce a given molecule in low concentration. Metabolic engineering of microbes is used to improve efficiency, generating greater quantities of a molecule of interest. Approaches include; creation of new metabolic enzymes and pathways or modification of existing pathways for optimal production of desired products. With the aim of making a biological system economically viable. It has been suggested that the most convenient and cost effective approach for large scale production of advanced biofuels may be the engineering of microorganisms (Peralta-Yahya and Keasling, 2010).

For biofuel production, metabolic engineering is used to direct the metabolic flux in an organism to the desired product. The production organism must first convert raw materials to central metabolites such as pyruvate and acetyl Coenzyme A (CoA), which are then diverted to fuel synthesis through engineered pathways (Lan and Liao, 2013). This can be done by either; constructing the resource utilization pathways into an organism that already produces the desired product, or construct the biofuel production pathways into a host organism capable of metabolizing the desired resource (Lan and Liao, 2013). Both approaches have demonstrated successful biofuel production.

1.7.1 Methods for engineering *Geobacillus*

A major barrier to development of thermophilic processes has been inability to genetically modify thermophile host strains. Until recently there has been a

lack of reliable methods for inducing competence, genetic material transfer, gene expression and genome integration in thermophiles (Taylor et al., 2011). The development of stable, reliable and reproducible genetic systems for thermophilic organisms is required to produce commercially viable organisms with high yields in valuable bioproducts (Peralta-Yahya et al., 2012). An increase in knowledge of thermophile physiology is also needed (Taylor et al., 2011). Recent genetic method development programmes have been successful in developing thermostable genetic systems in a range of commercially relevant organisms. However, improving the efficiency and yield of biofuels further will require more potent tools including techniques for stable maintenance of gene copy number and more precise control of mRNA and protein levels, especially in a dynamic manner that would automatically adjust the pathway according to its own metabolic flux (Peralta-Yahya et al., 2012).

Synthetic biology relies on characterized parts and tools to build new functionalities into organisms. Tools developed to date include a basic modular replicative plasmid set which has been described and applied to facilitate bioethanol production (Reeve et al., 2016; Sheng et al., 2017). This vector set included an origin of replication, selectable marker and reporter genes. The introduction of genetic material into a host strain is crucial to the establishment of a molecular toolkit. Electroporation is the most widely used transformation method applied to thermophilic prokaryotic species.

Efficient promoters for gene expression have been characterized for only a few thermophiles. The strongest and most commonly used promoter in *G. thermoglucosidasius* to date is the *G. stearothermophilus* lactate dehydrogenase promoter (P_{ldh}). P_{ldh} has been applied in the production of ethanol and isobutanol (Cripps et al., 2009 and Lin et al., 2014).

Sheng et al. (2017) established a gene knock-out/knock-in system to engineer the metabolism of *G. thermoglucosidasius*. This system utilises a temperature

sensitive replicon and a heterologous *pyrE* gene from *G. kaustophilus* as a counter-selection marker. Its use requires the initial creation of uracil auxotroph through deletion of the native *pyrE* gene using allele-coupled exchange (ACE) and selection for resistance to 5-fluoroorotic acid. Using this method cargo DNA can be integrated into the genome at the *pyrE* locus. This system was used to make in-frame deletions of two genes; lactate dehydrogenase and pyruvate formate lyase.

Allele Coupled Exchange (ACE) is a DNA integration method developed by Heap et al. (2011). Exploiting inactivation and subsequent reactivation of genes leading to a selectable phenotype. This technique utilises asymmetrical regions of homology to control the order of recombination events. The *pyrE* gene encodes orotate phosphoribosyltransferase, an enzyme involved in pyrimidine biosynthesis. Mutant strains defective in *pyrE* created using ACE become auxotrophs requiring a source of uracil to grow. 5-fluoroorotate (FOA) is toxic to *pyrE* mutant cells. Toxicity occurs as mis-incorporation of fluorinated nucleotides into DNA and RNA cause cell death. This phenotype is used as a selection marker. Similarly Suzuki et al. (2012) describe a counter selection system for *Geobacillus kaustophilus* through disruption of *pyrF* and *pyrR* genes.

In addition, recently a CRISPR method has been developed to manipulate *G. thermoglucosidasius*, not yet published. Application of CRISPR-Cas9 has the potential to offer a powerful tool for genome editing.

There are currently sixty-three *Geobacillus* genome sequences publicly available (Aliyu et al., 2016). The availability of genome sequences is essential for identification of genes of interest and the rapid advancement of rationally designed metabolic engineering. The availability of genome sequences combined with new molecular biology techniques enables rapid engineering of *Geobacillus* species.

1.7.2 Pathway control

Metabolic pathways can be controlled in several ways. The number of gene copies can be manipulated by increasing how many are integrated into the genome or by placing the gene in a vector with high or low copy number. The strength of constitutive and inducible promoters can be chosen to match the desired transcription initiation rate, and orthogonal, inducible promoters can be used to regulate parts of pathways, or multiple pathways independently. Synthetic terminators can be engineered to control transcription termination efficiencies. The metabolic pathway can be modulated at the translational level by inserting functional RNA segments into intergenic regions of operons to regulate the processing and stability of mRNAs that encode individual pathway enzymes. The strength of ribosome binding sites can be predicted and designed using computational methods, allowing the control of the protein translation efficiencies. Enzyme stability can be altered through the control of protein degradation rate using peptide tags (Peralta-Yahya et al., 2012).

These controls may be effective in engineered pathways, which are optimized on the basis of an initial set of conditions. However, the conditions may change when used in bioreactors. If the system cannot respond this could lead to suboptimal production of biofuels. To address this, Zhang et al. (2012) designed a dynamic sensor-regulator system which is able to adjust the metabolic pathway according to the metabolic status of the host and regulate product production. The group applied this technique to fatty acid ethyl ester biosynthesis in engineered *E. coli*. The acyl-CoA intermediate was monitored, controlling the expression of genes to regulate pathway intermediates dynamically. This approach increased yield three-fold and reduced the concentration of toxic intermediates.

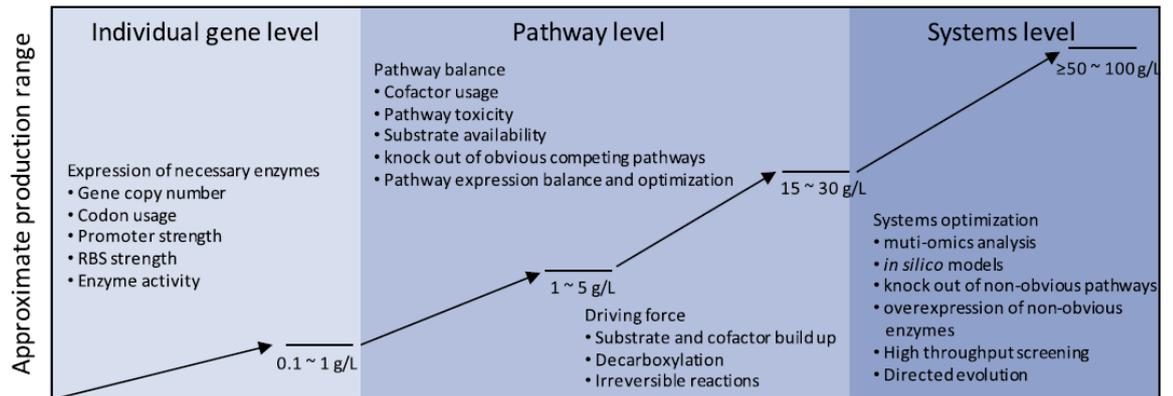


Figure 1.10 Systematic approach of strain development for the production of chemicals from engineered organisms (Lan & Liao, 2013). Production ranges shown are approximate values.

For optimal strain development a systems biology approach should be adopted to look at interactions in the system as a whole, further detailed in Figure 1.10. Advances in “Omics” technologies and reductions in the cost of sequencing mean the creation of genome scale metabolic models is more feasible. Figure 1.10 predicts that taking a systems based approach significantly higher production ranges could be achieved.

1.8 Optimising biofuel production

For biofuel production to be economically viable low cost and high volumes are required. Therefore production must be thoroughly efficient. Every opportunity to increase yield, titre and productivity needs to be explored. Almost all metabolic engineering strategies start with overexpressing the desired pathway genes, followed by knocking out competing pathways that drain precursors, products or cofactors. In addition increasing production of the precursor, circumventing or deleting the native regulatory loops or constructing artificial regulatory circuit can also be beneficial in increasing titre. Mutation and selection of desired phenotypes in the host organism may also improve yields.

1.8.1 Balancing co-factors

Genetically modifying metabolic pathways may cause cells to become unstable. Editing or knocking out pathways can cause an imbalance in co-factor ratios. This can have an impact on production, growth rates and cell survival. Understanding the electron flux in a host organism often the key to high-yield production (Liao et al., 2016).

Metabolic networks within a cell are interdependent and will be strongly influenced by levels of cellular metabolites. ATP/ADP, NAD⁺/NADH, NADP⁺/NADPH, and acyl-CoAs all play significant roles in regulating numerous cellular pathways (Lee et al., 2008). The relative ratios of these metabolites regulate pathway activity. An NADH/NADPH producing pathway must match a consumption pathway using the same redox co-factor for streamlined co-factor regeneration resulting in a balanced system. Identifying the driving force for pathway flux often leads to useful strategies for increasing production. Efficient metabolic pathways need to be driven by both kinetic and thermodynamic driving forces, which can be manipulated through the size of the metabolite or cofactor pool (Liao et al., 2016). Introduction of a new synthetic pathway requires amino acids, energy for synthesis and redox factors, which imparts a metabolic cost on recombinant protein synthesis (Howard et al., 2013).

1.8.2 Consolidated bioprocessing

The third stream in the Ricefuel project aims to bring the two previous components together, incorporating rice straw deconstruction and biofuel production into a single process capable of producing advanced biofuels from rice straw.

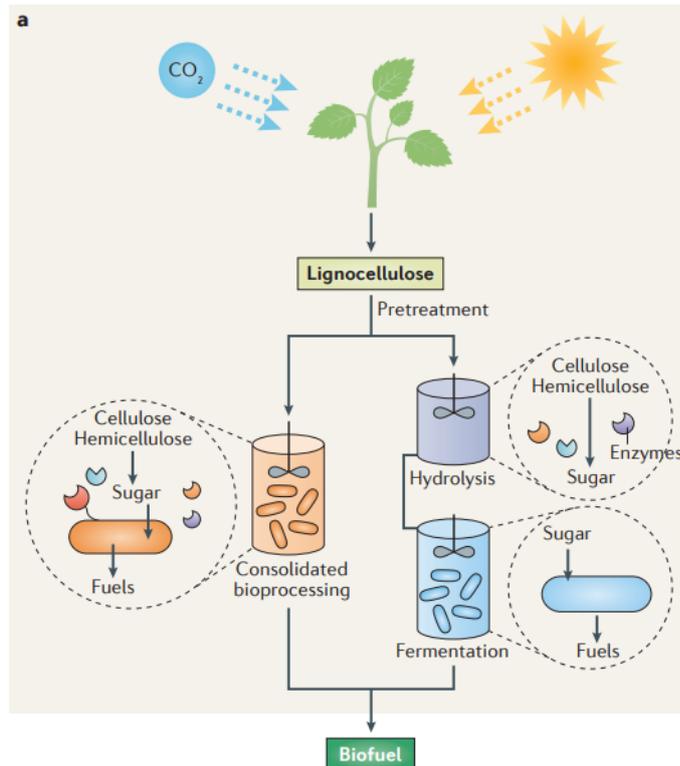


Figure 1.11 Consolidated bioprocessing (Liao et al., 2016). The lignocellulose of plant biomass can be converted to fuels through hydrolysis followed by fermentation, or through consolidated bioprocessing, which combines the two processes in one reactor.

Consolidated bioprocessing can use one or several organisms. Here the desire is for a consolidated bioprocess in a single organism; *G. thermoglucosidasius*. Native cellulolytic organisms have been engineered to produce ethanol. The cellulolytic thermophile *C. thermocellum* is a leading candidate for consolidated bioprocessing owing to its rapid cellulolytic function and growth rate. A less characterized cellulolytic thermophile, *C. bescii*, has also emerged as a potential candidate for consolidated bioprocessing, owing to its impressive rate of biomass deconstruction. This organism is not naturally ethanologenic, but has been engineered to produce ethanol (yielding 700 mg/l). Another approach is to engineer alcohol producers to digest cellulose. Examples of this approach include; *S. cerevisiae*, *E. coli* and *Z. mobilis* engineered to digest cellulose through expression of heterologous cellulolytic enzymes. However,

much work is still required to bring biofuel production by these engineered microorganisms to an industrial level.

As thermophilic bacteria, *Geobacillus* also offer advantages as a potential host for consolidated bioprocessing. Cellobiohydrolases and endoglucanases have been engineered to increase their optimal operating temperature to 70°C by directed evolution. These engineered enzymes performed better than their wild-type counterparts (Liao et al., 2016), demonstrating the benefit of using higher temperatures in cellulose hydrolysis.

1.8.3 Pathway flux

Enzyme activity is a common bottleneck in the engineering of metabolic pathways (Peralta-Yahya et al., 2012). Many native enzymes are regulated by metabolites at the post translational level. To overproduce enzymes to increase pathway flux may require feedback inhibition resistant mutants. In addition, synthetic protein scaffolds have been used to spatially assemble metabolic enzymes and to catalyse multistep reactions synergistically (Peralta-Yahya et al., 2012).

Protein engineering to increase enzyme activity may be an alternative to increasing the quantity of enzyme products. As protein synthesis is energy intensive, overexpression of the butanol production enzymes is likely to disrupt biosynthesis of other metabolites that indirectly favour butanol production (Zheng et al., 2009). Using rational design and directed evolution to protein engineer could result in enzymes with much higher turnover rates. Alternatively, artificial enzymes with new functions can be created by incorporating unnatural amino acids and computational protein design (Peralta-Yahya et al., 2012).

1.9 Challenges of biofuel production

To make biofuel innovations sustainable, commercialisation on a massive scale is needed. Advanced biofuels must be economically competitive with existing products, overcoming the primary economic drivers of feedstock price, and overall process productivity and yield (Peralta-Yahya et al., 2012). Commercialisation is supported by governments across the world with policies in place to encourage biofuel uptake. Further use of biofuels is also required to mitigate the effects of greenhouse gasses and as an alternative to finite fossil fuels.

1.9.1 Theoretical yields

When designing biofuel pathways and production processes, the theoretical limitations should be considered. In this case, the biosynthesis of *n*-butanol starts with glycolysis splitting sugar to pyruvate and then to acetyl-CoA. This is the starting metabolite for almost all biofuels. The decarboxylation of pyruvate to acetyl-CoA results in loss of one-third of the carbon in the form of CO₂ or formate. The loss of carbon is necessary to gain reducing power and ATP. However it limits the maximum theoretical carbon yield to 66.6% for any products that are derived from acetyl-CoA. This loss of carbon also releases GHGs into the environment.

In order to achieve total carbon conservation between glucose and acetyl-CoA the organism must fix CO₂ via the natural Wood–Ljungdahl pathway, or a synthetic non-oxidative glycolysis pathway (NOG), or a synthetic reverse glyoxylate shunt (rGS).

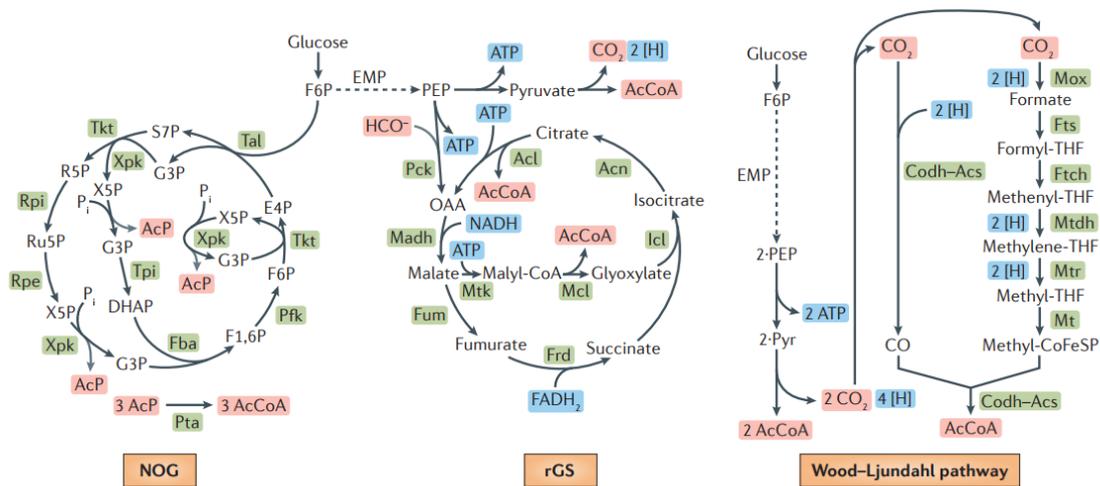


Figure 1.12. Carbon conservation pathways (Liao et al., 2016). AcCoA, acetyl-CoA; Acl, ATP-citrate lyase; Acn, aconitase; Codh–Acs, carbon monoxide dehydrogenase–acetyl-CoA synthase; CoFeSP, corrinoid iron–sulfur protein; DHAP, dihydroxyacetone phosphate; EMP, Embden–Meyerhof–Parnas pathway; F1,6P, fructose-1,6-bisphosphate; Fba, fructose 1,6-bisphosphate (FBP) aldolase; Frd, fumarate reductase; Ftch, formyl-THF cyclohydrolase; Fts, formyl-THF synthase; Fum, fumarase; Icl, isocitrate lyase; Madh, malate dehydrogenase; Mox, methanol oxidase; Mt, methyltransferase; Mtdh, methylene-THF dehydrogenase; Mtr, methylene-THF reductase; Pfk, phosphofructokinase; Pi, inorganic phosphate; Pta, phosphate acetyltransferase; R5P, ribose-5-phosphate; Rpe, R5P epimerase; Rpi, R5P isomerase; Ru5P, ribulose-5-phosphate; S7P, sedoheptulose-7-phosphate; THF, tetrahydrofolate; Tpi, triose phosphate isomerase.

Introduction of the pathways in Figure 1.12 could be used to increase the theoretical yield maximum. The Wood–Ljungdahl pathway is present in acetogens and methanogens. This pathway directly fixes two CO₂ molecules to synthesize acetyl-CoA. However, the pathway involves several complex and oxygen sensitive enzymes. Therefore, engineering this pathway remains a challenge. Alternatively, a non-oxidative glycolysis pathway (NOG pathway) or

a reverse glyoxylate shunt (rGS) can also be used to conserve carbon. The NOG pathway was engineered into *E. coli* and demonstrated the generation of acetyl-CoA through redox-neutral carbon rearrangement from sugars. This increased the theoretical yield of acetyl-CoA from two to three molecules per glucose (Bogorad et al., 2013).

1.9.2 Thermophilic production

Development of thermophilic strains able to produce butanol is not a straightforward process. Almost all microbes that naturally produce *n*-butanol are mesophilic. The enzymes from these strains are unlikely to be functional in a thermophilic host due to thermal instability. In addition, the tolerance of many thermophilic species to butanol is very low. It has been suggested that the toxicity of butanol is due to interactions within components of cell membranes (Vollherbst-Schneck et al., 1984). This effect may be exacerbated at higher temperatures with increase in membrane fluidity.

1.9.3 Aerobic production

Native *n*-butanol production from the genus *Clostridium* is limited by toxicity. Alternative host strains which are more tolerant to *n*-butanol concentrations include aerobic organisms. Development of an aerobic production process could overcome toxicity issues. Use of an aerobic organism would also enable ease of growth and manipulation in an aerobic environment and also prevent the need to maintain a strictly anaerobic fermentation process. An aerobic system could offer biological and practical advantages to biofuel production.

n-butanol, poly(3-hydroxybutyrate), isobutanol and isopropanol have been successfully produced under aerobic conditions by recombinant *E. coli*, requiring NADPH as reducing agent (Kataoka et al., 2015), (Choi et al., 1998; Saika et al., 2014; Valentin and Dennis, 1997; Yang et al., 2014) and (Baez et al., 2011; Hanai et al., 2007) respectively. Kataoka et al. (2015) demonstrated functionality of a CoA-dependent *n*-butanol synthetic pathway under aerobic

conditions in *E. coli*, producing; 4.07 g/l *n*-butanol in shaking flask and 8.60 g/l *n*-butanol in a bioreactor.

n-butanol is a highly reduced chemical, therefore oxygen-limiting conditions have been used to provide sufficient reducing power, in particular NADH, for biological production of *n*-butanol (Lee et al., 2008; Shen et al., 2011). However, the oxidative pentose phosphate pathway used to generate NADPH, a cofactor for *n*-butanol production, was reported to be more active under aerobic conditions (Peng and Shimizu, 2003). Therefore an NADPH-dependent synthetic pathway could allow for aerobic production of reduced compounds. In addition, Kataoka et al. (2015) reported increased *n*-butanol titre and glucose consumption, proportional to an increase in oxygen supply. The activation of glucose catabolism under high oxygen supply conditions could potentially enhance *n*-butanol production.

1.10 Aims of this study

The objectives of this work are first to develop molecular tools to enable metabolic engineering and genetic manipulation of *Geobacillus*. Currently the lack of tools available is limiting the research progress for Gram-positive thermophilic bacteria. Secondly the tools will be utilised in order to introduce a heterologous metabolic pathway to *Geobacillus*. Non-native production of the advanced biofuel and platform chemical *n*-butanol is the aim. Further to this, use of *Geobacillus* as a host strain will be explored. Characterisation of the relatively novel *G. thermoglucosidasius* to better understand, and where possible, improve upon its properties. This will enable evaluation of *Geobacillus* as a potential industrial organism.

CHAPTER 2

MATERIALS AND METHODS

2.1 Bacterial growth conditions

2.1.1 Bacterial strains

Escherichia coli (*E. coli*)

Table 2.1. *E. coli* bacterial strains used in this study.

Strain	Source	Properties
MG1655	Internal research group culture collection	Wild-type laboratory strain of <i>E. coli</i> K-12, cured of F plasmid and phage lambda and with a reduced growth rate in minimal media due to reduced levels of PyrE.
Top10	Invitrogen	Cloning strain used for efficient transformation and blue/white colour screening of recombinant clones. Genotype; F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) Φ 80/ <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ (<i>ara leu</i>) 7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (<i>StrR</i>) <i>endA1</i> <i>nupG</i> .
DH5alpha	ThermoFisher Scientific	Cloning strain with blue/white screening, <i>recA1</i> and <i>endA1</i> mutations for increased insert stability and to improve the quality and yield of plasmid DNA prepared from minipreps.
NEB Turbo	New England Biolabs	Used for cloning of potentially toxic genes. An <i>E. coli</i> K12 strain modified for rapid growth. Genotype; F' <i>proA</i> ⁺ <i>B</i> ⁺ <i>lacI</i> ^q Δ <i>lacZ</i> M15 / <i>fhuA2</i> Δ (<i>lac-proAB</i>) <i>glnV</i> <i>galK16</i> <i>galE15</i> <i>R(zgb-210::Tn10)</i> Tet ^S <i>endA1</i> <i>thi-1</i> Δ (<i>hsdS-mcrB</i>).
BL21 star (DE3)	ThermoFisher Scientific	Used for high-level expression of non-toxic recombinant proteins from low copy number, T7 promoter-based expression systems. High mRNA stability and protein yield. Genotype; F ⁻ <i>ompT</i> <i>hsdS_B</i> (<i>r_B⁻</i> , <i>m_B⁻</i>).
NEB Stable	New England Biolabs	Used for cloning of constructs containing repeat elements; direct repeats and inverted repeat sequences and for cloning unstable inserts. Genotype; F' <i>proA</i> + <i>B</i> + <i>lacI</i> ^q Δ (<i>lacZ</i>)M15 <i>zzf::Tn10</i> (<i>TetR</i>) Δ (<i>ara-leu</i>) 7697 <i>araD139</i> <i>fhuA</i> Δ <i>lacX74</i> <i>galK16</i> <i>galE15</i> <i>e14-</i> Φ 80/ <i>dlacZ</i> Δ M15 <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>nupG</i> <i>rpsL</i> (<i>StrR</i>) <i>rph</i> <i>spoT1</i> Δ (<i>mrr-hsdRMS-mcrBC</i>).

Geobacillus thermoglucosidasius (*G. thermoglucosidasius*)

The wild type strain used throughout this work is *G. thermoglucosidasius* NCIMB 11955. Sheng et al. published the strain sequence in 2016. *Geobacillus*

thermoglucoasidius may be referred to in literature as *G. thermoglucoasidans*, *Parageobacillus thermoglucoasidius* or *Parageobacillus thermoglucoasidans*.

Modified *G. thermoglucoasidius* strains used during this study; Δldh , $\Delta adhE$, *ter* integration + Δldh , $\Delta ldh\Delta pfl$, *ter* integration + $\Delta ldh\Delta pfl$, $\Delta ldh\Delta pfl pdh^{UP}$, *ter* integration + $\Delta ldh\Delta pfl pdh^{UP}$, $\Delta ldh\Delta pfl\Delta adhE pdh^{UP}$, *acc* integration.

***Bacillus subtilis* (*B. subtilis*)**

Bacillus subtilis strain 168. Laboratory strain of *B. subtilis*, sourced from the research group's internal culture collection.

2.1.2 Growth media

Tryptone soy agar (TSA)

Casein digest 15 g/l, Soybean digest 5 g/l, NaCl 5 g/l, No. 1 Bacteriological agar 15 g/l.

2SPYNG

Soy peptone 16 g/l, Yeast extract 10 g/l, Sodium chloride 5 g/l.

2SPY

Soy peptone 16 g/l, Yeast extract 10 g/l, Sodium chloride 5 g/l, Glycerol 10 g/l.
Used for *Geobacillus* transformation recovery.

Luria-Bertani (LB) agar

Tryptone 10 g/l, Yeast extract 5 g/l, NaCl 5g/l, No. 1 Bacteriological agar 10 g/l.

Luria-Bertani (LB) broth

Tryptone 10 g/l, Yeast extract 5 g/l, NaCl 5g/l.

Terrific Broth

Tryptone 12 g/l, Yeast extract 24 g/l, 100 mL potassium phosphate (0.17 M KH_2PO_4 , 0.72 M K_2HPO_4).

AYSE medium

$\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ (10 mM), K_2SO_4 (10 mM), citric acid (2 mM), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (1.25 mM), $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (0.02 mM), $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ (1.65 mM), $(\text{NH}_4)_2\text{SO}_4$ (20 mM), $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ (25 μM), $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (100 μM), $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ (50 μM), $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (5 μM), $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$ (10 μM), $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$ (16.85 μM), and H_3BO_3 (6.5 μM) in deionised water. 1% yeast extract and 1% sugar, as indicated. Filter-sterilised biotin was added to 12.5 μM after autoclaving. Medium adjusted to pH 7.0 and buffered with 2-Amino-2-hydroxymethyl-propane-1,3-diol (Bis-Tris) $\text{C}_4\text{H}_{11}\text{NO}_3$ (40 mM), 1,4-Piperazinediethanesulfonic acid, Piperazine-1,4-bis(2-ethanesulfonic acid), Piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) $\text{C}_8\text{H}_{18}\text{N}_2\text{O}_6\text{S}_2$ (40 mM) and 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) $\text{C}_8\text{H}_{18}\text{N}_2\text{O}_4\text{S}$ (40 mM).

UYSE medium

As ASYE with filter-sterilized urea (20 mM) replacing $(\text{NH}_4)_2\text{SO}_4$.

MOPS Minimal medium

10x MOPS mixture 100 ml/l, 0.132 M K_2HPO_4 10 ml/l, 1 mg/ml thiamine 0.1 ml/l.

AYSE and UYSE with no addition of yeast extract were also used as minimal media.

10X Medium A Base

Yeast extract 10 g/l, Casamino acids 2 g/l, Distilled water to 900 ml, autoclave, then add 50% w/v glucose (filter-sterilised) to 1 l.

10X Bacillus Salts

(NH₄)₂SO₄ 20 g/l, K₂HPO₄ 139 g/l, KH₂PO₄ 60 g/l, Na⁺ citrate 10 g/l, MgSO₄·7H₂O 2 g/l.

Medium A

10 ml 10X Medium A base, 10 ml 10X Bacillus salts, sterile water to 80 ml.

Medium B

10 ml Medium A, 0.1 ml 50 mM CaCl₂·2H₂O, 0.1 ml 250 mM MgCl₂·6H₂O.

Clostridia base medium (CBM) agar

MgSO₄ 7H₂O 200 mg/l, MnSO₄ H₂O 7.58 mg/l, FeSO₄ 7H₂O 10 mg/l, p-aminobenzoic acid 1 mg/l, biotin 2 µg/l, thiamine HCl 1 mg/l, casein hydrolysate 4 g/l, K₂HPO₄ 0.5 g/l, KH₂PO₄ 0.5 g/l, No. 3 bacteriological agar 15 g/l.

CBM1X

CBM, as above, 1% xylose addition following autoclaving, filter sterilised.

CBM1X uracil +

CBM1X, as above, with addition of uracil 20 µg/ml.

CBM1X 5-FOA

CBM1X, as above, with addition of 5-Fluoroorotic Acid (5-FOA) 300 µg/ml. 5-FOA sourced from Europa Bioproducts Ltd.

Soft top agar

Tryptone 10 g/l, NaCl 5 g/l, Technical agar No. 3 1.5% 6.5 g/l.

Electroporation buffer (EPB)

0.5 M sorbitol, 0.5 M mannitol, 10% (v/v) glycerol.

SOC medium

Supplied pre-made from both Invitrogen and New England Biolabs, was used for *E. coli* transformation recovery.

2x glycerol buffer for -80°C storage of bacterial stocks

65% glycerol (v/v), 0.1 M MgSO₄, 0.025 M Tris. Filter sterilised.

Unless otherwise indicated, all media was made up to volume using de-ionized water, pH adjusted to pH 7.0 using 5 M KOH and autoclaved or filter sterilised before use.

Antibiotic supplementation

The above media were supplemented with ampicillin (100 µg/ml for use with *E. coli*), kanamycin (50 µg/ml for use with *E. coli* and 12.5 µg/ml for use with *Geobacillus*) or spectinomycin (50 µg/ml for use with *E. coli* and 12.5 µg/ml for use with *Geobacillus*) where required.

2.1.3 Growth conditions

E. coli

Unless indicated otherwise, agar plates were incubated at 37°C or 30°C. Liquid cultures were incubated at 37°C shaking at 200 rpm. Frozen stocks were kept at -80°C, either as bead stocks according to manufacturer's instructions or in 1ml 1X glycerol stock (2X concentrated glycerol stock: 65% glycerol, 0.1 M MgSO₄, 0.025 M Tris, pH 8.0).

G. thermoglucosidasius

Unless indicated otherwise, agar plates were incubated at 52°C or 60°C in a bag to prevent drying out. Liquid cultures were incubated at 52°C with shaking at 250 rpm. Frozen stocks were stored at -80°C, either as bead stocks according to manufacturer's instructions or in 1 ml 1X glycerol stock (2X concentrated glycerol stock: 65% glycerol, 0.1 M MgSO₄, 0.025 M Tris, pH 8.0).

B. subtilis

Unless indicated otherwise, agar plates were incubated at 37°C. Liquid cultures were incubated at 37°C shaking at 200 rpm. Frozen stocks were kept at -80°C, as bead stocks according to manufacturer's instructions.

Growth of all bacterial cultures in liquid media was monitored by measuring optical density at 600 nm (OD₆₀₀) using a Biomate 3 spectrophotometer (Thermo Scientific).

2.1.4 Cell preparation**Preparation of *E. coli* electro competent cells**

Competent *E. coli* cells were prepared to allow uptake of vectors during transformation using an electroporation protocol. *E. coli* cultures were grown overnight in Luria broth (LB). This pre-culture was used to inoculate fresh LB medium in 100 ml baffled conical flasks and cultured for 3 hours to OD 600nm of 0.6. The culture was then chilled and centrifuged at 4°C for 15 minutes at 5000 x g. The cells were washed and centrifuged again before re-suspension in 10% glycerol. The cells were then centrifuged again and re-suspended in 10% glycerol. Cells were aliquot into 45 µl and frozen at -80°C.

Preparation of *E. coli* chemically competent cells

Competent *E. coli* cells were prepared to allow uptake of vectors during transformation using a heat shock protocol. Top10 *E. coli* cultures were grown overnight in Luria broth (LB). This pre-culture was used to inoculate fresh LB medium in 100 ml baffled conical flasks and cultured until mid-exponential phase O.D 600nm 0.3-0.4. The culture was then chilled on ice for 30 minutes. Culture was centrifuged at 4°C for 10 minutes at 3500 x g. The cell pellet was re-suspended in 10 ml chilled 100 mM CaCl₂·2H₂O and kept on ice for a further 30 minutes. Culture was centrifuged at 4°C for 10 minutes at 3500 x g. The cell pellet was re-suspended in 1.5 ml of chilled 100 mM CaCl₂·2H₂O and kept chilled for 3 hours. A final centrifugation before resuspension in 1 ml of 50%

(v/v) sterile glycerol and water. Cells were aliquot into 100 μ l and frozen at -80°C.

Preparation of *Geobacillus* electro competent cells

Competent *Geobacillus* cells were prepared to allow uptake of vectors during transformation using electroporation protocol. *Geobacillus* strains were grown in pre-warmed 2SPYNG medium in baffled conical flasks (inoculated from a 10 ml overnight culture) until early-exponential phase (OD_{600} nm 2.0) at 52°C with shaking at 250 rpm. The culture was held on ice for 10 minutes. Cells were harvested by centrifugation at 4,500 x g for 20 minutes at 4°C. All subsequent steps were carried out on ice and using chilled equipment and solutions. Cells were re-suspended and washed with filter sterilised electroporation medium (0.5 M sorbitol, 0.5 M mannitol, 10% glycerol) prior to harvesting by centrifugation; this was repeated 3 times with 30 ml, 20 ml, 10 ml & 10 ml ice-cold electroporation buffer, and the final resuspension volume was adjusted to give a final concentration of approximately 40 OD_{600} nm/ml. 60 μ l aliquots of the cell solution were stored at -80°C.

2.2 Laboratory reagents

Unless stated, all chemicals and laboratory reagents were sourced from; Oxoid, Sigma Aldrich or ThermoFisher Scientific.

Fast Digest restriction enzymes were purchased from ThermoFisher Scientific, DreamTaq Green PCR Master Mix was sourced from ThermoFisher Scientific. Other DNA polymerases, Antarctic phosphatase, calf intestinal phosphatase, dNTPs (deoxyribonucleoside triphosphates), DNA markers and associated reaction buffers were sourced from ThermoFisher scientific or New England Biolabs. NEBuilder DNA assembly kits and site-directed mutagenesis kits were sourced from New England Biolabs. Oligonucleotide synthesis was carried out by Sigma-Aldrich. T4 DNA ligase was sourced from Promega. Complete EDTA-

Free Protease Inhibitor Cocktail Tablets were manufactured by Roche (Mannheim, Germany).

2.3 DNA purification

Plasmid DNA extraction was performed using either QIAGEN's QIAprep spin miniprep kit or NEB's Monarch plasmid miniprep kit. Genomic DNA extraction was carried out using the Gen Elute bacterial genomic DNA kit (Sigma). The Zymoclean gel DNA recovery kit from Cambridge Bioscience or NEB's Monarch gel extraction kit were used for isolating DNA from agarose gels. NEB's Monarch reaction cleanup kit was used following PCR. Recovered DNA concentration and purity was measured using a NanoDrop Lite (Thermo Scientific) measuring nucleic acid concentration at 260 nm and purity using the 260/280 nm ratio. DNA was stored at -20°C.

2.4 Bioinformatic tools

DNA database search

Searches for translated nucleotide databases were routinely performed using the tBLASTn algorithm of the Basic Local Alignment Search Tool (BLAST). (<http://www.ncbi.nlm.nih.gov>)

Protein database search

Searches for protein databases were performed using the Blastp algorithm of the Basic Local Alignment tool (BLAST).

(<http://www.ncbi.nlm.nih.gov>)

Metabolic pathway database search

Information regarding metabolic pathways was explored using the Kegg Pathway Database.

(<http://www.genome.jp/kegg/pathway.html>)

Enzyme database search

Information relating to enzymes was obtained using BRENDA.

(<http://www.brenda-enzymes.info/>)

Sequence data handling and plasmid map design

Sequence data were routinely handled using VectorNTI, ApE and SnapGene.

Plasmid maps were designed using ApE and SnapGene. Alignments of two or more than two sequences were performed using either BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) or Align X of VectorNTI.

Oligonucleotide design and synthesis

Oligonucleotides for PCR and sequencing applications were designed either manually or using Primer3 software (<http://primer3.ut.ee>) or Integrated DNA technologies (IDT) software (<https://eu.idtdna.com>).

NEBuilder HiFi DNA assembly cloning method

The NEBuilder online assembly tool (<https://nebuilder.neb.com/>) was used to design the constructs and primers required for DNA assembly.

2.5 Synthetic genes

Genes were synthesised by Biomatik, Eurofins Gene strand and GeneArt by ThermoFisher Scientific. Codon optimisation for *G. thermoglucosidasius* was applied as detailed in appendix A8. Codon optimisation table. Genes were synthesised in a BioBrick BB-2 format, with an RBS, as detailed, and a stop (ochre: TAA) codon.

2.5.1 Tags

The synthetic genes contain either a Flag-tag or a HIS-tag at the C-terminus for detection of expression. A Flag-tag is a polypeptide protein tag with the sequence motif DYKDDDDK. Histidine tagged genes have an addition of six histidine residues fused to the C-terminus of the recombinant protein.

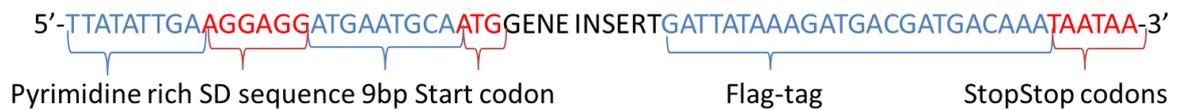


Figure 2.1. The structure of synthetic genes used in this work. This example includes the synthetic RBS and a Flag-tag.

2.6 Cloning strategy

2.6.1 Vectors

Table 2.2. Plasmids used in this work.

Name	Features	Use
pMTL6000 series	Modular vector system containing a Gram-positive replicon, Gram-negative replicon, marker gene and an application specific module.	Modular shuttle vector compatible with <i>E. coli</i> and <i>G. thermoglucosidasius</i> .
pTMO372	Vector containing the reporter gene <i>pheB</i> , from <i>G. stearothermophilus</i> inserted after the <i>idh</i> promoter. Colonies containing the <i>pheB</i> gene turn yellow in contact with catechol.	<i>G. thermoglucosidasius</i> shuttle vector and reporter plasmid.
pJH12	Vector with kanamycin resistance (Kan ^R).	Cloning vector.
pJ201	Vector containing Kan ^R , standard Biobrick prefix and suffix restriction enzyme sites.	Biobrick compatible cloning vector.
pMTL_L5k	Modular vector comprising <i>pyrE</i> from <i>G. kaustophilus</i> and a kanamycin resistance marker.	In-frame deletion using $\Delta pyrE$ <i>G. thermoglucosidasius</i> strain.
ACE_INT2	Vector containing the <i>pyrE</i> gene, lacking a start codon, as a left homology arm (further detailed in section 2.9). Biobrick compatible.	ACE chromosomal integration vector. Used to integrate DNA at the <i>pyrE</i> locus in a $\Delta pyrE$ strain of <i>Geobacillus</i> .
pCR TM 2.1 (Invitrogen)	TA cloning vector with blue-white selection for screening of successful insert.	Invitrogen TA Cloning [®] Kit ligates PCR products containing deoxyadenonine (A) tail to this plasmid vector.
pET17b	Amp ^R , T7 promoter.	<i>E. coli</i> high expression vector for use with BL21(De3) host.

2.6.2 Oligonucleotide primers

The oligonucleotide primers used in this work are listed in appendix A1.

2.6.3 Biobrick cloning

In a manner similar to traditional cloning, the BioBrick method uses restriction endonucleases to generate DNA fragments with specific complementary end sequences that can be joined together with a DNA ligase, prior to transformation. This is a low cost and versatile approach, however it is dependent on the presence of specific restriction sites. Biobrick assembly allows different parts to be assembled using restriction enzymes while retaining the restriction sites so they can be used again for further assembly. The parts are flanked by a standard prefix on the 5' end and standard suffix at the 3' end. The prefix has the configuration of 5'-EcoRI-NotI-SpeI-3' and the suffix 5'-NheI-NotI-PstI-3'. The restriction sites are unique which ensures their sole purpose of addition or removal of components. Since NheI and SpeI have compatible overhangs but their ligation returns a sequence unrecognizable by both, the final product will resume the original standard. However, once the vector has been ligated a six base pair benign protein scar (Ala-Ser) is left.

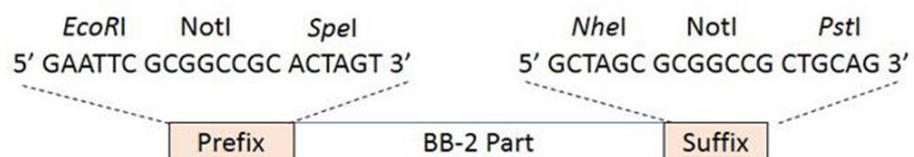


Figure 2.2. The standardised format of a BioBrick part. The DNA fragment has a prefix (EcoRI, NotI, SpeI) and a suffix (NheI, NotI, PstI).

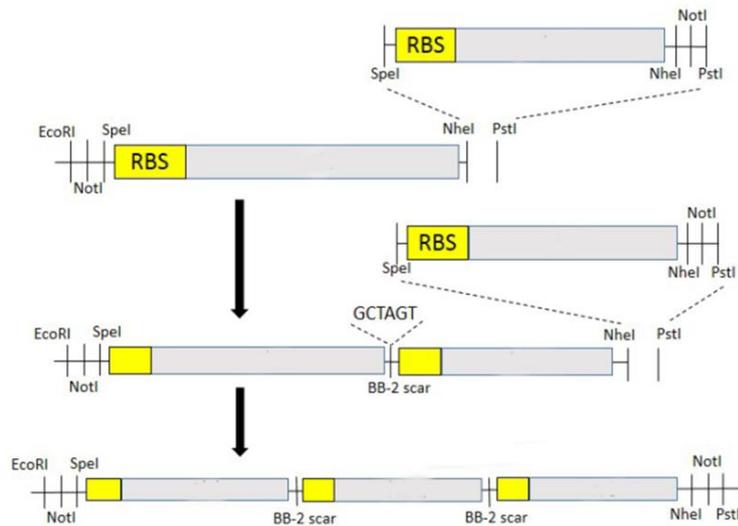


Figure 2.3. Diagram of the BioBrick cloning method. Assembly of multiple parts is possible by utilising the compatible sticky ends of SpeI and NheI. Assembly in this manner allows the construction of operons, leaving a benign Ala-Ser scar between each assembled part.

2.6.4 NEBuilder HiFi DNA assembly

PCR cloning amplifies the vector and DNA fragment/fragments of interest by PCR and ligates the two together without the use of restriction enzymes. Seamless cloning or gene assembly techniques allow scar less insertion of one or more DNA fragments into a vector. Commercial systems include; Gibson Assembly, In-Fusion, GeneArt, Golden Gate and NEBuilder HiFi. These methods are more expensive than traditional cloning however they offer efficient assembly of multiple fragments in an ordered and seamless manner.

Here NEBuilder HiFi DNA Assembly was routinely used for seamless assembly of up to 6 DNA fragments. Constructs and primers were designed using the NEBuilder online assembly tool (<https://nebuilder.neb.com/>). With overlapping regions of 15–80 bp. The vectors were linearized by either restriction enzyme or PCR. The master mix contained; exonuclease to create single-stranded 3' overhangs that facilitate the annealing of fragments that share complementarity at one end (the overlap region), polymerase to fill in

gaps within each annealed fragment, DNA ligase to seal nicks in the assembled DNA. The end result is a double-stranded fully sealed DNA molecule that can be transformed into *E. coli*.

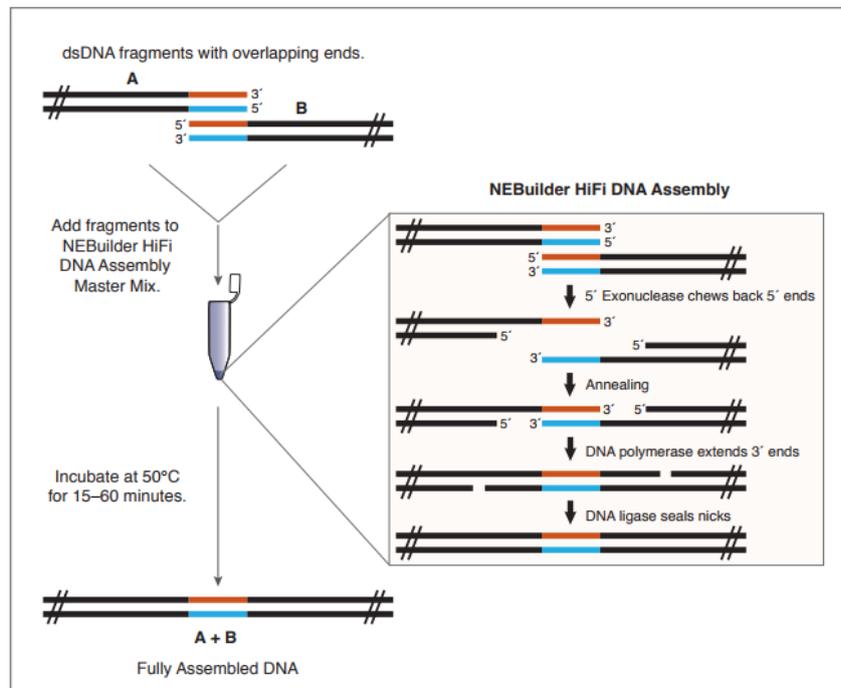


Figure 2.4. Diagram of the NEBuilder HiFi DNA Assembly Method.

Protocol was followed as per the manufacturer's instructions. Briefly; restriction digestion to linearize the vector and PCR amplification of inserts or PCR amplification of all fragments. PCR was carried out with primers designed using the NEBuilder online tool. Additional sequences, such as RBS, terminators or restriction enzymes sites were included in an introduced spacer region between fragments, as required. Q5 high fidelity 2x Master Mix polymerase was used. Following amplification the PCR products and digested vector, if applicable, were run on a gel to confirm amplification and correct fragment size. Fragment were then recovered from the gel and the DNA purified. DNA concentration was measured on a Nanodrop. The fragments were assembled on ice to the following parameters.

Table 2.3. HiFi DNA assembly protocol (NEB). Ratio and concentration of DNA fragments and reagents used during HiFi cloning experiments.

	Recommended Amount of Fragments Used for Assembly		
	2–3 Fragment Assembly*	4–6 Fragment Assembly**	Positive Control †
Recommended DNA Ratio	vector:insert = 1:2	vector:insert = 1:1	
Total Amount of Fragments	0.03–0.2 pmols* X μ l	0.2–0.5 pmols** X μ l	10 μ l
NEBuilder HiFi DNA Assembly Master Mix	10 μ l	10 μ l	10 μ l
Deionized H ₂ O	10-X μ l	10-X μ l	0
Total Volume	20 μ l††	20 μ l††	20 μ l

Samples were incubated in a thermocycler at 50°C for 15 minutes when 2 or 3 fragments were assembled or 60 minutes when 4–6 fragments were assembled. Following incubation the reaction was transformed into DH5alpha competent *E. coli* cells, protocol as described in 2.6.5.

2.6.5 Cloning protocol

Polymerase chain reaction (PCR)

PCR reactions were carried out using either Phusion DNA polymerase (NEB), Q5 DNA polymerase (NEB) or Dream *Taq* polymerase Master Mix (ThermoFisher Scientific). Purified DNA was used as template with 100 μ M primer concentration, appropriate reaction buffer, dNTP and enhancer addition where required. Reaction made up to desired volume using nuclease free water. PCR reactions were carried out in a Professional Trio thermocycler (Biometra). Cycling conditions varied according to the polymerase used, primer T_m and amplicon length targeted. For PCR colony screening a small amount of bacterial colony was used to provide the DNA template. Colonies were picked using a sterile toothpick.

Restriction digest

A typical restriction digest reaction mixture contained; 1 μ l enzyme; 1 μ l plasmid DNA (dependent on concentration); 2 μ l compatible buffer; made up to 20 μ l total volume using nuclease free water. Reaction briefly spun, then

incubated at 37°C for 1 hour. The integration vector was then dephosphorylated to prevent self ligation; 2 µl 10x Antarctic phosphatase reaction buffer and 1 µl Antarctic phosphatase (5 units) or 1 µl calf intestinal phosphatase was added to the digestion and incubated for a further 1 hour at 37°C. Following incubation, digested samples were separated immediately on agarose gels.

Agarose gel electrophoresis

Electrophoresis in 1% (w/v) agarose gel for DNA visualisation. TAE buffer (40 mM Tris (2-Amino-2-hydroxymethyl-propane-1,3-diol) acetate (pH 8.0), 1 mM EDTA) supplemented with SYBR safe. DNA was prepared by dilution with 6x loading dye (50% (v/v) glycerol, 50 mM EDTA, pH 8.0, 0.05% (w/v) bromophenol blue). TAE buffer was used as electrophoresis tank buffer. DNA was run at a constant 100V voltage for 45 to 120 minutes, until the desired band separation was obtained. DNA was visualised using a UV trans-illuminator. Standard 2 log or 1 kb Plus GeneRuler DNA Ladder (Thermo Scientific) was used to estimate the size of DNA fragments.

Ligation

Digested gene inserts were ligated into cut plasmid vectors using the Promega rapid ligation kit, according to the manufacturer's protocol, or using T4 DNA ligase. A range of insert to vector ratios were used. Ligations were carried out either; 37°C for 30 minutes, room temperature for 60 minutes or 4°C overnight.

Transformation

For transformation of *E. coli* cells heat shock was carried out using 100 µl aliquots of chemically-competent *E. coli* cells. Cells were removed from storage at -80°C and left to thaw on ice for 2 minutes. 10 µl of ligation reaction was added and incubated on ice for 30 minutes. The cell aliquots were then incubated at 42°C for 30 seconds, then immediately cooled on ice for 2 minutes. The cells were then re-suspended in 900 µl of SOC medium and

MATERIALS AND METHODS

incubated for a 1 hour recovery period, shaking at 37°C. The cells were then plated onto LB agar plates containing the appropriate selection antibiotic and incubated at either 30°C or 37°C overnight.

Electroporation of *E. coli* was carried out using the following procedure. For plasmids prepared from miniprep, 1-2 µl of 100-500 ng/µl was added to 45 µl electrocompetent cells prior to transformation. For ligation mixtures, salts were removed prior to transformation by pipetting the mixture onto dialysis discs with 0.025 µm pores (Millipore, UK) dialyzing against distilled water for 30 minutes. A 5 µl aliquot of dialysed ligation mix was added to 45 µl electrocompetent cells. Electroporation was carried out using a 2 mm gap electroporation cuvette inside a MicroPulser (BioRad) set to 200 Ω, 2.5 kV and 25 µFD. After electroporation, 800 µl LB medium was added to the cells, mixed, and then transferred to a pre-warmed Falcon tube. Cells were allowed to recover at 37°C, shaking at 200 rpm for 1-3 hours. Cells were then plated on LB agar containing appropriate antibiotics.

Geobacillus transformation was carried out by electroporation of electrocompetent *Geobacillus* cells. 60 µl cell aliquots were removed from storage at -80°C and left to thaw on ice for 5 minutes. 1 mm gap electroporation cuvettes (BioRad) and the plasmid preparation to be transformed were also cooled on ice. 3 µl of plasmid preparation added to the thawed cell aliquots. The cell aliquots were then transferred to chilled cuvettes, placed a micropulser electroporator (BioRad) and subjected to a 5 ms electrical pulse of 2500 voltage, capacitance µF 10, resistance 600. Time constants ranged 5.0-5.3. The cells were immediately re-suspended in 1 ml of pre-warmed (55°C) 2SPY medium and grown with shaking at 250 rpm and 52°C for 4 hours without selection. 200 µl of cells were then plated onto TSA agar plates (pre-warmed at 52°C) containing the appropriate selection antibiotic, and grown at 52°C overnight.

For generation of competent cells and transformation of *B. subtilis*; 10 mL Medium A was inoculated with *B. subtilis* grown overnight on LB agar. Cultured for 90 minutes past stationary phase. 50 µl sub cultured into 0.45 mL pre-warmed Medium B. Allowed to grow for a further 90 minutes at 37°C. 1 µg DNA was added and incubated at 37°C for 30 minutes. Plated on selective medium.

Sanger sequencing

Sanger sequencing of DNA was carried out by Source BioScience (Nottingham, UK).

2.7 Detection of protein expression

2.7.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Overnight cultures adjusted equivalent to OD₆₀₀ 0.05, then centrifuged at 16,000 x g for 5 minutes. The cell pellet was re-suspend in 20 µl 1x SDS sample buffer. Samples and protein marker were boiled for 10 minutes at 100°C. Samples were centrifuged at 16,000 x g for 20 minutes and the supernatant recovered. Samples loaded into NuPAGE® Novex® 4-12% Bis-Tris pre-cast mini gels (Invitrogen) in MES SDS running buffer (50 mM 2-(N-morpholino) ethane sulphonic acid, 50 mM Tris, 0.1% (w/v) SDS, 1 mM EDTA, [pH 7.3]) Gel run at 120V for 2 hours in a Novex XcellIII minicell. Precision Plus Kaleidoscope protein ladder (Bio-Rad) was used to estimate protein size.

To visualize NuPAGE gels after electrophoresis, gels were carefully removed from casing, rinsed with water, then Coomassie blue stain (40% methanol, 10% glacial acetic acid, 0.05% Coomassie brilliant blue R-250) was poured to cover the gel and shaken slowly for 1 hour at room temperature. Gels were then destained using destaining solution (12.5% isopropanol and 10% acetic acid) overnight at room temperature until protein bands were clearly visible.

2.7.2 Western Blot of tagged proteins

Two commonly used epitope tags were used in this work for protein detection by Western blot analysis using tag-specific antibodies. FLAG- and His- tags were added as C-terminal translational fusions. Overnight cultures adjusted equivalent to OD₆₀₀ 0.4, then centrifuged at 16,000 x g for 5 minutes. The cell pellet was re-suspend in 20 µl 1x SDS sample buffer. Samples, protein marker and positive control protein were boiled for 10 minutes at 100°C. Samples were centrifuged at 16,000 x g for 20 minutes and the supernatant recovered providing the soluble cell lysate fraction. Samples loaded into NuPAGE® Novex® 4-12% Bis-Tris pre-cast mini gels (Invitrogen) in MES SDS running buffer (50 mM 2-(N-morpholino) ethane sulphonic acid, 50 mM Tris, 0.1% (w/v) SDS, 1 mM EDTA, [pH 7.3]) Gel run at 120V for 2 hours in a Novex XcellIII minicell. Precision Plus Kaleidoscope protein ladder (Bio-Rad) was used to estimate protein size. Recombinant Flag fusion protein (Alpha Diagnostic International), or a known functional His-tagged protein, was used as a positive control. Wild type cells were included as a negative control.

Following electrophoresis, the proteins were blotted onto a 0.2 µm PVDF membrane using the Trans-Blot® Turbo™ Transfer System (Bio-rad). A mixed molecular weight programme; 1.3A, 25V for 7 minutes was used for the transfer process. The membrane was then blocked with 30 ml Tris-Buffered Saline (TBS) (0.05 M Tris pH 8.0, 0.15 M NaCl) with 1.5% milk powder for 1 hour gently shaking at room temperature before being replaced with 30 ml TBS with 1.5% milk powder with antibody. Either 3 µl FlagTag antibody and anti-flag HRP conjugate (Cell Signalling) or Mouse-anti-Penta-His (Qiagen) and anti-mouse-HRP (Promega) were used for Flag and His tagged genes, respectively. Gel incubated with gently shaking at 4°C overnight for antibody binding. The following day, the membrane was washed with TBS with TWEEN®20 six times for 10 minutes each wash. Then 3,3',5,5'-Tetramethylbenzidine (TMB) (Sigma Aldrich) was applied to the membrane for antibody detection.

2.8 Enzyme Assays

2.8.1 Cell lysis

Geobacillus cells were lysed to release protein using BugBuster Protein Extraction Reagent (Novagen). BugBuster reagent diluted to 1x in 50 mM sodium phosphate buffer with EDTA-free protease inhibitor (Roche) and 20 mg/ml lysozyme. Overnight culture was centrifuged at 4,500 x g for 5 minutes, then re-suspended in BugBuster reagent to OD₆₀₀ 10. The cells were lysed, slowly shaking for 20 minutes at room temperature, then centrifuged at 16,000 x g for 20 minutes at 4°C. The supernatant was recovered and kept chilled on ice.

Protein concentration were measured using Bradford assay from Bio-Rad, following the manufacturer's instructions. BSA was used as the standard.

Wild type cell extract and addition of no substrate were used as negative controls for all assays, unless otherwise indicated.

2.8.2 Reporter gene assays

PheB

PheB is used as a reporter gene in *Geobacillus*. The *pheB* gene product; catechol 2,3-dioxygenase, cleaves the aromatic ring of catechol to form 2-hydroxymuconic-semialdehyde (HMSA). This compound has a strong yellow colour and therefore can be used as an indicator for the expression of the *pheB* gene. Direct qualitative screening of colonies expressing *pheB* was carried out by applying 100 mM catechol directly to the cells and observing a colour change.

A quantitative assay was used to determine the level of *pheB* gene expression by measuring the rate of production of the 2-hydroxymuconic-semialdehyde with a spectrophotometer. HMSA has an absorbance coefficient of 0.33 mMcm⁻¹ at 375 nm. 995 µl of 50 mM sodium phosphate buffer at pH 7.2 with 0.33 mM catechol were pre-warmed to 55° C for 10 minutes in a quartz cuvette.

After zeroing with the same buffer blank, the reaction was initiated by the addition of 5 μ l cell lysate and the increase in absorbance at 375 nm was recorded using a Specord 250 spectrophotometer. Absorbance readings were taken over a one minute period. The same protocol was followed using a plate reader platform. The absorbance of samples in a 96 or 48 well-plate was measured using a BMG LabTech ClarioStar with automated injection of the substrate.

Green fluorescent protein

Here two green fluorescent proteins are used as reporter genes. Superfolder green fluorescent protein (sfGFP) is a thermostable variant of GFP (Pédelacq et al., 2006). Blanchard and co-workers demonstrated the use of sfGFP in *Geobacillus stearothermophilus* (Blanchard et al., 2014). Enhanced consensus green protein variant 123 (eCGP123) is an extremely thermostable GFP (Paul et al., 2011). Expression of the GFP proteins was measured using a BMG LabTech Clariostar instrument. Either whole cell culture or extracted cell lysate was measured, as indicated. Fluorescence was measured using an excitation wavelength of 470 nm to 495 nm, detecting the fluorophore emission at 510 \pm 5 nm. With the fluorescence intensity quantified in Relative Fluorescence Units (RFU).

LicB

A thermostable lichenase protein (β -1,3-1,4-glucan-d-glycosylhydrolase) encoded by the licB gene from *C. thermocellum* is used here as a reporter protein. Cultures are incubated with lichenan (Icelandic Moss) and the lichenase gene activity is detected by Congo Red pH indicator.

Culture centrifuged at 4,500 x g for 5 minutes, 100 μ l supernatant added to 400 μ l of the appropriate sterile culture medium containing 0.2% lichenan. A control experiment using 100 μ l medium only was included. Samples incubated at 50°C for 10 minutes. 100 μ l of 0.2 M NaOH added to stop the reaction. 5 μ l

of 0.1% Congo Red solution added. Samples mixed thoroughly by inverting the tube. 200 μ l of 2M NaCl added. Absorbance of the samples at 530 nm was measured using a ClarioStar plate reader (BMG LabTech). Blank solution contained; 500 μ l medium, 100 μ l 0.2 M NaOH, 5 μ l Congo Red and 200 μ l 2 M NaCl. A positive reaction is determined by a reduction in absorbance at 530 nm, corresponding to a drop in pH. Supernatant is tested as lichenase is excreted by the cells.

2.8.3 *n*-butanol pathway assays

The following assays were performed using a SpectraMax M3 microplate reader (Molecular Devices).

Thiolase (Thl)

Thiolase activity was determined by monitoring the disappearance of acetoacetyl-CoA. The reverse thiolytic cleavage reaction was measured spectrophotometrically by acetoacetyl-CoA decrease at 303 nm as described by Hartmanis and Gatenbeck (1984). 303 nm is the characteristic absorption band of an enolate complex formed by acetoacetyl-CoA with Mg^{2+} . The reaction mixture contained; 100 mM Tris-HCl, pH 8.0, 10 mM $MgSO_4$, 100 μ M acetoacetyl-CoA, 100 μ M CoA and cell extract. The assay was carried out at 52°C in an aerobic environment. The multi well assay plate was pre-warmed to temperature. The reaction was initiated by the addition of the cell extract. Absorbance was recorded every 3 seconds with the plate shaken for 3 seconds before the first read and between each read. The molar absorption coefficient used was 16.9 mM^{-1} (Tanaka et al., 2011).

3-hydroxybutyryl-CoA dehydrogenase (Hbd)

Hbd activity was determined by monitoring the decrease of absorption at 340 nm, corresponding to consumption of NADH. The reaction mixture contained 100 mM 3-(*N*-morpholino) propanesulfonic acid (MOPS), pH 7.0, 200 μ M NADH, 100 μ M acetoacetyl-CoA and cell extract. The assay was carried out at

52°C in an aerobic environment. The multi well assay plate was pre-warmed to temperature. The reaction was initiated by the addition of the cell extract. Absorbance was recorded every 3 seconds with the plate shaken for 3 seconds before the first read and between each read. The molar extinction coefficient used was 6,220 M⁻¹cm⁻¹.

Crotonase (Crt)

Crotonase activity was determined by monitoring the decrease of absorption at 263 nm, corresponding to disruption of the α - β saturation of crotonyl-CoA. The assay mixture contained 100 mM Tris-HCl pH 7.6, 100 μ M crotonyl-CoA and cell extract. The assay was carried out at 52°C in an aerobic environment. As 263 nm is in the UV range a quartz cuvette was used. The quartz cuvette was pre-warmed to temperature. The reaction was initiated by the addition of the cell extract. Absorbance was recorded every 3 seconds. The molar extinction coefficient used was 6,700 M⁻¹cm⁻¹.

Trans-2-enoyl CoA reductase (Ter)

Ter activity was determined by monitoring the forward direction reaction by a decrease in NADH measured at 340 nm with crotonyl-CoA as the substrate. The reaction mixture contained 100 mM potassium phosphate buffer, pH 6.2, 200 μ M NADH, 100 μ M crotonyl-CoA, and crude extract. The assay was carried out at both 52°C and 60°C in an aerobic environment. The multi well assay plate was pre-warmed to temperature. The reactions were initiated by the addition of the cell extract. Absorbance was recorded every 3 seconds with the plate shaken for 3 seconds before the first read and between each read. The molar extinction coefficient used was 6,220 M⁻¹cm⁻¹.

Aldehyde/alcohol dehydrogenase (AdhE2)

AdhE2 activity was determined by monitoring the decrease of absorbance at 340 nm corresponding to the consumption of NADH or NADPH. The reaction mixture contained 100 mM Tris-HCl, pH 7.5, 5 mM dithiothreitol (DTT), 200 μ M

NADH and 0.1 mM butyryl-CoA for the butyraldehyde dehydrogenase reaction (BYDH) and 50 mM butyraldehyde for the butanol dehydrogenase reaction (BDH). The assay was carried out at 52°C in an aerobic environment. The multi well assay plate was pre-warmed to temperature. The reactions were initiated by the addition of the cell extract. Absorbance was recorded every 3 seconds with the plate shaken for 3 seconds before the first read and between each read. The molar extinction coefficient used was 6,220 M⁻¹cm⁻¹.

Thioesterase

Thioesterase activity was measured spectrophotometrically. The reaction mixture contained 60 mM potassium phosphate buffer (pH 7.4), 100 μM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB), 0.08 mg bovine serum albumin, 20 μM butyryl-CoA and crude enzyme from cell extract. Reaction was started with the addition of enzyme. Reduction of DTNB by the CoA liberated in the thioesterase reaction was measured at 412 nm. One unit of enzyme activity was defined as the amount of enzyme catalyzing the cleavage of 1 nmol of acyl-CoA per min. The molar extinction coefficient of reduced DTNB was considered as 13,600 M⁻¹ cm⁻¹.

2.9 Mutant strains

Modified *G. thermoglucosidasius* strains described in Sheng et al. (2017) were used in this work. Strains used; LS001: Δldh , LS003: Δldh , pdh^{up} and LS242: Δldh , pdh^{up} , Δpfl . In addition a $\Delta adhE$ mutant strain was also used here. This mutation was created by Sheng, L (unpublished work).

Modified *G. thermoglucosidasius* strains with heterologous *ter* and *acc* gene insertions into the genome were used here. The *ter* integration was carried out by Sheng, L and the *acc* integration was carried out by Bashir, Z. The integrations were carried out as described by Sheng et al. (2017).

2.9.1 Genomic deletions

In-frame deletion of *spoA* and *sigF* genes was carried out to generate two further mutant strains. Deletions were carried out according to the method described in Sheng et al. (2017). Briefly; the protocol utilised a *pyrE* minus strain, previously created by Sheng et al., (2017). The $\Delta pyrE$ strain provides a counter-selection marker by uracil auxotrophy and resistance to 5-fluoroorotic acid (5-FOA). Knock-out vectors were constructed comprising approximately 500 bp from downstream and upstream of the target gene forming the left homology arm (LHA) and right homology arm (RHA) and transformed into the $\Delta pyrE$ *G. thermoglucosidasius* strain. Homologous recombination of either the LHA or the RHA between the chromosome and in-frame deletion vector results in two types of single crossover integrants; the left cross and right cross, both of which are selected using the plasmid borne antibiotic selection marker by re-streaking on antibiotic selection plates. Further recombination of single crossover integrants, depending on which homology arm, will result either in revertant wild type, or an in-frame deletion mutant, which can be selected using plasmid mediated toxicity of the counter-selectable marker. Two re-streaks on CBM1X without uracil at 60 – 62°C, followed by re-streaking colonies on CBM1X with 20 µg/ml uracil and 300 µg/ml 5-FOA. Double crossover results in gene deletion. Colonies were screened by colony PCR using primers flanking the desired knock-out region. WT genomic DNA used as positive control. Colony PCR using PyrE_C1_F and PyrE_C2_R to ensure a truncated *pyrE* gene is retained. Following creation of the desired mutants, the *pyrE* allele is then restored back to wild type, using ACE and selection for uracil prototrophy.

2.10 Culture conditions

2.10.1 Starter culture preparation

To improve successful and consistent growth of *G. thermoglucosidasius* culture, correct preparation of a starter culture is required. A -80°C stock of the required strain was plated onto TSA +/- antibiotic as appropriate, allowed to grow for approximately 10 hours at 52°C. A single colony was then re- streaked

for whole plate coverage and allowed to grow for approximately 10 hours at 52°C. Approximately half a plate of growth was suspended in 5 ml medium. The concentration of the cell suspension determined and an appropriate volume used to inoculate growth medium to the desired starting inoculum concentration. A starting inoculum concentration of OD 0.1 was commonly used.

2.10.2 Batch culture

For aerobic growth experiments 250 ml baffled flasks were used with 50 ml pre-warmed growth medium +/- antibiotic, carbon source or other supplementation as appropriate. Duran Erlenmeyer flasks with four baffles were used to interrupt the laminar flow to create a turbulent flow. This increases the liquid and gas exchange surface resulting in increased oxygen intake. The flasks were sealed with membrane screw caps to allow continuous gas exchange. The cultures were subject to shaking at 250 rpm. Cultures were generally incubated at either 52°C or 60°C.

For microaerobic growth experiments 50 ml falcon tubes were used with 50 ml pre-warmed growth medium +/- antibiotic, carbon source or other supplementation as appropriate. The caps were sealed using parafilm. This limited the oxygen availability from the headspace and limited the transfer of oxygen into the sealed tube. The cultures were subject to shaking at 250 rpm. Cultures were generally incubated at either 52°C or 60°C.

For anaerobic growth experiments 250 ml serum flasks were used with 50 ml pre-warmed growth medium +/- antibiotic, carbon source or other supplementation as appropriate. The cultures were prepared and inoculated in an anaerobic cabinet. The medium was purged with nitrogen before transfer into the cabinet. Samples were taken by syringe injection. The cultures were subject to shaking at 250 rpm. Cultures were generally incubated at either 52°C or 60°C.

Bioreactor culture was carried out in a 1.5 L vessel. A New Brunswick BioFlo 115 bioreactor was used with a working volume of 1 L. UYSE medium was used with an initial glucose concentration of 20 g/L. The bioreactor was inoculated with a starter culture of exponential phase cells to OD_{600} 0.1. The reactor was maintained at 55°C. A medium pH of 7 was maintained by automatic addition of 3 M NaOH. Airflow rate was kept at 1-2 vessel volume per minute. Agitation 1200 rpm. Any evaporated medium was condensed and returned to the vessel. Antifoam was added during the experiment as required. Samples for OD and solvent analysis were taken with a syringe through a valve port.

2.11 Analytical procedures

2 x 1.5 ml samples were collected during culture. One sample was used for pH measurement, the second was used for metabolite analysis. Culture was centrifuged at 21,000 x g for 5 min at 4°C. If not processed immediately, the supernatant was stored at -20°C until further analysis was performed.

2.11.1 High Performance Liquid Chromatography (HPLC)

If required, the supernatant was thawed before centrifugation at 21,000 x g for 5 minutes at 4°C. The supernatant was filtered through 0.2 µm syringe filters (Whatman, GE Healthcare). Mix 50:50 sample and mobile phase diluent. Mobile phase contained; 0.005 M H_2SO_4 in purified ELGA water with 50 mM valeric acid added for use as an internal standard. 300 µl of the sample was loaded into HPLC vials containing glass inserts and closed with split caps. Standard concentrations of 0 to 150 mM were prepared to quantify results.

Metabolites were analysed using a Dionex Ulti Mate 3000 HPLC System (Thermo Scientific). A Bio-Rad Aminex HPX-87H (BioRad) column, a refractive index and diode array detector at UV 210 nm. A flow rate of 0.5 ml/min was used for 55 minutes. A column temperature of 35°C was used. 0.005 M H_2SO_4 was used as mobile phase. The injection volume was 20 µl. Signal analysis was

performed using the Chromeleon 7.2 Chromatography Data System (Thermo Scientific).

2.11.2 Gas Chromatography (GC)

If required, the supernatant was thawed before centrifugation at 21,000 x g for 5 minutes at 4°C. The supernatant was filtered through 0.2 µm syringe filters (Whatman, GE Healthcare). 5 µl 10 M H₂SO₄ was added to 500 µl of sample supernatant and vortexed shortly. 500 µl of propyl-propionate containing 50 mM valerate (prepared with 549 µl valeric acid in 100 ml of propyl-propionate) was added in a fume hood and vortexed intensely for 10 seconds. The extraction mix was centrifuged at 21,000 x g for 1 minute to remove denatured proteins produced by the pH drop. 300 µl of the top layer loaded into a vial containing a glass insert and sealed with a snap-cap. Standard concentrations of 0 to 150 mM were prepared to quantify results.

Solvent analysis was carried out using a Thermo Scientific Focus GC. The following parameters were used; Column: Thermo Scientific TR-FFAP 30 m x 0.25 mm x 0.25 µm. Carrier: Hydrogen at 0.8 ml/min flow rate. Oven Temp: 50°C (hold for 1 min); ramp to 210°C at 40°C/min; hold at 210°C for 1 minute. Inlet: 240°C with a split ratio of 50. Injection: 1 µl. Detector: FID at 270°C.

2.11.3 Gas Chromatography Mass Spectrometry (GC-MS)

Free fatty acids (FFA) were analyzed by GC-MS; 0.4 ml of culture supernatant was supplemented with 50 µl of 10% (w/v) NaCl and 4 µl of mixture of 10 mg/ml each of C7, C14 and C17 fatty acids as internal standard. The solution was further acidified with 50 µl of glacial acetic acid followed by addition 200 µl of ethyl acetate. The mixture was centrifuged at 16,000 x g for 10 min and esterified with ethanol by adding 900 µl of a 30:1 mixture of ethanol and 37% (v/v) HCl to 100 µl of the organic phase, and incubated at 55°C for 1 hr. The fatty acyl ethyl esters were extracted by adding 500 µl dH₂O and 500 µl hexane and vortexing for 20 seconds. The top hexane layer extract (300 µl) was

analyzed by GC-MS in full scan mode using Agilent 7890 GC with an Agilent 7000 Triple Quadrupole GC-MS system. The column used was Zebron ZB-5MS (length: 30 m; diameter: 0.25 mm; film thickness: 0.25 μm). The injection volume was 1 μl without split mode and oven temperature ramped from 40 to 325°C (initial ramp at the rate of 5°C per min from 40 to 75°C and then 10°C per min until 325°C) and the total run time was 41 min. Compound identities were determined via GC retention times of known standards (FAEES mix from Sigma Aldrich) and verified by mass spectra using NIST library.

2.12 Plasmid segregational stability

Plasmid segregational stability was assessed using a modified method previously described by Pennington (2006). *G. thermoglucosidasius* were transformed with plasmids containing different replicons and selected on TSA plate with antibiotic supplementation. The following day, single colonies were picked and inoculated at 52°C for 16 hours in 10 ml 2SPYNG with antibiotic. 100 μl of culture was used to inoculate 10 ml 2SPYNG again with antibiotics for 12 hours. This marked the start of the stability experiment (0 h). From this point on, cells were inoculated as previously however in a non-selective manner with no antibiotic. This re-inoculation was repeated at 12 hour intervals for 72 hours. After each 12 hours of inoculation without selection pressure, serial dilutions were carried out for each broth from 10^{-1} to 10^{-7} in fresh 2SPYNG pre-warmed for 30 minutes. A 100 μl aliquot of each dilution was plated out on non-selective TSA plate pre-dried for 1 hour at 37°C. On the following day, one hundred single colonies from the TSA plates were replica plated using a 1 μl sterile inoculation loop on TSA plates with and without antibiotic. After 24 hours, all colonies were counted to enumerate total CFU and antibiotic resistant CFU. The percentage plasmid loss was calculated using the difference between number of non-resistant and resistant colonies. Plasmid retained per generation was calculated with the equation \sqrt{Rn} and plasmid lost per generation as $1-\sqrt{Rn}$ where n is the number of generation and R is the

percentage of cell population retaining the plasmid. The experiment was carried out incubating the cultures at 52°C or 60°C throughout.

2.13 Plasmid copy number

Plasmid copy number per chromosome estimates were determined by real time quantitative PCR (qPCR). Samples were taken after 24 hours culture in 2SPYNG (250 rpm) at 52°C to determine copy number at stationary phase, and immediately subject to total DNA extraction. Total DNA was extracted using the Gen Elute bacterial genomic DNA kit, following the manufacturer's instruction. DNA concentration and purity was measured using a NanoDrop Lite (Thermo Scientific).

qPCR amplification using two primer sets targeting unique plasmid and genomic DNA specifically for quantitation of plasmid copy number was carried out. A 120 bp region in the *lepA1*, elongation factor 4 gene, was used as the chromosomal target for qPCR amplification. PCR primers were designed using IDTDNA online software (<https://eu.idtdna.com>). The qPCR primers were designed between 20 and 24 bases in length and took into consideration T_m value and specificity. Primers free of strong secondary structures and self-complementarity were used. Melting temperatures, equal and close to the optimum 62°C were used. qPCR amplicons of <200bp in length were designed for each replicon.

LuminoCt® SYBR® Green qPCR ReadyMix™ (Sigma) was used according to the manufactures instruction. Briefly; reactions were prepared on ice. 20 µl reactions containing 10 µl 1x ready mix, 0.2 µM each forward and reverse primer, 4 µl DNA, to volume with PCR grade water, were loaded into a white 96 well PCR plate. Each reaction was run in triplicate. The plate was sealed with a ThermalSeal film (Sigma). A Lightcycler 480 II instrument was used for amplification and detection. Cycling parameters; Initial denaturation 94°C for 2 min followed by 40 cycles of denaturation 94°C for 15 sec and annealing,

extension and fluorescence read 57°C for 1 min. Cycle threshold (Ct) values were calculated automatically by the Lightcycler 480 software. Relative quantification based on the genomic amplification was used to determine copy number per chromosomal copy ratio, assuming the chromosomal value to be one.

2.14 *n*-butanol toxicity

2.14.1 Determining tolerance

The growth characteristics of *G. thermoglucosidasius* were observed using the methods as described in 2.10. The cell density was measured by spectrophotometer at OD₆₀₀ over the course of the experiment to obtain a growth curve. To determine the toxicity of *n*-butanol to *G. thermoglucosidasius* 2SPYNG medium was supplemented with *n*-butanol at a range of concentrations. The minimum inhibitory concentration (MIC) was defined as the highest concentration in which *G. thermoglucosidasius* culture can grow without detrimental effect, in comparison to unexposed culture.

2.14.2 Directed evolution

Directed evolution of *G. thermoglucosidasius* culture was carried out by continual sub-culture with exposure to sub-lethal levels of *n*-butanol, increasing in concentration over time. 10 ml of pre-warmed 2SPYNG medium was inoculated with 50 µl of *G. thermoglucosidasius* culture and allowed to grow overnight at 52°C with shaking at 250 rpm. The following day the culture was used to inoculate 5ml fresh 2SPYNG. This sub-culture process was repeated daily. Multiple experiments were carried out simultaneously. One set of sub-cultured strains was not exposed to *n*-butanol. To the other sub-cultures *n*-butanol was added to the growth medium. To the initial culture 1% (v/v) *n*-butanol was added. This was increased by 0.1% (v/v) daily. Samples of the final culture were stored at -80°C.

2.15 Whole genome sequencing

Whole genome sequencing was carried out by MicrobesNG (<http://www.microbesng.uk>), at the University of Birmingham which is supported by the BBSRC (grant number BB/L024209/1). High throughput Illumina MiSeq using 2x250 bp paired-end reads.

2.15.1 Sequencing data analysis

Read mapping

Paired-end reads were mapped against the published *G. thermoglucosidasius* NCIMB 11955 genome (Sheng et al., 2016) in CLC Genomics Workbench 8.5.1 (Qiagen, DK).

The following parameters were used for mapping:

- masking mode = no masking
- mismatch cost = 2
- cost of insertions and deletions = linear gap cost
- insertion cost = 3
- deletion cost = 3
- insertion open cost = 6
- insertion extend cost = 1
- deletion open cost = 6
- deletion extend cost = 1
- length fraction = 0.5
- similarity fraction = 0.8
- global alignment = No
- auto-detect paired distances = yes
- non-specific match handling = map randomly

Variant analysis

CLC Genomics Workbench 8.5.1 (Qiagen, DK) was used to detect variations in the sequencing data when mapped against the published *G.*

thermoglucoasidarius NCIMB 11955 genome (Sheng et al., 2016) as the reference strain. Single nucleotide variants (SNV), multi-nucleotide variants (MNV), small to medium sized insertions, deletions or replacements, as well as large structural variants, were detected.

The following parameters were used for variant detection:

ploidy = 1
ignore positions with coverage above = 100,000
restrict calling to target regions = not set
ignore broken pairs = yes
ignore non-specific matches = reads
minimum coverage = 10
minimum count = 2
minimum frequency (%) = 40.0
base quality filter = yes
neighbourhood radius = 5
minimum central quality = 20
minimum neighbourhood quality = 15
read direction filter = no
relative read direction filter = yes
significance (%) = 1.0
remove pyro-error variants = yes
in homopolymer regions with minimum length = 5
with frequency below = 0.7

2.16 Microscopy

2.16.1 Light microscopy

Light microscopy was used for routine observation of cells and also to determine if sporulation had occurred. Cultures were viewed using an Eclipse Ci microscope (Nikon). Images were captured with Nikon's digital sight DS-U3.

Endospore staining was carried out to differentiate bacterial spores from vegetative cells. Schaeffer-Fulton staining as described by Shaeffer and Fulton (1933) is a differential stain which selectively stains bacterial endospores. A droplet of cell culture was added to a microscopy slide and allowed to air-dry. The sample was heat-fixed above a gas flame and covered with a piece of absorbance paper, after which the slide was flooded with 50 g/l malachite green (4-[(4-dimethylaminophenyl)phenyl-methyl]-N,N-dimethylaniline) and heated to steam twice. The absorbance paper was removed and the slide was washed with tap water, after which it was flooded with 25 g/l safranin in 95% ethanol for 30 seconds, washed with tap water, dried with paper and examined under light microscope. Endospores appear bright green, whilst vegetative cells appear pink.

2.16.2 Confocal microscopy

Confocal fluorescence microscope imaging was utilised to observe the structure of *G. thermoglucosidasius* biofilm. The biofilms were stained with live/dead staining using propidium iodide (PI) and SYTO9 nucleic acid binding dyes. PI is a red intercalating stain that is membrane impermeant and is therefore excluded by live cells. SYTO9 is a green intercalating membrane permeant stain and will stain all cells, provided they contain nucleic acid. The dyes were used to show the presence of viable cells within the biofilm structure. Live cells, being exclusive of PI, will stain SYTO9 positive and fluoresce in green. PI has a stronger affinity for nucleic acid and so when the two stains are present within a cell, SYTO9 will be displaced from nucleic acid and the cells will fluoresce in red.

Fluorescence microscopy was performed using a Zeiss LSM 700 Confocal Laser Scanning Microscope. PI was detected using an excitation/emission maxima of 493/636 nm. Syto9 was detected using an excitation/emission maxima of 485/498 nm (for DNA detection). Z-stacks were obtained to show 3D surface

topography of the sample. Measurement result analysis and image assembly were carried out using the LSM 700 ZEN software.

2.17 Prophage Induction

Prophage are bacteriophage genomes integrated into a host chromosome. Bacteria carrying a prophage are termed lysogens. Lysogeny is widespread in bacteria, contributing to bacterial diversity and evolution. Here isolation of prophage via induction of potential *Geobacillus* lysogens was attempted. Prophage induction to enter the lytic cycle occurs in response to either favourable growth conditions or DNA damage. The prophage is excised from the bacterial genome and the lytic pathway resumes.

In order to test for induction during favourable growth conditions; *Geobacillus* wild type culture was grown to exponential phase in liquid 2SPYNG culture at 52°C. 5 ml culture was centrifuged at 3,000 x g for 12 minutes at 4°C. The supernatant was recovered and neutralised to pH 7 with 0.1 M NaOH. The supernatant was filtered through a membrane filter of pore size 0.45 µm. The supernatant was stored at 4°C. Petri dishes were prepared with 15 ml TSA bottom agar. 500 µl overnight wild type culture was mixed with 3 ml soft top agar, poured over the set bottom agar and swirled to cover the plate. 500 µl of the sterile supernatant was spread over the solidified top agar. The supernatant was allowed to absorb into the agar before overnight incubation at 52°C. The following day the agar plates were inspected for zones of clearing (plaques).

In order to test for induction caused by DNA damage; exposure to UV light, heat stress and aged cultures were used. The previous protocol was modified to treat the cell culture prior to obtaining the supernatant. For bacteriophage induction by UV light; 1 ml aliquots of exponential phase cells were centrifuged at 16,000 x g for 2 minutes. The cell pellet was re-suspended in 500 µl PBS. The cells were then exposed to UV radiation for 20 minutes using the CX-2000

(UVP). For heat stress cultures were incubated at 100°C for 20 minutes. For aged culture the cells were incubated at 52°C for up to 5 days prior to use. The supernatant of the treated cultures was plated onto agar plates prepared as described previously. The plates were incubated at 52°C overnight and inspected for plaques.

CHAPTER 3

DEVELOPMENT OF MOLECULAR
TOOLS

3.1 Introduction

Thermophilic microorganisms have several operational advantages as production hosts over mesophilic organisms, such as; lower cooling costs, reduced contamination risk and a process temperature matching that of commercial hydrolytic enzymes, enabling simultaneous saccharification and fermentation at higher efficiencies and with less enzyme. However, exploitation of the potential of thermophilic organisms will not be achieved until a range of effective gene tools are developed to allow the engineering of thermophilic strains. *Geobacillus* is an attractive chassis organism for high temperature biotechnological application. Its implementation in industrial use is currently limited by a lack of available genetic tools to effectively modify the host. The development of genetic tools for thermophilic organisms is crucial to fully understand their metabolic versatility and to establish a thermophilic production platform for biofuel production.

To date the genetic tools which have been developed for *G. thermoglucosidasius* application include; Compatible expression vectors (Reeve et al., 2016), transformation by electroporation protocol (Cripps et al., 2009), identification of a strong promoter (Lin et al., 2014), and gene knock-out and knock-in techniques (Sheng et al., 2017; Bacon et al., 2017).

In this work further tools for the metabolic engineering of *G. thermoglucosidasius* are developed, with the aim of enabling implementation of desired modifications to a host, resulting in an *n*-butanol production strain.

3.2 Reporter systems

The manipulation of gene expression in a microbial host is an essential requirement for all metabolic engineering experimentation. In order to determine expression levels, a selection of functional and preferably quantifiable reporter systems are needed. For use with *Geobacillus*, thermophilic reporter genes encoding stable products are required. In this

work the reporter systems described are used for the measurement of promoter strength, RBS strength, to determine the application of expression systems and to determine vector presence and gene expression in host cells.

3.2.1 PheB

The *pheB* gene from *G. stearothermophilus* has been exploited as a reporter gene for *Geobacillus* spp. (Bartosiak-Jentys et al., 2012). The gene product, catechol 2,3-dioxygenase (C23O), catalyses the conversion of catechol to 2-hydroxymuconic semi aldehyde (HMSA) which is vivid yellow in colour. This can be used as a qualitative and a quantitative reporter system.

When exposed to catechol, cells containing the *pheB* gene produce C23O to catalyse the dioxygenolytic metacleavage of the catechol aromatic ring to yield HMSA, which is bright yellow in colour. A rapid change in colour provides a distinctive visual result enabling a quick and easy qualitative result.

Here the *pheB* gene from *G. stearothermophilus* was inserted into the pTMO372 *Geobacillus* shuttle vector using the BioBrick cloning method. Various promoters were cloned into the vector, also via the Biobrick cloning method, to drive expression of the reporter gene. The pTMO372+*pheB* vector was then transformed into *G. thermoglucosidasius*. Transformed colonies were exposed to catechol, on plate, for qualitative visualisation. In order to determine the application and temperature range of the PhB reporter system for quantitative measurement, the rate of production of 2-hydroxymuconic-semialdehyde was measured with a spectrophotometer. HMSA has an absorbance coefficient of 0.33 mMcm⁻¹ at 375 nm. Change in absorbance is indicative of the gene expression level. Transformed cells were cultured across a temperature range 50 to 70°C.

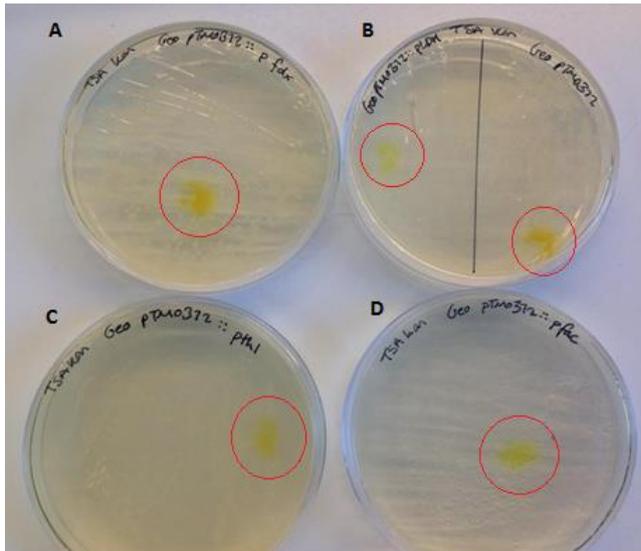


Figure 3.1. Qualitative visualisation of *pheB* positive cultures. *G. thermoglucosidasius* was transformed with the plasmid pTMO372 containing the *pheB* gene driven by the following promoters; A. *fdx*, B. *ldh* and *idh*, C. *thl* and D. *fac*. Single colonies of positive transformants were re-streaked, resulting in pure cultures containing the *pheB* gene. Following overnight incubation at 52°C single areas of each plate, highlighted here in a red circle, were exposed to 100 mM catechol. The cultures can be identified as expressing catechol 2,3-dioxygenase by the yellow colouration.

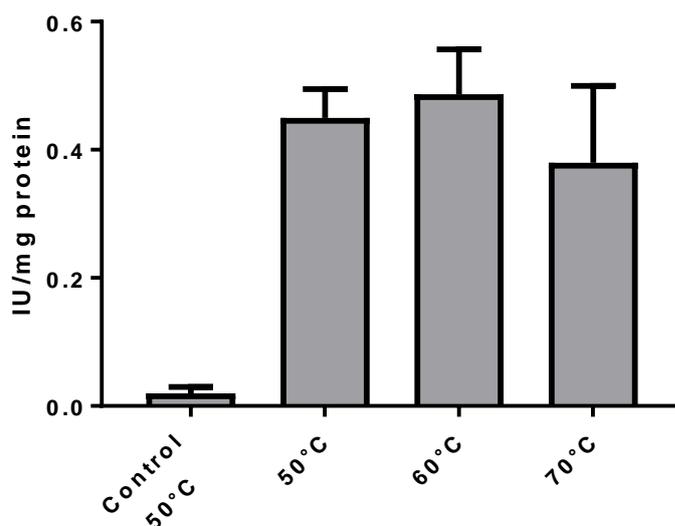


Figure 3.2. Functional quantitative use of the PheB reporter system across a temperature range. *G. thermoglucosidasius* cultures transformed with a vector containing the *pheB* gene, driven by *idh* promoter, were grown overnight at the temperature indicated. The control contained an empty vector with no *pheB* gene present and was incubated at 50°C. Extracted cell lysates were assayed, at the corresponding incubation temperature. Result determined by International Unit (IU)/mg protein value. IU measured as the average slope of absorbance at 375 nm over a one minute period after exposure to catechol. Results adjusted to standardise protein concentrations, with protein concentration measured by Bradford assay. Samples were tested in triplicate. Analysis of variance (ANOVA); f-ratio value: 1.24628, p-value: 0.352644, therefore the result is not significant at $p < .05$.

Although this reporter system has previously been described, its use as a direct qualitative assay on intact colonies has not been described previously. In addition the functionality of the system across a temperature range has not previously been confirmed. This demonstrates the *pheB* reporter gene system is functional for application in *G. thermoglucosidasius* in the temperature range 50°C to 70°C. No significant difference was observed across the temperature

range tested, indicating thermophilic temperatures up to 70°C will not influence the assay performance.

3.2.2 Green fluorescent proteins

Green fluorescent proteins (GFP) are commonly used reporters in mesophiles. Here two green fluorescent proteins have been identified and demonstrated applicable for use in *G. thermoglucosidasius*. Results presented subsequently in this chapter.

Superfolder GFP (sGFP) is a thermostable variant of the green fluorescent protein originating from the jellyfish *Aequorea victoria*. sGFP folds efficiently when fused to poorly folded polypeptides (Pedelacq et al., 2006). sGFP exhibits an improved tolerance to circular permutation, greater resistance to chemical denaturants, and improved folding kinetics relative other GFP variants (Pedelacq et al., 2006). Cava et al. (2008) demonstrated sGFP to be functional *in vivo* at 70°C in the extreme thermophile *Thermus thermophilus*. In this work the sGFP gene was codon optimised for *G. thermoglucosidasius*. The optimised protein is functional in *Geobacillus*, showing strong brightness and thermostability.

Enhanced consensus green protein variant 123 (eCGP123) is an extremely thermostable GFP (Paul et al., 2011). eCGP123 is a synthetic construct, derived from consensus engineering and a recursive evolutionary process. eCGP123 exhibits negative reversible photoswitching properties. This protein has not previously been described as a reporter, or for any other application. Its use as a molecular tool is novel.

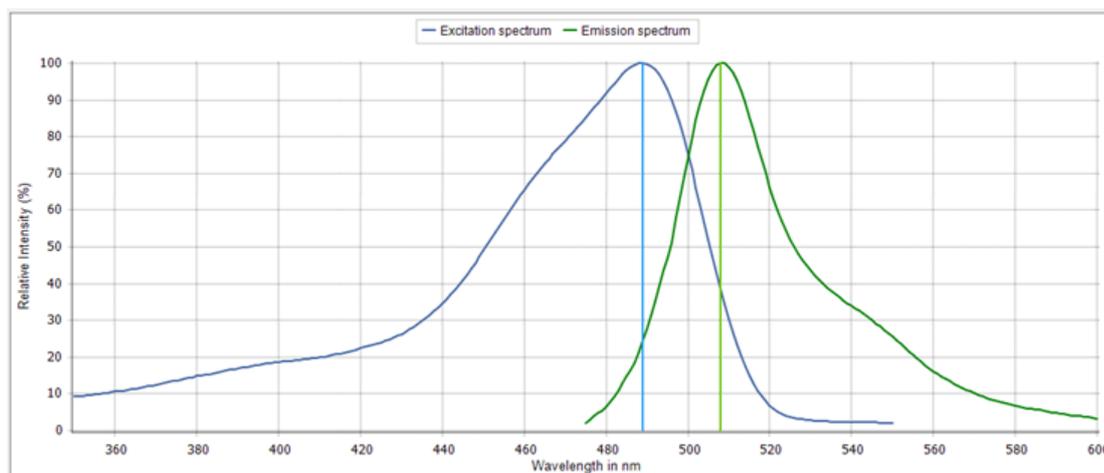


Figure 3.3. Excitation and emission spectra of sGFP and eCGP123. *G. thermoglucosidasius* cultures expressing plasmid borne sGFP and eCGP123. Main excitation peak at 489nm and emission maximum at 510 +/-5 nm. Measured on BMG LabTech Clariostar instrument. It was noted that sGFP emission peak was at the top end of the range, whereas the eCGP123 emission peaked at the bottom of the range.

3.2.3 LicB

A thermostable lichenase protein (β -1,3-1,4-glucan-d-glycosylhydrolase) encoded by the *licB* gene from *Clostridium thermocellum* is used here as a reporter protein, as previously described by Schimming et al. (1992). Cultures are incubated with lichenan and the lichenase gene activity is detected by Congo Red pH indicator.

The enzyme is reported to have high resistance to heat (temperature optimum around 80 °C), ethanol, ionic detergents and is active over a broad pH range (pH 5–12) (Berdichevets et al., 2010). These properties make LicB a suitable candidate for use with *G. thermoglucosidasius*. Results presented subsequently in this chapter (Figure 3.20) demonstrate successful application in *G. thermoglucosidasius*.

3.3 Promoters

Promoters are crucial control elements for *in vivo* gene expression. Promoters play a key role in determining the level of gene transcription. In order to manipulate the expression of genes in a microbial host, promoters with characterised expression profiles are required. For the optimization of metabolic pathways, the fine-tuning of gene expression is likely necessary. The development of a promoter library is therefore desirable, for selection of an appropriate promoter. Here the pheB assay is used as a quantifiable reporter gene system to screen promoter activity.

To provide a set of characterised promoters for *G. thermoglucosidasius* engineering, a range of constitutive promoters from *Geobacillus* sp., in addition to known strong promoters from *Clostridia*, were evaluated. The promoters were used to drive the expression of the pheB gene, using the reporter assay as described previously. The promoter strength was measured in both *G. thermoglucosidasius* and *E. coli*. The vector used contained a strong transcriptional terminator, T1T2, to ensure any expression was not the result of overrun from a previous promoter. The synthetic RBS was used for all promoters to ensure consistency in sequence and spacing to allow comparison between the promoters.

The gene source of the promoters tested:

- 1) P_{ldh} *G. stearothermophilus* lactate dehydrogenase promoter
- 2) P_{idh} *G. thermoglucosidasius* isocitrate dehydrogenase promoter
- 3) P_{ala} *G. thermoglucosidasius* alanine dehydrogenase promoter
- 4) P_{suc} *G. thermoglucosidasius* succinate dehydrogenase promoter
- 5) P_{thl} *Clostridium acetobutylicum* thiolase promoter
- 6) P_{fdx} *Clostridium sporogenes* ferredoxin promoter
- 7) P_{fac} A synthetic construct of *Clostridium pasteurianum*'s ferredoxin promoter with a lac operator

Promoter sequences can be found in Appendix A3.

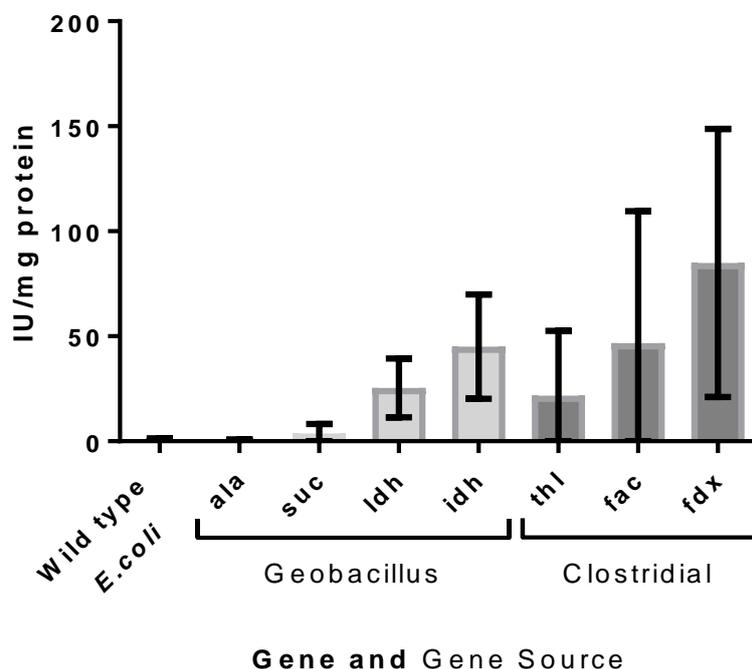


Figure 3.4. Promoter strength in *E. coli* as measured by *pheB* expression. Measurement of constitutive promoter activity in *E. coli* was carried out using the PheB reporter assay. Following overnight incubations, cells were lysed and supernatant used in the assay performed at 37°C. Result determined by International Unit (IU)/mg protein value. IU measured as the average slope of absorbance at 375 nm over a one minute period after exposure to catechol. Results adjusted to standardise protein concentrations, with protein concentration measured by Bradford assay. All measurements performed in biological and technical triplicate, with the average and standard error of nine values shown. Samples tested; Wild type *E. coli* containing no *pheB* gene; *E. coli* cultures containing pTMO372 plasmid with the *pheB* gene under the transcriptional control of the promoters as labelled.

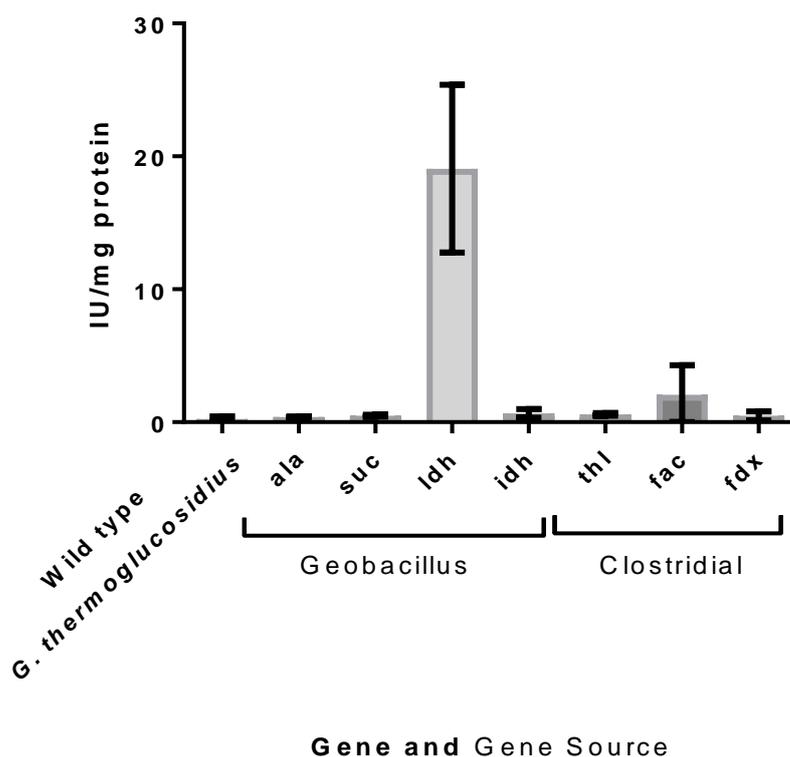


Figure 3.5. Promoter strength in *G. thermoglucosidasius* as measured by *pheB* expression. Measurement of constitutive promoter activity in *G. thermoglucosidasius* was carried out using the PheB reporter assay. Following overnight incubations, cells were lysed and supernatant used in the assay performed at 55°C. Result determined by International Unit (IU)/mg protein value. IU measured as the average slope of absorbance at 375 nm over a one minute period after exposure to catechol. Results adjusted to standardise protein concentrations, with protein concentration measured by Bradford assay. All measurements performed in biological and technical triplicate, with the average and standard error of nine values shown. Samples tested; Wild type *G. thermoglucosidasius* containing no *pheB* gene; *G. thermoglucosidasius* cultures containing pTMO372 plasmid with the *pheB* gene under transcriptional control of the promoters as labelled.

The levels of expression are much higher in *E. coli* than *Geobacillus*, across all of the promoters tested. None of the promoters originating from clostridia

bacteria; P_{thl} , P_{fdx} and P_{fac} , showed strong activity in *Geobacillus* (Figure 3.5). Of the promoters originating from *Geobacillus* species, P_{ldh} is by far the strongest. With expression from P_{ldh} being 28.5 times greater than that from P_{idh} . However P_{idh} is not the strongest promoter in *E. coli*, with P_{idh} expression 1.8 times greater. P_{ldh} will be selected as the promoter of choice for strong gene expression in *Geobacillus*.

The differing expression levels identified in *Geobacillus* and *E. coli* are a desirable feature which can be utilised. A promoter active in *Geobacillus* but not in *E. coli* would enable ease of construction in *E. coli* before transferring a vector into *Geobacillus* for integration and expression. It would also be advantageous for working with toxic genes. As low *E. coli* expression will reduce stress caused to the host and therefore would likely improve efficiency of the cloning and propagation stages in *E. coli*.

Following this initial screening, further *Geobacillus* gene promoters were screened (work carried out by Lau, M. and not presented here). Additional strong promoters were identified in this work, thus expanding the library of characterized promoters further. One additional promoter of note is that of the gene encoding the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPD). P_{GAPD} was found to be a strong promoter in the subsequent screening of *Geobacillus* promoters and is used in this work.

3.4 RBS

Following promoter driven gene transcription, protein translation is initiated by recruitment of a ribosome to the ribosome binding site (RBS) of the mRNA transcript. Expression levels can be influenced at this stage. Protein synthesis commences with translation initiation. Initiation depends on a multitude of factors including; the hybridization of 16S rRNA to the RBS; the binding of the tRNA^{fMet} to the start codon; the spacing between RBS and the start codon; and

the presence of RNA secondary structures that occlude the 16S rRNA binding site (Taton et al., 2014).

The 3' end of 16S rRNA contains the anti-Shine-Dalgarno sequence, which binds the Shine-Dalgarno (SD) sequence through base pair complementation. The *Geobacillus* consensus SD sequence is AGGAGG, as is many other bacteria including *E. coli* (Shine & Dalgarno, 1975). Changes in the RBS sequence can weaken or strengthen the 16s rRNA affinity to the RBS, allowing for regulation.

A synthetic RBS for use in *Geobacillus* was previously developed (Sheng, 2014). The synthetic RBS sequence was based on the RBS sequence from *G. stearothermophilus* lactate dehydrogenase (*ldh*) gene. One nucleotide substitution was made to modify the SD sequence. The resulting RBS sequence was predicted to provide an effective translation initiation signal.

Here the synthetic RBS sequence, in addition to further *Geobacillus* RBS sequences, were measured. Screening was carried out in order to characterise translation efficiencies and to increase the number of RBS sequences available for use when engineering *Geobacillus*. The RBS sequences tested are *Geobacillus* native, all originating from genes encoding enzymes of the glycolysis pathway, apart from the gene coding for isocitrate dehydrogenase which is found in the citric acid cycle. The latter promoter was included in the screen as it has previously been used in *Geobacillus* metabolic engineering efforts. The RBS sequences from glycolysis pathway genes were chosen as their encoded enzymes are predicted to be present at high levels in the cell due to their native function. Their genes are therefore likely to possess both effective transcriptional and translational signals.

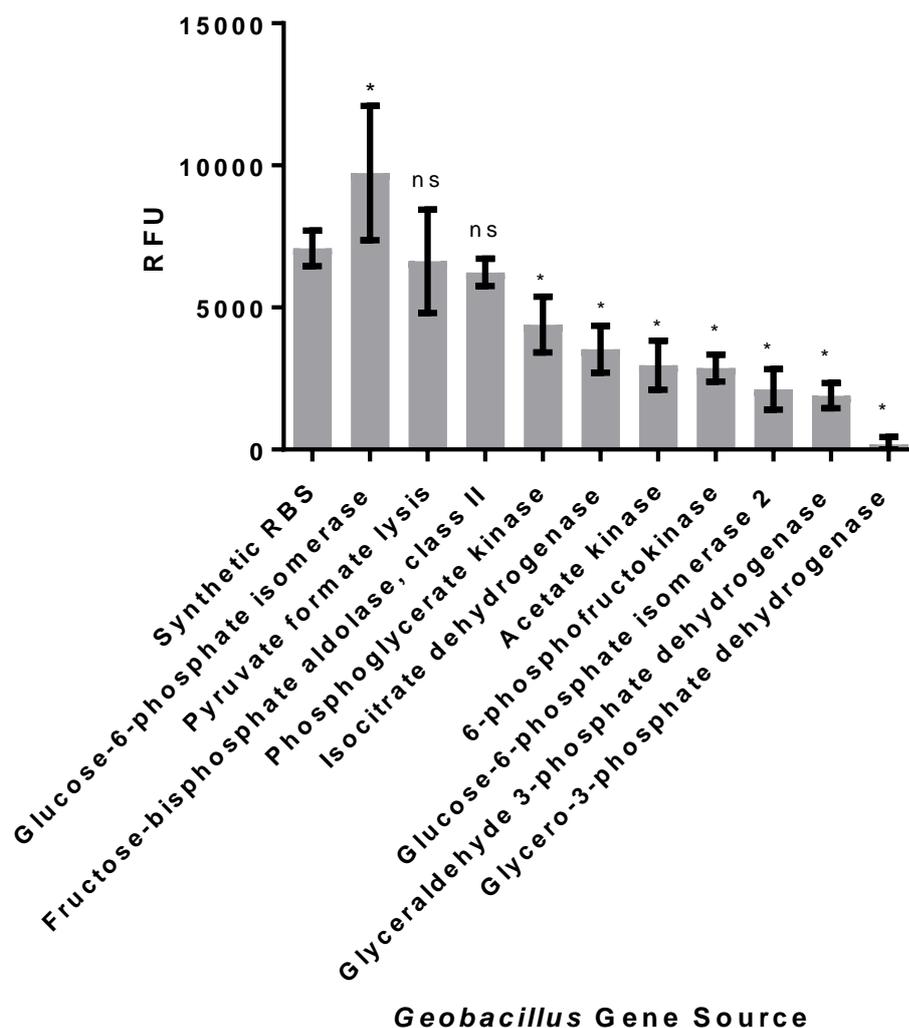


Figure 3.6. RBS screening using sGFP assay. Eleven RBS sequences, from the *Geobacillus* genes encoding the indicated enzymes, were cloned into a standardised expression vector; pMTL61224, including a strong T1T2 terminator and P_{ldh} . The RBS sequence design included 9 bp prior to the shine-delgarno sequence ending with the ATG start codon. The sGFP assay was carried out as previously described. All GFP fluorescence readings were corrected for medium autofluorescence and divided by corrected OD_{600} readings to give an estimate of GFP per cell. The data was analysed by two tailed t test to determine significance. The result was deemed significant when $p < 0.05$. *Signifies a significant difference when compared to the synthetic RBS; ns signifies no significant difference in comparison to the synthetic RBS sequence; RFU relative fluorescent units.

Only one of the RBS sequences screened was found to be significantly stronger than the synthetic RBS construct designed. This confirmed the synthetic RBS is, as predicted, a suitable strong initiator of translation for use in engineering work. In addition, two further RBS sequences were found to be similar in comparison to the synthetic RBS. The sGFP expression levels from the remaining seven RBS screened were significantly lower. This screening has provided a range of RBS with differing strengths which can be selected from, as appropriate, for application in *Geobacillus* work.

This screening experiment has shown the RBS sequence affects gene expression, resulting in a range of sGFP expression levels. The spacing between the SD sequence and the translation initiation codon is one factor influencing initiation. Numerous studies have suggested that a distance of 7-10 nucleotides between the SD site and the start codon is optimal for translation initiation in *E. coli* (Curry & Tomich, 1988; Chen et al., 1994). Previous analysis of the *G. thermoglucosidasius* C56-YS93 genome, found 9 bp to be the most common spacing in essential pyruvate metabolism genes (Sheng, 2014).

Table 3.1. Sequences of the screened RBS. The RBS included in the screening are listed here. The order of relative strength denotes the sGFP expression measured. The Shine-Delgarno sequences and the ATG start codon are highlighted in blue and green, respectively. The distance from the end of the SD sequence to the start of the start codon is listed.

Order of relative strength	Gene source	Sequence	Distance from SD sequence to start codon (bp)
1	Glucose-6-phosphate isomerase	GGAAGAGAAAGGAGAGT GTTATATG	8
2	Synthetic RBS	TTATATTGAAGGAGGATGTCTAGAATG	9
3	Pyruvate Formate Lysis	CAGATGCATGGGAGGGAAACTGTTATG	9
4	Fructose-bisphosphate aldolase, class II	TGAAGCATAAGGAGGATTTTACCATG	8
5	Phosphoglycerate kinase	CAATCAAAAAGGGGGCATAGCGCCATG	9
6	Isocitrate dehydrogenase	AAAATATATTGGAGGTTGTTTCTAGAATG	11
7	Acetate kinase	TGATGATTAAGGGAGTTGTTTGAATATG	11
8	6-phosphofructokinase	AACGAATTGAGGTGGCAACATG	4
9	Glucose-6-phosphate isomerase 2	ATTTTTGTTTGGAGGATCGAGCGTATG	9
10	Glyceraldehyde 3-phosphate dehydrogenase	TATGATGAAAGGAGAACATGAATG	6
11	Glycero-3-phosphate dehydrogenase	GGAGATCTTAGGAGAGACCATGCATG	8

Although based on a limited sample size, the results of this RBS screen give an initial indication that the distance has no correlation with translation efficiency. The strongest RBS sequences did fall within the 7-10 bp optimum of *E. coli*, however as did the weakest. This indicates other factors play a role in determining the translation efficiency of the RBS sequence. The composition, rather than length, of the sequence is a determinant of strength. The sequence upstream of the SD sequence also effects ribosomal binding and therefore translation initiation efficiency.

A set of RBS sequences from *Clostridium sporogenes* were previously screened for use in clostridial spp. (data not published). Here the strongest clostridia RBS characterised was compared to the *Geobacillus* synthetic RBS, to determine if the *C. sporogenes* RBS sequences are compatible with *Geobacillus*.

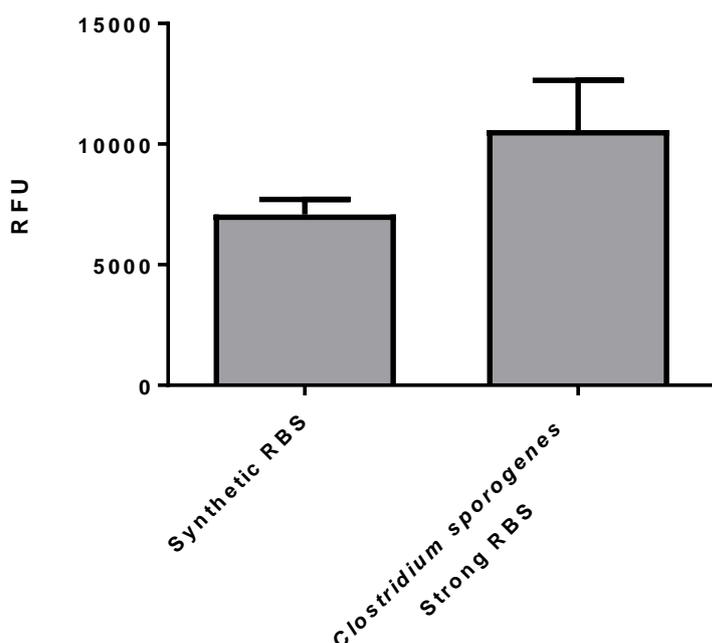


Figure 3.7. Comparative sGFP expression from a strong *C. sporogenes* RBS and the *Geobacillus* synthetic RBS. Two RBS sequences, as indicated, were cloned into a standardised vector; pMTL_gSlimS_Idh_RBS_T1T2_sGFP and expressed in *G. thermoglucosidasius*. The sGFP assay was carried out as previously described. All GFP fluorescence readings were corrected for medium autofluorescence and divided by corrected OD₆₀₀ readings to give an estimate of GFP per cell. All measurements were performed in triplicate, with the average and standard error, shown here. RFU; relative fluorescent units.

The *C. sporogenes* RBS was found to be stronger than the synthetic RBS construct with a statistically significant difference in RFU measured, shown in Figure 3.7. This indicates the *C. sporogenes* RBS sequences are compatible for use in *Geobacillus*. Expanding the RBS library of available components. The remainder of the *C. sporogenes* RBS set, were subsequently screened for strength characterisation (work carried out by Lau, M. and not presented here) and are available for use.

3.5 Orthogonal expression system

A system enabling modification to heterologous components without disrupting the host's own machinery required for cell survival, is explored here. Such 'orthogonal' expression systems have been used by researchers previously with reported high levels of expression (An & Chin, 2009). Orthogonal systems are uncoupled from evolutionary constraints and therefore are abstracted from cellular regulation, making use of heterologous elements, such as phage T7 polymerase. The advantage of using such polymerases resides in the fact that they are specifically targeted to the promoter employed in front of the transgene. Orthogonal systems are uncoupled from evolutionary constraints, as they rely on heterologous components which are not subject to native regulation by host cell machinery. Another advantage of using alternative polymerases is the specific targeting of the promoter in driving the expression cassette.

Here a system of components from *Clostridium difficile* are used. The system comprises an RNA polymerase sigma factor and a promoter recognised by the RNA polymerase sigma factor driving a gene of interest. The *tcdB* gene encodes Toxin B and is located in a pathogenicity locus along with *tcdR*. TcdR is a group 5 RNA polymerase sigma factor which directly activates the *tcdB* toxin gene expression. As *Geobacillus* does not harbour these components natively, expression will only be induced when both components are present. Zhang et al. (2015) demonstrated use of this orthogonal expression system to generate transposon mutants in *C. acetobutylicum*. Here the system is tested in *Geobacillus* to determine if the system is compatible in an alternative bacterial host.

The PheB reporter assay, as described previously, was used as a quantifiable reporter gene system to investigate the functionality of the *tcdB/tcdR* orthogonal expression system. The *tcdR* gene was integrated into the *G. thermoglucosidasius* chromosome at the *pyrE* locus using the ACE

methodology, as described by Heap et al. (2011). The P_{tcdB} promoter is positioned upstream of *pheB* on the pTMO372 plasmid.

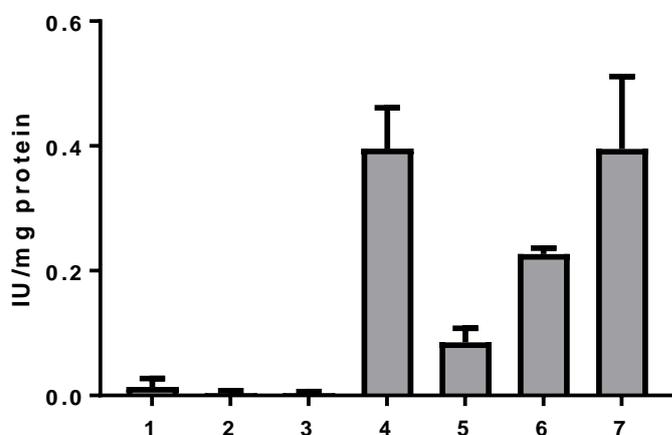


Figure 3.8. Use of the *tcdR/tcdB* orthogonal expression system in *G. thermoglucosidasius*. Graph showing *pheB* expression in *Geobacillus*, as determined by International Unit (IU)/mg protein value. IU measured as the average slope of absorbance at 375 nm over a one minute period after exposure to catechol. Results adjusted to standardise protein concentrations, with protein concentration measured by Bradford assay. Cultures were grown in CBM1X minimal medium. All measurements performed in triplicate, with the average and standard error, shown here. 1. Wild type *G. thermoglucosidasius* control containing an empty vector with no *pheB* gene, 2. Plasmid with no promoter in wild type *Geobacillus*, 3. Plasmid with no promoter in *Geobacillus* with *tcdR* integrated, 4. Plasmid containing P_{idh} in wild type *Geobacillus*, 5. *Geobacillus* with *tcdR* integrated in the genome, 6. Plasmid containing P_{tcdB} promoter in wild type *Geobacillus*, 7. Plasmid containing P_{tcdB} in *Geobacillus* with *tcdR* integrated.

Results here demonstrate the orthogonal system is functional in *Geobacillus*. To our knowledge, this is the first time *tcdB/tcdR* orthogonal expression has been demonstrated in a non-clostridial host organism. The *pheB* expression resulting from the orthogonal system gave the same level of expression as the

non-regulated constitutive P_{idh} . This result is comparable to that of *C. acetobutylicum* and *C. sporogenes*, where the *tcdB/tcdR* expression levels were equivalent to the strong P_{fdx} (Zhang et al., 2015). Expression of the two constructs containing only one component of the system, Figure 3.8. columns 5 and 6, show the system is not tightly regulated.

Other group 5 RNA polymerase sigma factors which could be used include; BotR, TetR or UviA. The application of transposon mutagenesis to generate a mutant library, as described previously (Zhang et al., 2015) could be applied to *G. thermoglucosidasius* using this technique. This system could also be utilised for cloning of toxic genes. The *tcdA* and *tcdB* promoters are poorly recognised by the *E. coli* RNA polymerase. This allows vectors carrying the promoters to be easily manipulated in *E. coli* without any potential toxic or other growth side effects arising from unwanted expression of the nucleic acid operably linked to the *tcdA* and *tcdB* promoter.

3.6 Development of the modular vector system

Replicative plasmids are basic tools in molecular microbiology. Plasmids are circular, double-stranded DNA molecules, distinct from chromosomal DNA. Plasmids are commonly used to transfer DNA into a host strain for the production of recombinant proteins. Cells containing a plasmid are typically selected and maintained using an antibiotic resistance marker encoded by the plasmid. The relevant antibiotic is added to the growth medium, creating a selective pressure. Plasmids are replicated during cell division and segregate to the daughter cells.

E. coli is used as a host for molecular cloning, prior to transfer of the plasmid into the organism of interest. Plasmids therefore need to include all the features required for replication and maintenance in both hosts. To enable ease of manipulation and reduce burden on the cell, plasmids should be optimised to contain the minimum DNA. The essential DNA sequences include;

a DNA replication origin, an antibiotic resistance gene and a gene of interest or region in which exogenous DNA fragments can be inserted.

The pMTL60000 vector series was recently introduced by Sheng et al. (2017). Here a collection of modules to expand the pMTL60000 series for *G. thermoglucosidasius* are designed and tested. The additional modules developed have extended the plasmid kit to now provide multiple options for each component. A user can select components based on experimental requirements. Rare 8 bp type II restriction enzyme recognition sites between each component allow simple modification to the plasmid constituents. Demonstrated here for the first time, with the correct combination of elements co-transformation of two plasmids can be performed in *Geobacillus*. This has not previously been reported in any *Geobacillus* sp.

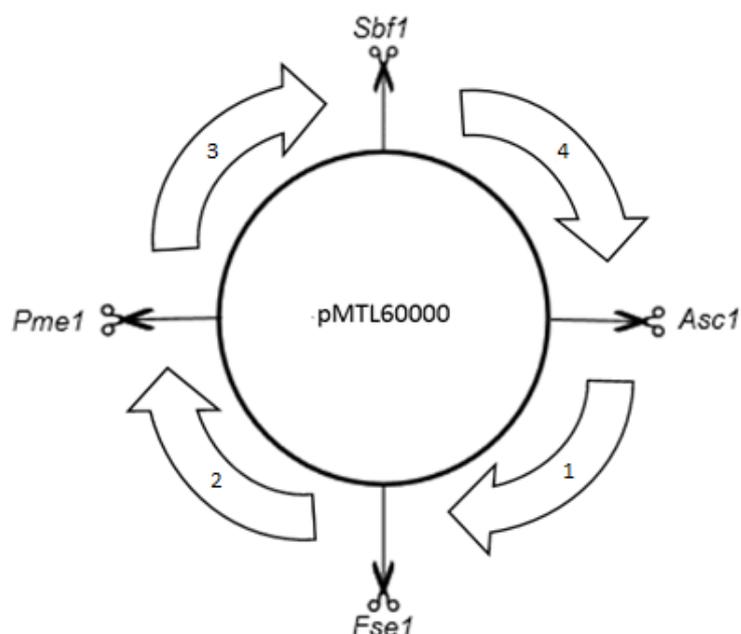


Figure 3.9. The pMTL60000 vector series format. Numbered arrows indicate the standardised arrangement of plasmid components. 1. Gram-positive replicon, 2. Marker, 3. Gram-negative replicon, 4. Application specific module. Sbf1, Asc1, Fse1 and Pme1 are 8 bp type II restriction enzyme sites between the components positioned as indicated.

Table 3.2. Component modules of the pMTL60000 vector series. All modules will be detailed subsequently.

1. Gram-positive replicon	2. Marker	3. Gram-negative replicon	4. Application specific module
0. Spacer	0.	0.	0. Spacer
1. UB110.1	1. Kanamycin	1. ColE1 +oriT	1. MCS
2. UB110.2	2. Thermostable Kanamycin	2. ColE1	2. P _{ldh} + MCS
3. UB110.3	3. Spectinomycin	3. p15a	3. P _{gapd} + MCS
4. NCI001		4. p15a+oriT	4. MCS + sGFP reporter
5. NCI002			5. MCS + PheB reporter
6. BST1			6. MCS + LicB reporter
7. GEOTH02			7. MCS + eCGP123 reporter

A standardised naming convention of; ‘pMTL6’ followed by the numbered combination of modules, in the correct order, is used. This provides the plasmid name as a shorthand recognition for the whole plasmid sequence.

3.6.1 Gram-positive replicons

Bacterial plasmids are able to replicate as autonomous genetic elements within a host. The replicon is comprised of an origin of replication (ORI) and control elements. The ORI is the point from which replication starts. A plasmid encoded replication-initiation (Rep) protein interacts with specific sequences at the ORI. The Rep protein initiates replication by nicking of the plasmid and recruitment of its host’s transcriptional machinery.

There are three replication mechanisms for circular plasmids; theta type (Θ), strand displacement, and rolling circle (RC). During theta and strand displacement replication occurs by melting of the parental strands, followed by RNA priming and initiation of DNA synthesis by covalent extension of primer RNA. DNA synthesis is continuous on the leading strand and discontinuous on

the lagging strand. Replication can occur either single or bidirectional. Whereas during rolling circle replication one of the DNA strands is cleaved at a double strand origin initiating leading strand synthesis. This first nick generates a 3'-OH end, from which a polymerase synthesises the leading strand. A second nick results in a highly unstable, single stranded DNA intermediate. From the single strand origin the alternative or lagging strand is synthesised, forming the double stranded plasmid DNA.

For propagation in *G. thermoglucosidasius* functional Gram-positive and thermostable replication origins are required. Here five Gram-positive replicons, were characterised, for use with *G. thermoglucosidasius*.

Table 3.3. Gram-positive replicon source plasmids. RC; Rolling Circle, Θ ; Theta Replication.

Replicon name	Size (bp)	Source	Replication mechanism
1. UB110.1	1428	<i>Staphylococcus aureus</i> cryptic plasmid pUB110.	RC
2. UB110.2	1378	Modified UB110.1 (Sheng et al., 2017).	RC
3. UB110.3	1205	Modified UB110.1 (Sheng et al., 2017).	RC
4. NCI001	2908	<i>G. thermoglucosidasius</i> megaplasmid pNCI001.	RC
5. NCI002	3559	<i>G. thermoglucosidasius</i> megaplasmid pNCI002.	RC
6. BST1	3056	<i>Bacillus stearothermophilus</i> cryptic plasmid pBST1.	Θ
7. GEOTH02	2205	<i>Parageobacillus thermoglucosidasius</i> C56-YS93 plasmid pGEOTH02.	RC

3.6.1.1 Plasmid segregational stability

Plasmid DNA is replicated upon cell division where each daughter cell receives at least one copy of the plasmid. When all daughter cells maintain at least one plasmid during cell division, the culture is segregational stable. If plasmid free cells are generated then the culture is segregationally unstable. The degree of instability is determined by the rate of plasmid loss per generation. If the plasmid is expressing a particular gene or genes of interest, segregational instability can lead to a significant loss in expression level and therefore reduced productivity.

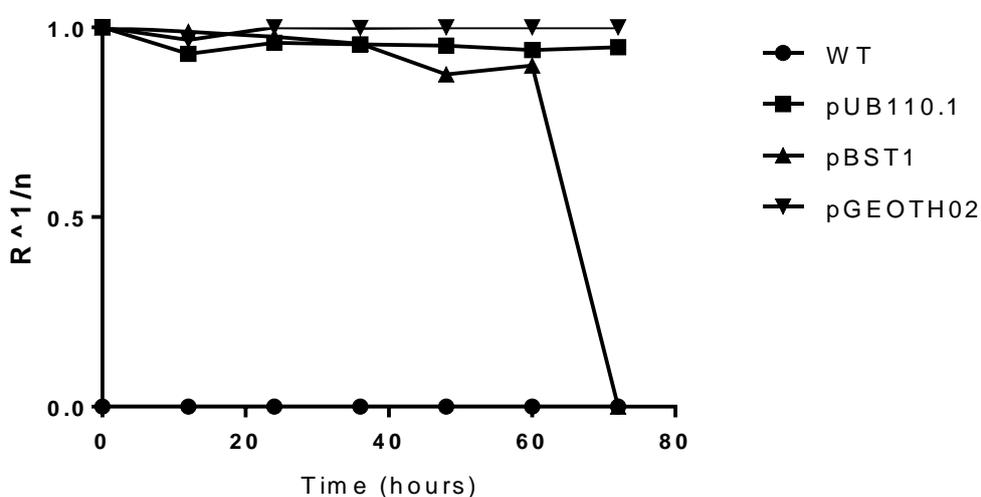


Figure 3.10. Plasmid retention per generation at 52°C. Plasmid segregational stability assays were carried out as described in section 2.12. *G. thermoglucosidasius* was transformed with plasmids containing the replicons indicated. Incubation was carried out at 52°C. The percentage plasmid loss was calculated using the difference between number of non-resistant and resistant colonies. Plasmid retained per generation was calculated with the equation R^n and plasmid lost per generation as $1 - R^n$, where n is the generation number and R is the percentage of cell population retaining the plasmid. Given the 1% inocula used, and assuming the cultures reach maximum cell density in 12 h, we took the number of generations per 12 h to be 6.64 ($2^{6.64} = 100$).

Table 3.4. Plasmid presence in culture after 72 hours at 52°C.

Replicon	Plasmid retention per generation ($R^{1/n}$)	Percentage loss (%)
WT	N/A	N/A
UB110.1	0.948172	5.18281
BST1	0	100
GEOH02	1	0

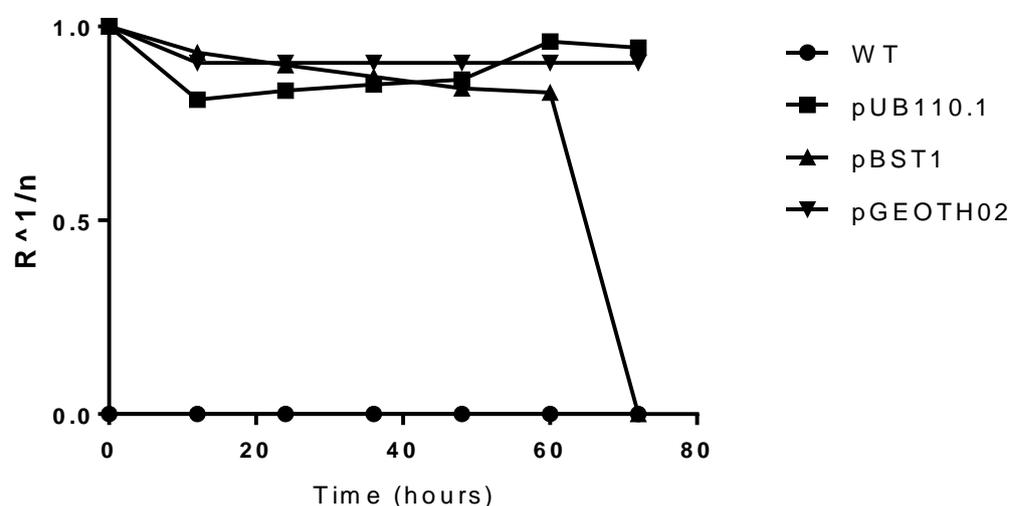


Figure 3.11. Plasmid retention per generation at 60°C. Plasmid segregational stability assays were carried out as described in section 2.12. *G. thermoglucosidasius* was transformed with plasmids containing the replicons indicated. Incubation was carried out at 60°C. The percentage plasmid loss was calculated using the difference between number of non-resistant and resistant colonies. Plasmid retained per generation was calculated with the equation $R^{1/n}$ and plasmid lost per generation as $1 - R^{1/n}$, where n is the number of generation and R is the percentage of cell population retaining the plasmid. Given the 1% inocula used, and assuming the cultures reach maximum cell density in 12 h, we took the number of generations per 12 h to be 6.64 ($2^{6.64} = 100$).

Table 3.5. Plasmid presence in culture after 72 hours at 60°C.

Replicon	Plasmid retention per generation ($R^{1/n}$)	Percentage loss (%)
WT	N/A	N/A
UB110.1	0.945479	5.452144
BST1	0	100
GEOETH02	0.906474	9.352638

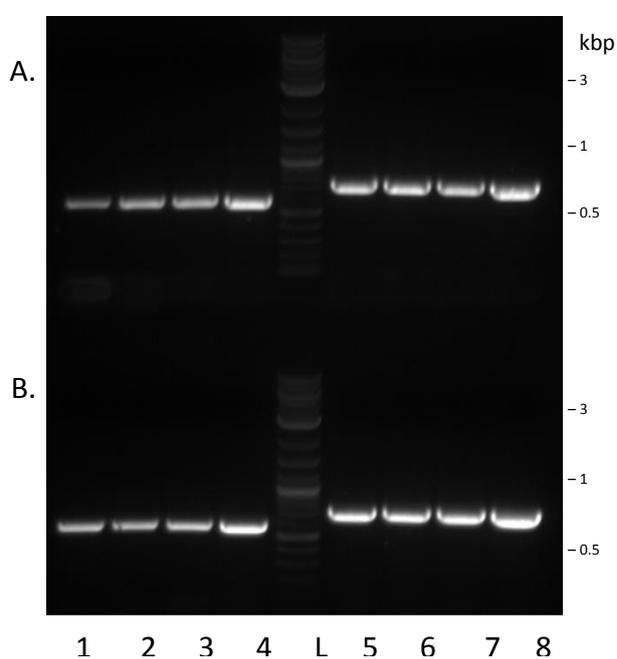


Figure 3.12. Visualisation of colony PCR confirming plasmid presence at the end of the stability experiment. Four colonies of each; A. pUB110.1 and B. pGEOETH02 were selected from plates grown after 72 h incubation at 60°C. Two target regions of the plasmid were amplified in each colony; 1 – 4. 600bp region of ColE1, 5 – 8. 687bp region of GFP. PCR protocol as previously described, used 55°C annealing temperature and 1 minute extension time. L indicates 2-log ladder.

Here three of the plasmids were tested at both 52°C and 60°C, with wild type cells as a control. The plasmids harbouring the replicons GEOTH02 and UB110.1 were found to be highly segregationally stable at both temperatures. Both cultures maintained high plasmid retention after 72 hours without antibiotic selection. After 72 hours 5.2% and 5.5% of cells containing the UB110.1 replicon lost the plasmid when incubated at 52°C and 60°C respectively. This data is concurrent with previously published studies of the pUB110 plasmid. Seyler et al. (1993) recorded 95-100% segregational stability of pUB110 in *B. subtilis*. The results presented here demonstrates this holds true for *Geobacillus*, where at higher temperatures 5% loss was observed.

Cultures transformed with the plasmid containing the GEOTH02 replicon were found to be highly segregationally stable. No plasmid loss was observed after 72 hours from cultures incubated at 52°C. After 72 hours at 60°C 9.4% of cells lost the plasmid. This high level of stability could be a result of the native function of the replicon in maintaining the pGEOTH02 plasmid in its host *G. thermoglucosidasius* C56-YS93. However, as *G. thermoglucosidasius* C56-YS93 has a temperature range of 55 to 75°C and an optimum temperature of 65°C (Brumm et al., 2015) this would suggest the replicon would confer stability at the higher temperature. It could be the replicon plays a role in maintaining the plasmid stability along with other factors such as essential or beneficial gene presence.

Plasmids harbouring the BST1 replicon were found to be highly unstable. At both 52°C and 60°C 100% of the culture lost the plasmid in 72 hours. This property, although not ideal for gene expression plasmids, could be utilised for molecular engineering applications. For strategies where DNA is delivered to the cell then the plasmid needs to be lost, such as gene integration and knock out and CRISPR vectors, this replicon choice is ideal.

To expand upon the selection of replicon properties on offer, modifications were made to the stable UB110 replicon, resulting in the three variants now available. It had previously been reported that a 358 bp incompatibility region that resides 5' to the UB110 gene acts as a trans-acting element involved in the control plasmid replication (Maciag et al., 1988). Two derivative replicons were constructed in which the 5' incompatible region was reduced from 412 bp to 362 bp and to 189 bp, to give three variants UB110.1, UB110.2 and UB110.3 respectively (work carried out by Sheng, L.) Both of the new plasmids exhibited a significant increase in segregational instability with less than 10% of the cells retaining the plasmid after 72 h, as shown in Figure 3.13. (Sheng et al., 2017). This indicates this region adjacent to the replicon plays a role in plasmid segregational stability.

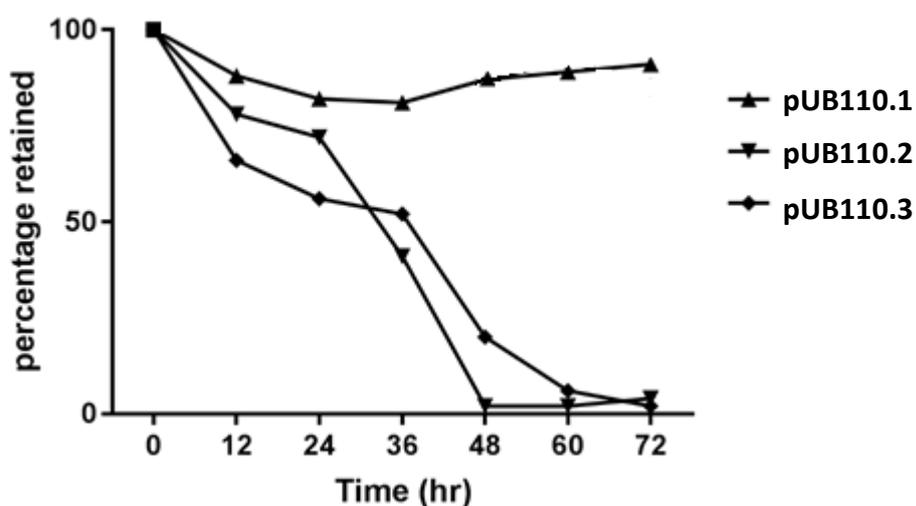


Figure 3.13. Plasmid retention per generation of plasmids containing UB110 replicons following modification (Data from Sheng et al., 2017). Modification to reduce the incompatibility region of the UB110 replicon resulted in reduced segregational stability at 52°C. Replicons UB110.2 and UB110.3 can be utilised for applications where instable plasmids are required.

Although no experimental data is presented here, both replicons NCI001 and pNCI002 are expected to be highly stable. The replicons originate from *G.*

thermoglucoasidarius' native megaplasms, which are continuously maintained in culture. The replicons are expected to play a significant role in maintenance of the megaplasms, other factors may also influence the plasmid stability. Such as a toxin/antitoxin system present on megaplasmid pNCI001. There could also be other essential or beneficial genes encoded on the megaplasms, providing a cellular advantage to the host.

3.6.1.2 Plasmid copy number

Plasmid encoded gene expression and recombinant protein production levels are determined by the plasmid copy number. In order to predict gene expression levels, quantification of the plasmid copy number is insightful. Factors affecting the copy number of a plasmid are regulation of the replicon, plasmid size and growth conditions. Regulation of Rep protein expression in turn regulates its binding to the ORI and therefore controls initiation of plasmid replication. The replicon can be regulated by antisense RNA or iteron binding groups. Low copy number or stringent plasmids have tighter control of replication.

In order to gain an understanding of the properties of the Gram-positive replicons available in the vector series presented here, plasmid copy number was investigated. To determine the plasmid copy numbers real-time quantitative PCR (qPCR) was used. In this work, a total DNA isolation protocol followed by qPCR amplification using two primer sets targeting the plasmid and genomic DNA specifically for quantitation of plasmid copy number was carried out. LuminoCt® SYBR® Green qPCR ReadyMix™ uses a fluorescent dye to quantitatively measure the amplification of DNA during qPCR. SYBR Green fluorescent dye binds to double-stranded DNA molecules by intercalating between the DNA bases. Fluorescence is measured after each amplification cycle to determine DNA amplification. Relative quantification based on the genomic amplification is used to determine copy number per chromosomal copy ratio, assuming the chromosomal value to be one. A region in the *lepA1*,

elongation factor 4 gene, was used as the chromosomal target for qPCR amplification. Full details of the protocol used have been previously detailed.

Table 3.6. qPCR primers. PCR primers were designed using IDTDNA online software (<https://eu.idtdna.com>). The qPCR primers were designed between 20 and 24 bases in length and took into consideration T_m value and specificity. Primers free of strong secondary structures and self-complementarity were used. Melting temperatures, equal and close to the optimum 62°C were used.

Replicon	Primer	Positions			Dimer			ΔG (kcal/mole)	
		Start	End	Length	T_m (°C)	GC(%)	Hairpin	For	Rev
UB110.1	Forward	632	652	21	62.3	42.9	0.00	-4.34	-5.89
	Reverse	748	727	22	62.3	45.5	0.00		-4.88
NCI001	Forward	1062	1082	21	61.6	47.6	-0.05	-2.90	-5.33
	Reverse	1139	1116	24	62.0	41.7	0.00	-4.56	-3.14
NCI002	Forward	864	883	20	61.8	50.0	-0.31	-4.80	-3.47
	Reverse	1010	990	21	61.7	47.6	0.00		-6.53
BST1	Forward	1854	1873	20	61.8	50.0	0.00	-4.34	-5.96
	Reverse	2003	1984	20	61.6	50.0	-0.24		-4.48
GEOTH02	Forward	1270	1292	23	62	47.8	0	-4.88	-4.42
	Reverse	1377	1358	20	61.8	50	0		-2.24
Chromosomal lepA1 gene	Forward	363	383	21	61.7	52.4	0.00	-4.34	-4.41
	Reverse	482	462	21	61.9	47.6	0.00		-4.88

Table 3.7. Size of qPCR amplification products. qPCR amplicons of <200bp in length were designed for each replicon.

Replicon	Amplicon Region		
	Start	End	Length (bp)
UB110.1	632	748	117
NCI001	1062	1139	78
NCI002	864	1010	147
BST1	1854	2003	150
GEOH02	1270	1377	108
Chromosomal <i>lepA1</i> gene	363	482	120

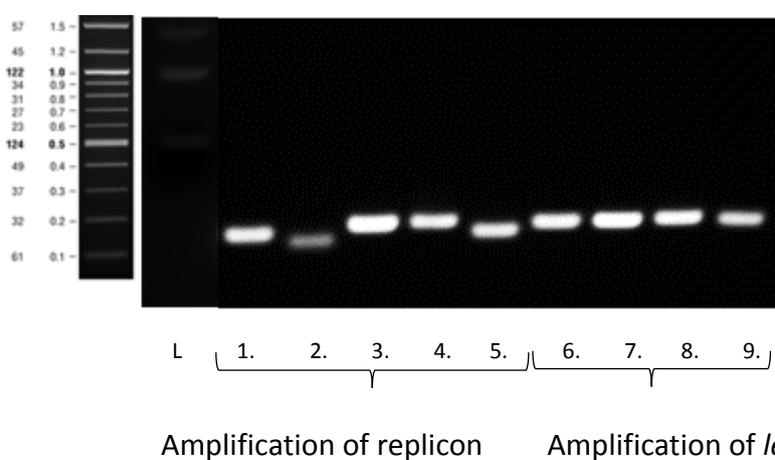


Figure 3.14. Visualisation of plasmid and chromosomal target amplification following a standard PCR protocol. PCR was carried out to confirm the primers are specific to the desired target, in order to ensure accurate quantification in qPCR experimentation. PCR protocol as previously described with annealing temperature of 57°C and an extension time of 10 seconds. Template DNA used was genomic DNA of wild type *G. thermoglucosidasius* (lanes 2, 3 and 7) or genomic DNA from *G. thermoglucosidasius* transformed with plasmids

containing the following Gram-positive replicons; UB110.1 (lanes 1 and 6), BST1 (lanes 4 and 8) and GEOTH02 (lanes 5 and 9). Primer sets used; 1. UB110.1, 2. NCI001, 3. NCI002, 4. BST1, 5. GEOTH02, 6 to 9., *lepA1*. L indicates 2-log ladder.

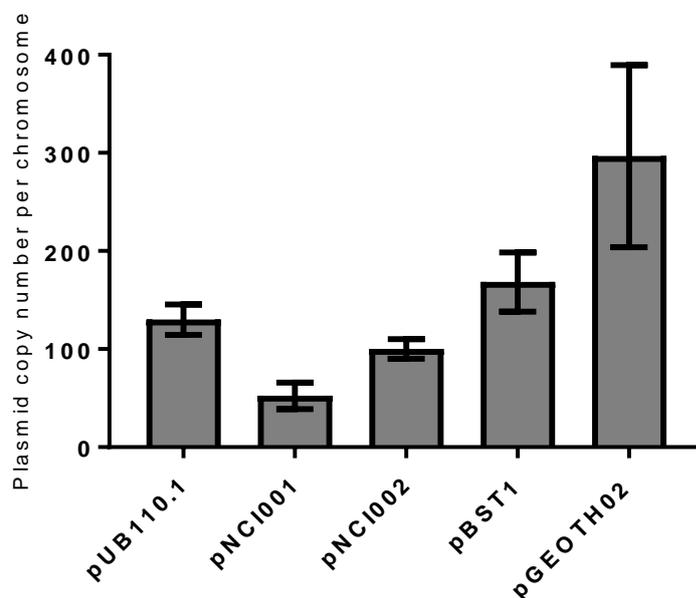


Figure 3.15. Estimation of plasmid copy number per chromosome. The qRT-PCR protocol described previously was used to estimate the relative copy number of plasmids containing the Gram-positive replicons indicated. *G. thermoglucosidasius* cultures were transformed with plasmids UB110.1, BST1 and GEOTH02. *G. thermoglucosidasius* was used for quantification of plasmids pNCI001, pNCI002 and the chromosomal standard. Cells were grown at 52°C for 24 hours prior to DNA extraction. The average and standard deviation from triplicate experiments are shown.

Results show all three of the heterologous replicons result in higher copy number than the native megaplasmids. The copy number of NCI002 is twice that of NCI001, with averages of 100 and 52 respectively. As pNCI001 is larger its lower copy number may be due to the size of the plasmid, as it will confer a greater metabolic burden on the cell.

Plasmids containing the UB110.1, BST1 and GEOTH02 replicons are present with approximately 130, 168 and 297 copies per cell, respectively. The relatively high copy number associated with the BST1 replicon was unexpected. As theta type plasmids are usually maintained at a low 1 to 10 copy number per cell (Kananavičiūtė et al., 2014). In addition BST1 was found to have low segregational stability, Figures 3.10 and 3.11. High plasmid copy number means balanced distribution of plasmids between daughter cells after cell division is more likely, therefore high copy number plasmids have an increased chance of maintaining segregational stability. This high value copy number could be due to a lack of control elements. Regulatory mechanisms could be encoded in an up or downstream sequence not included in the replicon region included here.

Plasmids containing the GEOTH02 replicon were found to have the highest plasmid copy number. High copy number is a desirable feature for expression vectors and therefore this replicon could be useful for *Geobacillus* expression vectors where high levels of protein production is desired.

The results presented here offer an estimation of the replicon properties with regard to copy number. The relative concentrations between the replicons tested can be assumed. The methodology however is limited by differing efficiencies of plasmid and chromosomal DNA extraction and also from variability during the cells growth phases. The stage of sampling will causes difference in result, with large variations in copy number recorded across lag to exponential and stationary phases (Lee et al., 2005).

Here stationary phase cells are tested, which is when secondary metabolite product formation takes place, such as *n*-butanol. Therefore the plasmid copy number at stationary phase will give an indication as to the plasmid gene expression levels and protein production rates at this time. However dependent on application, gene expression at differing stages of growth may be desired, in this case the sample point should reflect this.

3.6.1.3 Host range

Plasmids, as vehicles for horizontal gene transfer, are able to move between bacterial strains and species. Plasmids can be categorized as having a narrow or broad host range with regard to their ability to transfer and be maintained in distantly related bacterial hosts. The transfer of narrow host range plasmids can be limited by the formation of mating pairs, the avoidance of the recipient's restriction system or the correct expression of its replication and maintenance systems in the recipient (Thomas & Nielsen, 2005).

The modular Gram-positive replicons were tested in *Bacillus subtilis* strain 168 to determine transfer and functionality in an alternative host. Demonstrating functionality of the replicons in *B. subtilis* would broaden the range of application of the vectors rather than limit the series to use only in *Geobacillus*.

An initial antibiotic sensitivity assay was carried out to determine the sensitivity of *B. subtilis* 168 to spectinomycin, further detailed in section 3.6.2. A working concentration of 200 µg/ml was used.

B. subtilis cultures were transformed with plasmid pMTL6x320, using the *B. subtilis* transformation method, as described previously. X representing one of the five Gram-positive replicons, 1. UB110.1, 2. NCI001, 3. NCI002, 4. BST1 and 5. GEOTH02. A control transformation using no vector and an untransformed control were included. An additional control plasmid was transformed. In the control vector, pMTL60320, the Gram-positive replicon was replaced with a spacer. This control was included to ensure the ColE1 replicon did not allow for plasmid expression in *B. subtilis* 168. Transformations were carried out in triplicate. Following transformation plasmid presence was selected for using antibiotic selection. All of the transformations including a Gram-positive replicon resulted in colonies the presence of antibiotic. Untransformed and no vector control cultures did not produce any colonies. The ColE1 control using pMTL60320, did not produce any colonies following transformation.

This result indicated the plasmids were functional and expressing the antibiotic resistance gene in *B. subtilis* cells. No colonies on the control plates showed it was the presence of the Gram-positive replicon allowing for replication in *B. subtilis* and not due to ColE1 host range. To confirm this result, colony PCR was carried out to detect plasmid presence.

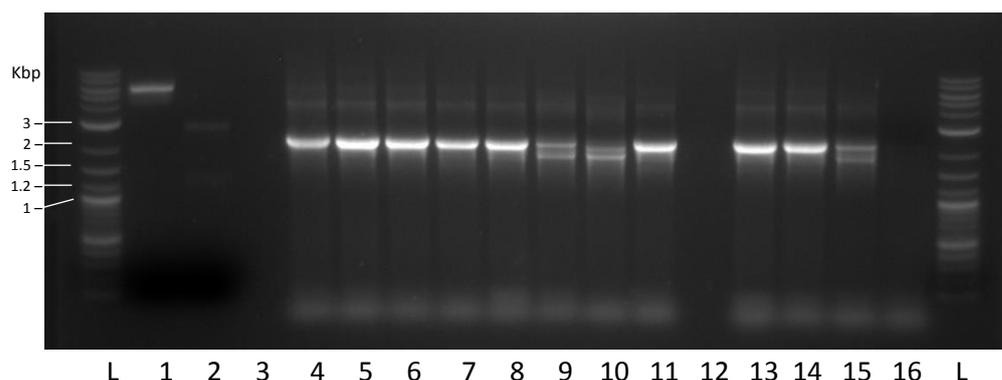


Figure 3.16. Colony PCR confirmation of plasmid presence in *B. subtilis*.

Colony PCR was carried out as previously described, with extension time of 1.5 minutes and a melt temperature of 54°C. Primers aad_F and M13_F were used to amplify the remainder of the plasmid (aad9, ColE1 and spacer) minus the replicon gene with an expected size 2271bp. Colony PCR was carried out in triplicate. L indicates 2-log ladder, 1. Positive control using plasmid DNA known to amplify with 6kbp product. The remaining lanes used *B. subtilis* colonies transformed with the following replicons; 2. 3. 4. UB110.1, 5. 6. 7. NCI001, 8. 9. 10. NCI002, 11. 12. 13. BST1 and 14. 15. 16. GEOTH02.

A further check for replicon compatibility used the sGFP fluorescent reporter assay to confirm gene expression in the *B. subtilis* host. Plasmids pMTL6x324 containing the sGFP reporter gene, and one of the five available Gram-positive replicons, were transformed into *B. subtilis*. Transformed colonies were grown overnight in liquid culture and fluorescence was measured.

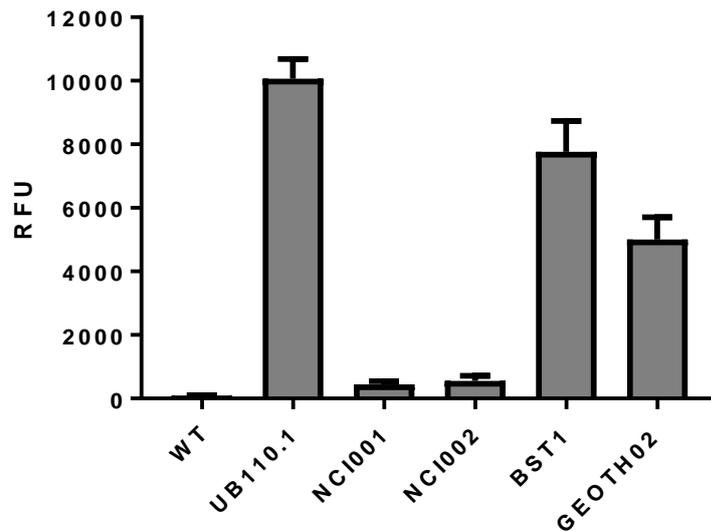


Figure 3.17. Detection of plasmid based gene expression in *B. subtilis*. The sGFP reporter assay was used, as previously described, to confirm plasmids containing the Gram-positive replicons as indicated are functional in a *B. subtilis* host. Cultures were grown overnight at 37°C. RFU values were normalised to OD 1.

This result demonstrates the Gram-positive replicons presented here are not specific to *Geobacillus* and are also functional in *B. subtilis*. This result further gives indication as to the expression levels of the plasmids in *B. subtilis*, however this result cannot be quantified. The vector series can be considered a more diverse tool, not only limited to use in *Geobacillus*. Further investigation of additional and more diverse strains is required to understand the extent of compatible host range. Further screening of alternative species would determine if the replicons confer narrow or broad host range functionality to a plasmid.

3.6.1.4 Compatibility

Multiple plasmids are able to co-exist within a cell. However if plasmids share the same replication control mechanism, they can be incompatible. In this case, both plasmids contribute to the total copy number and are regulated together.

In this situation it is likely that one of the plasmids will be out-copied and lost during cell division.

Co-transformation of multiple plasmids is desirable as a molecular tool. Use of multiple plasmid hosts for heterologous gene expression enables, ease of plasmid construction and take up. Use of multiple plasmids can also help to reduce plasmid size when working with many genes. In order to develop co-transformation in *Geobacillus*, compatible replicons must be identified.

Table 3.8. Compatibility of Gram-positive replicons.

	1. UB110.1	2. UB110.2	3. UB110.3	4. NCI001	5. NCI002	6. BST1	7. GEOTH02
1. UB110.1		x	x	✓	✓	x	✓
2. UB110.2	x		x	✓	✓	x	✓
3. UB110.3	x	x		✓	✓	x	✓
4. NCI001	✓	✓	✓		✓	✓	✓
5. NCI002	✓	✓	✓	✓		✓	✓
6. BST1	x	x	x	✓	✓		✓
7. GEOTH02	✓	✓	✓	✓	✓	✓	

Through transformation of plasmids containing each of the replicons, a trial and error approach was adopted to determine which of the replicons are compatible and to identify any incompatibilities. It is known that the NCI001 and NCI002 megaplasmids are compatible with the other replicons from single transformation experiments. It is also given the megaplasmid replicons are compatible with each other. Of the remaining replicons UB110 and BST were found to be incompatible. Co-transformation resulted in curing of the BST plasmid. Apart from this identified incompatibility, all other replicons can be used in combination and will be maintained within a cell.

3.6.2 Marker

The second module in the vector series is a marker gene. Here antibiotic resistance genes are used for selection of plasmid-containing bacteria by providing a survival advantage to the bacterial host. Growth in presence of the appropriate antibiotic creates a selective pressure ensuring the bacteria retain

the plasmid DNA for survival, despite the added replication burden. For thermophilic use, choice of antibiotic is limited as both the antibiotic and resistance mechanism must be heat stable. Here the choice of marker gene for use with *Geobacillus* has been extended. Two additional antibiotic resistance genes have been tested and put into the modular format.

Two genes conferring resistance to kanamycin were previously available and reported for use with *Geobacillus* (Taylor et al., 2008; Sheng et al., 2017). A kanamycin adenylyltransferase gene (*kan*) and a thermostable version (*kanHT*). The *kan* gene does not confer resistance to kanamycin above 60°C, whereas the thermostable *kanHT* is functional up to 70°C (Liao and Kanikula, 1990).

There was a desire for additional marker genes to provide alternative options and also to enable co-transformation of multiple plasmids within a cell. The antibiotic spectinomycin and its resistance gene *aad*, were first considered. Spectinomycin is an aminocyclitol antibiotic which inhibits bacterial protein biosynthesis by reversibly binding to the 30S ribosomal subunit (Kehrenberg et al., 2005). The spectinomycin resistance gene (*aad*) encodes an enzyme which inactivates the drug by adenylylation at the 9-OH position of the spectinomycin actinamine ring (Kehrenberg et al., 2005). Zhou and co-workers first published the use of spectinomycin for a thermophilic organism (Zhou et al., 2016). In this work, a spectinomycin resistance gene was used successfully for both selection of transformants at 52°C and integrants at 68°C in *Geobacillus*.

Here first the sensitivity of *Geobacillus* to the antibiotic was determined in order to establish an appropriate working concentration which would prevent all WT growth. Following this the *aad* resistance gene, in modular format was transformed into *G. thermoglucosidasius* on a pMTL61320 plasmid to determine if the gene conferred resistance to the antibiotic. Experimentation was carried out at both 52°C and 60°C.

Table 3.9. Minimum inhibitory concentrations (MIC) of spectinomycin.

	37°C		52°C		60°C	
	Liquid culture	Agar plated	Liquid culture	Agar plated	Liquid culture	Agar plated
<i>E. coli</i>	45 µg/ml	50 µg/ml				
<i>G. thermoglucosidasius</i>			12.5 µg/ml	12.5 µg/ml	75 µg/ml	75 µg/ml
<i>B. subtilis</i> 168	100 µg/ml	200 µg/ml				

Liquid culture MIC was determined by inoculating LB (*E. coli* and *B. subtilis*) or 2SPYNG (*G. thermoglucosidasius*) media containing a range of spectinomycin concentrations to OD₆₀₀ 0.2 with starter culture. Agar plated MIC was determined by plating 200 µl of starter culture onto LB (*E. coli* and *B. subtilis*) or TSA (*G. thermoglucosidasius*) containing a range of spectinomycin concentrations. The MIC in Table 3.9 represents the lowest concentration resulting in no observed colonies on plate or no change in OD₆₀₀ observed, after 48 h for agar plated and liquid cultures respectively. The working spectinomycin concentration for all subsequent experimentation was taken as; 50 µg/ml for use with *E. coli*, 12.5 µg/ml for use with *Geobacillus* at 52°C and 200 µg/ml for use with *B. subtilis*.

These results indicate spectinomycin is slightly heat sensitive, as higher concentrations are needed to inhibit *G. thermoglucosidasius* at 60°C than at 52°C. Some liquid cultures were found to have lower MIC than agar plated cells, this could be due to diffusion of the antibiotic in the agar.

Plasmid pMTL61320, containing the spectinomycin resistance marker, was constructed and transformed into *E. coli*, *G. thermoglucosidasius* and *B. subtilis* 168. Following transformation the cells were grown on spectinomycin selection plates at the concentrations determined here. The resulting colonies indicated that the *aad* resistance gene confers resistance to spectinomycin and can

therefore be used as an alternative marker gene. To confirm this result colony PCR was performed to show plasmid presence.

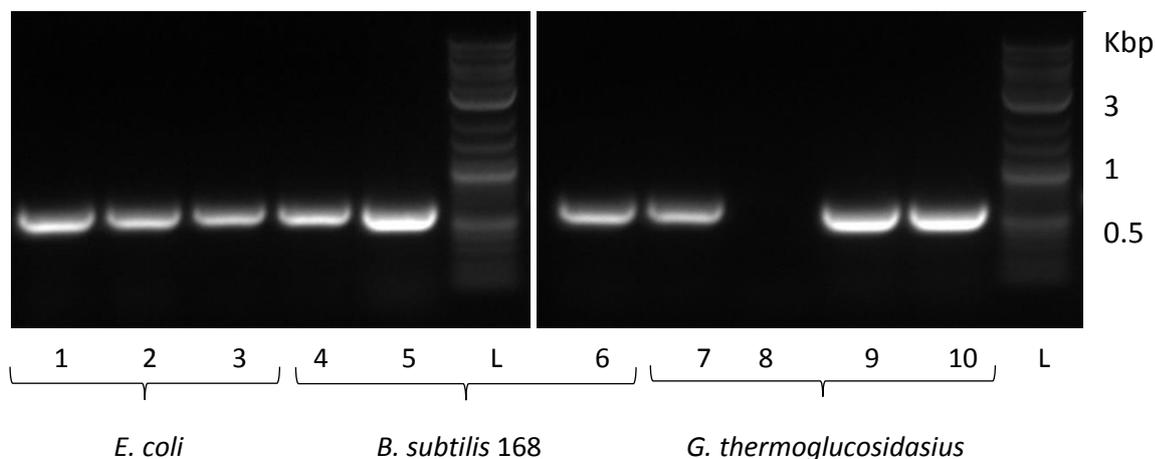


Figure 3.18. Colony PCR confirming aad derived spectinomycin resistance.

Visualisation of bands amplified following colony PCR of 1-3 *E. coli*, 4-6 *B. subtilis* 168 and 7-10 *G. thermoglucosidasius* from spectinomycin selection plates. PCR primers; RepB_F4 and RepB_R4, amplified a 613 bp region of the UB110.1 replicon gene.

A second antibiotic, gentamicin, was also investigated. Gentamicin was selected following reports of very high heat stability. Gentamicin has a broad spectrum of action, it is also an aminocyclitol antibiotic inhibiting bacterial protein biosynthesis by binding to the 30S ribosomal subunit. Gentamicin resistance by aminoglycoside modifying enzymes is encoded by N-acetyltransferase genes (*aac*). *Aac* catalyses acetyl CoA-dependent acetylation of an amino group. Use of gentamicin and an associated resistance gene has, to date, not been reported for use in *Geobacillus*. Here an *aac* gene from a *Pseudomonas* expression vector was found non-functional in *Geobacillus*. Resistance was conferred in *E. coli*. This is likely due to incompatibility of the gene in a Gram-positive host. This antibiotic is a promising option however an *aac* gene from a Gram-positive source should be tested.

Table 3.10. Minimum inhibitory concentrations (MIC) of gentamicin. MIC were determined as described above.

	37°C		52°C		60°C	
	Liquid culture	Agar plated	Liquid culture	Agar plated	Liquid culture	Agar plated
<i>E. coli</i>	3 µg/ml	4 µg/ml				
<i>G. thermoglucosidasius</i>			2 µg/ml	2 µg/ml	2 µg/ml	2 µg/ml

Other antibiotics were also considered; Chloramphenicol is available as a vector module, however it has not been used in this work as the antibiotic is reported to be only moderately thermostable (Taylor et al., 2008). Tetracycline is also available, however there have been reports of natural tetracycline resistance in *Geobacillus*. *G. stearothermophilus* strain (TK05) harbours the plasmid pSTK1 which contains a tetracycline resistance gene (Hoshino et al., 1985). Therefore this antibiotic is not ideal for use in *Geobacillus* as strains may be resistant to the antibiotic. Additional antibiotics which could be investigated for thermophilic use in future work include; tobramycin and vancomycin.

3.6.3 Gram-negative replicons

The third module in the vector series is the Gram-negative replicon which enables plasmid replication in *E. coli*. An *E. coli* shuttle vector is used for ease of construction and as a host for plasmid amplification. Further detail regarding replicon function and mechanism has been described in section 3.6.1.

In the modular series presented here ColE1, p15a and both with addition of oriT are available for use. The two replicons are compatible. ColE1 has a copy number of ~15-20 and p15a has a slightly lower copy number of ~10 (Wu & Liu, 2010).

An OriT (Origin of Transfer) is required for conjugation. This application has not been reported in *G. thermoglucosidasius* to date, however conjugation of *G.*

kaustophilus has been reported (Tominaga et al., 2016). Development of a conjugation protocol for *G. thermoglucosidasius* could be investigated as an alternative to the electroporation method in future work.

3.6.4 Application specific module

The fourth, and final module in the vector series is the application specific module. The other modules have encoded the necessary components for the plasmid to replicate in *E. coli* and *Geobacillus*. This module allows for plasmid application. The multiple cloning site (MCS) is a short segment of DNA containing several restriction enzyme sites, enabling easy insertion of DNA fragments by restriction enzyme digestion and ligation. In expression plasmids, the MCS is located downstream of a promoter, such that when a gene is inserted within the MCS, its expression will be driven by the promoter. Here P_{ldh} and P_{gapd} ; the strong *ldh* and *gapd* promoters, are available as a module for ease of cloning. The reporter genes are also present as application modules, for expression studies.

3.7 Co-transformation

Development of additional modules for each component in the vector series enabled construction of plasmids with wholly unique sequences. Such plasmids which are compatible both with *G. thermoglucosidasius* and each other, allow co-transformation of multiple plasmids in the same host cell. Co-transformation of multiple plasmids is a desirable molecular tool, offering advantages for ease of plasmid construction and take up when working with multiple genes. Here differing components for all four of the necessary plasmid parts were selected and constructed, creating two unique plasmids. Each containing a different reporter gene.

DEVELOPMENT OF MOLECULAR TOOLS

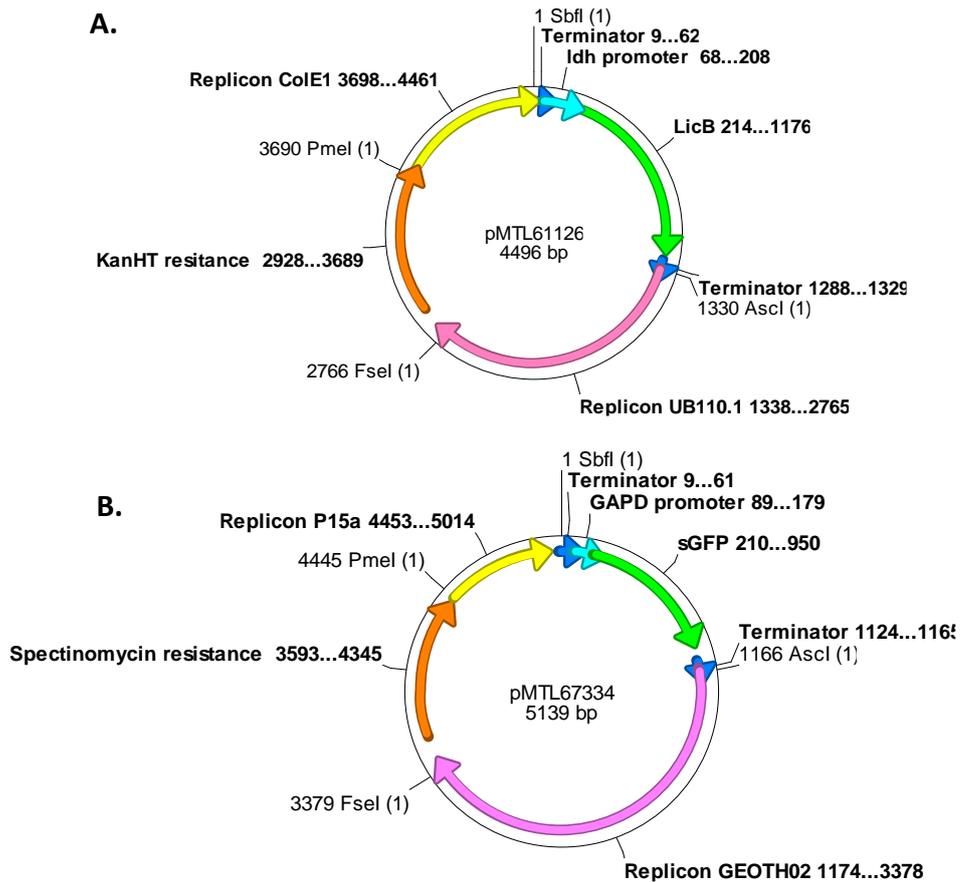


Figure 3.19. Maps of the plasmids co-transformed in *E. coli* and *G. thermoglucosidasius*. A. pMTL61226, B. pMTL67334.

Following construction of the above plasmids, they were transformed into both *E. coli* and *G. thermoglucosidasius*. Development of the methodology found *E. coli* could be co-transformed by introducing the plasmids together in a one-step heat shock. However, *Geobacillus* required two-step sequential transformation. First one plasmid was transformed by electroporation, following positive confirmation of plasmid presence, competent cells were made. Then the competent cells were transformed with the second plasmid. PCR was carried out to confirm presence of both plasmids and assays of both reporter genes were carried out to confirm gene expression.

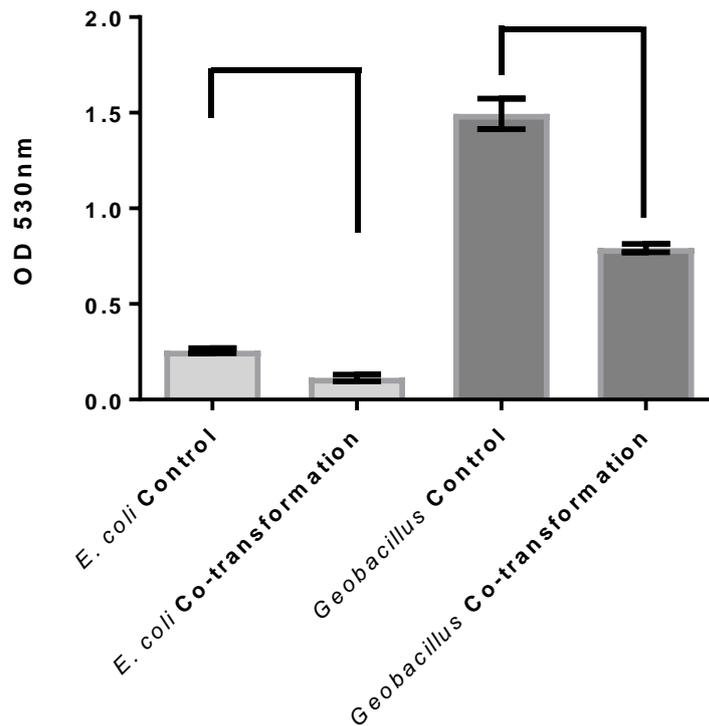


Figure 3.20. LicB assay of co-transformed *E. coli* and *G. thermoglucosidasius* culture. The LicB assay was carried out as described previously, with Congo Red used to detect pH change caused by lichenase activity in culture supernatant. Confirmed positive co-transformed cultures were compared with untransformed wild type cultures. The assay was carried out in triplicate with average and standard error shown. The reduction in OD at 530nm was significant for both, with the P value for *E. coli* 0.0003 and *G. thermoglucosidasius* 0.0001, therefore both significantly different at $P < 0.05$.

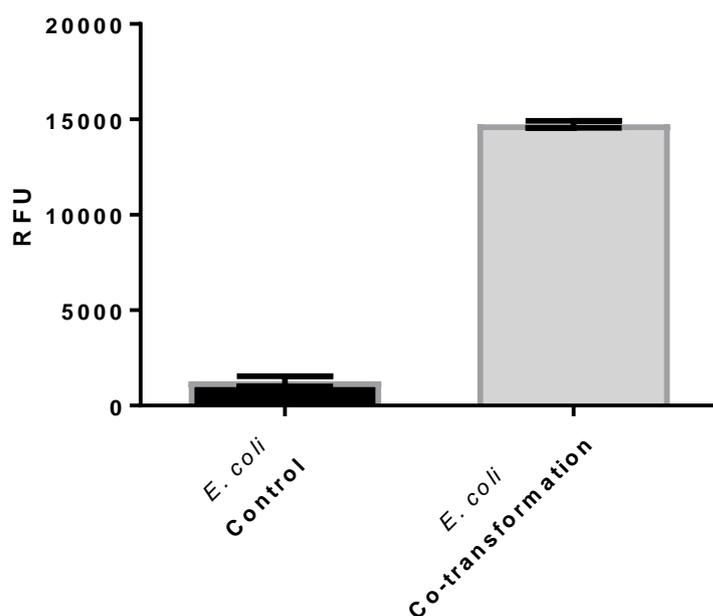


Figure 3.21. sGFP assay of co-transformed *E. coli* culture. The sGFP assay was carried out as described previously. Confirmed positive co-transformed cultures were compared with un-transformed *E. coli* Top10 wild type cultures. The assay was carried out in triplicate with average and standard error shown. The difference in RFU gave a P value <0.0001 and therefore is significantly different at $P < 0.05$.

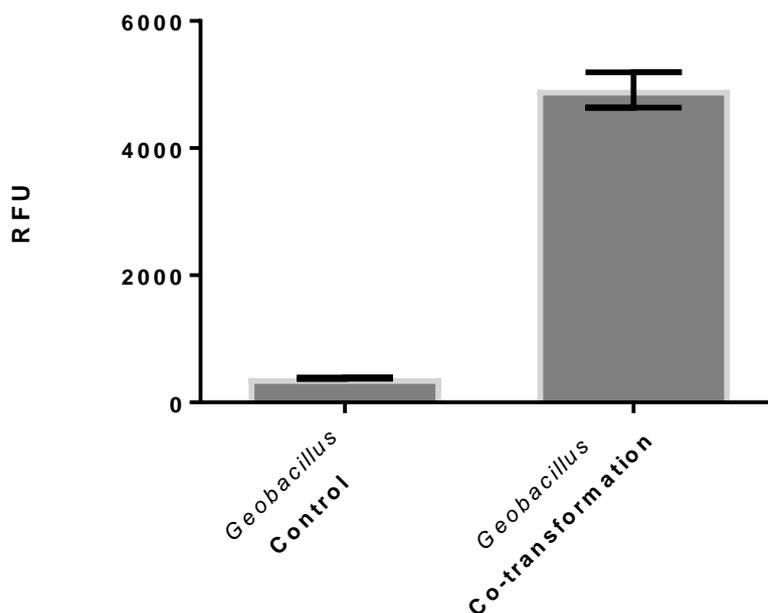


Figure 3.22. sGFP assay of co-transformed *G. thermoglucosidasius* culture.

The sGFP assay was carried out as described previously. Confirmed positive co-transformed cultures were compared with un-transformed *G. thermoglucosidasius* wild type cultures. The assay was carried out in triplicate with average and standard error shown. The difference in RFU gave a P value of 0.0012 and therefore is significantly different at $P < 0.05$.

These results show effective co-transformation and subsequent gene expression of reporters from two plasmids introduced to both *E.coli* and *Geobacillus*. *E.coli* can be transformed in one-step procedure. For *Geobacillus*, plasmids must be transformed sequentially.

This is, to our knowledge, the first demonstration of heterologous gene expression from co-transformation of multiple plasmids in *Geobacillus*. This system can be used as a tool to express genes of interest.

3.8 Mega plasmid free strain

G. thermoglucosidasius NCIMB 11955 harbours two megaplasmids; pNCI001 (83,925 bp, 43.5% GC) and pNCI002 (47,893 bp, 39.0% GC) (Sheng et al., 2016).

There are 81 coding sequences on pNCI001, and 55 on pNCI002 (Sheng et al., 2016). This puts a burden on cellular processes. Therefore generation of a megaplasmid free strain is desirable, in order to evaluate if the reduced burden creates a more efficient host strain for product production.

Attempts to knock out the plasmids were initially carried out by introducing a second plasmid containing the same replicon, pMTL64320 and pMTL65320. The introduced plasmids also harboured the spectinomycin resistance gene. It was thought that growing the transformed *Geobacillus* with antibiotic selection would cure the cell of the megaplasmid, with the plasmid conferring antibiotic resistance remaining. This was found not to be the case. *Geobacillus* transformed with pMTL65320, produced initial colonies following transformation, however they were unstable with none able to survive re-streak or growth in liquid culture. The *Geobacillus* cells transformed with pMTL64320 were able to survive in the presence of antibiotic, however the megaplasmids also remained stable in the cell.

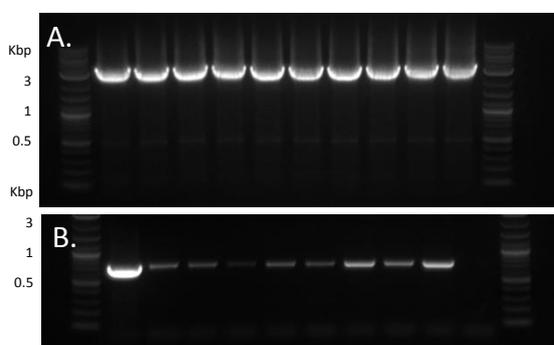


Figure 3.23. Visualisation of PCR showing both plasmid and megaplasmid presence in *G. thermoglucosidasius*. PCR of 10x *G. thermoglucosidasius* colonies following transformation with pMTL64320. Each colony amplified both A. 3433 bp region on the pMTL64320 plasmid and B. 750 bp region of the pNCI001 megaplasmid. PCR was carried out as previously described using primers A. ColE1_F2 and Rep01_R and B. ToxF and ToxR. Annealing temperatures of A. 57°C and B. 65°C and extension times of A. 3 minutes and

B. 1 minute were used. PCR was conducted after two re-streaks to enable any outcompeting of the plasmids over generations.

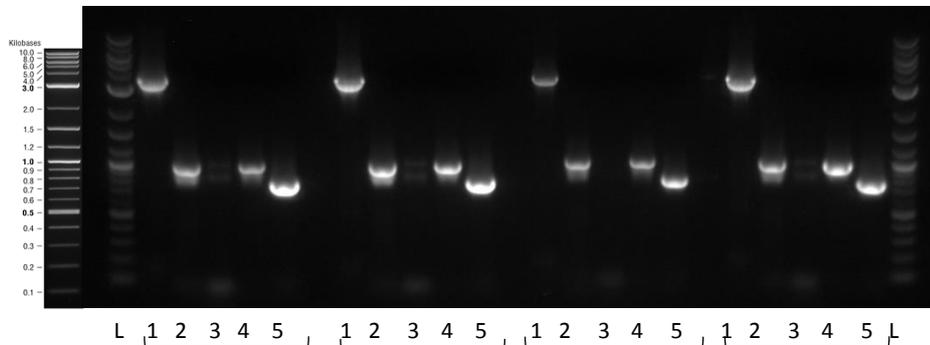


Figure 3.24. PCR confirmation showing transformation of pMTL64330 did not result in loss of pNCI001. To further confirm this result 4 colonies were screened for 5 regions across the plasmids; 1 and 2 on pMTL64330 and 3, 4, 5 on pNCI001. Primers used; 1. ColE1_F2 and Rep01_R = 3433bp, 2. ColE1_F2 and M13F = 1kb, Primers in Megaplasmid 01; 3. F1 and R1 = 1kb, 4. F2 and R2 = 1019 bp, 5. F3 and R3 = 985 bp. PCR was carried out as previously described with annealing temperature of 60°C and extension time of 1 minute. PCR was conducted after two re-streaks to enable outcompeting of plasmid over generations. All regions amplified for all four of the colonies, indicating both plasmids are complete within the cells screened here.

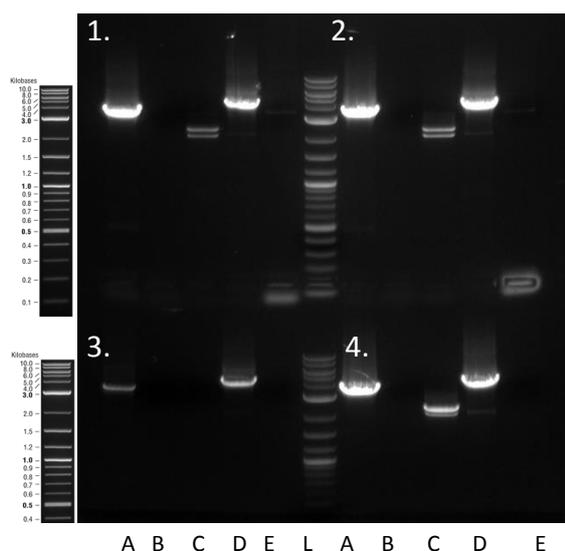


Figure 3.25. Recombination analysis. Following previous results suggesting presence of both plasmids pMTL64320 and NCI001 in *G. thermoglucosidasius*, further PCR analysis was carried out on the transformed strain. Using one primer on the megaplasmid and one on the introduced plasmid to determine if the two plasmids were discrete or if recombination had occurred. PCR was carried out as previously described. Four colonies were tested with five primer sets. Primers used; A. MegaF and ColE1_F2, B. MegaF and ColE1_R, C. MegaR and ColE1_F2, D. MegaR and ColE1_R, E. MegaF and MegaR. An annealing temperature of 58°C and extension time of 3 minutes were used. Amplification indicates recombination has occurred. Bands 3A, 3D, 4A and 4D were sent for sequencing. The sequencing results confirmed recombination of the two plasmids pMTL64320 and NCI001 had occurred.

Investigations found the two plasmids; pMTL64320 and NCI001 recombined to horizontally transfer the antibiotic resistance gene, along with the rest of the pMTL plasmid, into the megaplasmid. The replicon provides an extensive homologous region where recombination can occur. This suggests *G. thermoglucosidasius* has efficient recombination machinery.

After initial attempts to cure *G. thermoglucosidasius* of its megaplastids using incompatibility of the replicons proved unsuccessful, attempts were made to knock out the megaplastids using the ACE knock out method (work carried out by Habgood, R). This also proved unsuccessful. Finally development of a CRISPR-Cas9 deletion method was used to remove the megaplastids (work carried out by Lau, M. Unpublished method, not detailed here). Single loss strains and a double loss strain were created. I was then able to demonstrate transformation of pMTL64327 and pMTL65327 into the respective megaplastid defective strains.

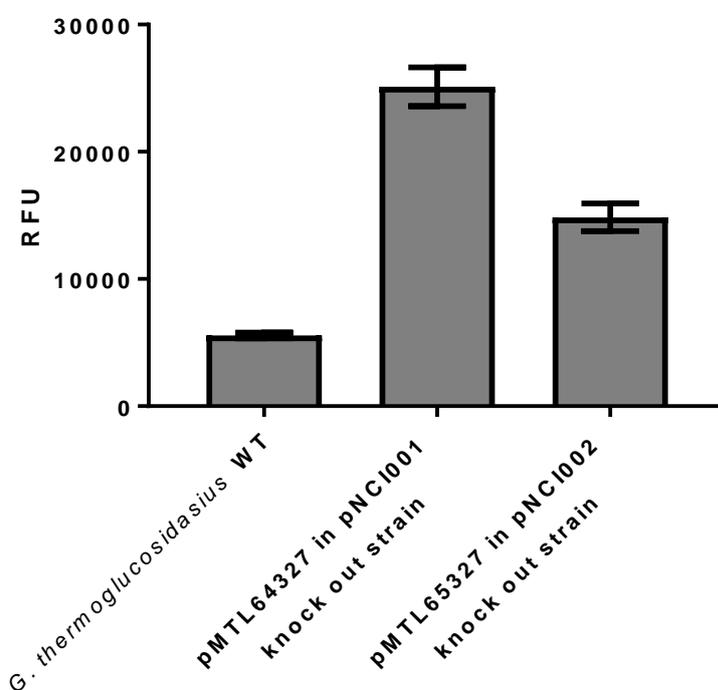


Figure 3.26. Assay showing plasmid based GFP expression in megaplastid knock out strains. pMTL64327 and pMTL65327 plasmids, containing the eCGP123 fluorescent protein, were transformed into the respective megaplastid knock out strains. The GFP assay was carried out as described previously, with RFU normalised to OD₆₀₀ 1. 10 colonies of each strain were screened with average and standard deviation shown here. This demonstrates use of the NCI001 and NCI002 replicons in the two *Geobacillus* megaplastid knock out strains.

Attempts to work with the double loss mutant were unsuccessful as the strain exhibited poor growth. This, in addition to the difficulties encountered when curing *G. thermoglucosidasius* of pNCI001 and pNCI002, suggest pNCI002 encodes a gene essential to *G. thermoglucosidasius* survival. This also explains the high stability of the megaplasmid. After searching the genome sequences, a toxin/antitoxin system was identified in pNCI001, ensuring its maintenance in the cell (further discussed in 3.9).

Further work is needed to understand the genes present on both megaplasmids, and the benefit provided to the cell. If the essential components of the megaplasmids are confirmed, then the unessential components could be removed, reducing the cellular burden. Such a strain would be an attractive host strain for industrial use. As the cell will have less burden, cellular resources can be diverted to increasing production of the compound of interest.

Alternatively, the megaplasmids in current or reduced size, offer an attractive site for gene integration. As an alternative to chromosomal integration. Integrating genes into the megaplasmids would provide the advantages of chromosomal integration, being naturally highly stable and requiring no antibiotics for maintenance. The megaplasmids also have a higher copy number than chromosome, offering the benefit of higher expression levels. The recombination observed could be utilised for this purpose.

3.9 Toxin/Antitoxin module

G. thermoglucosidasius megaplasmid pNCI001 encodes a toxin/antitoxin (TA) system. The toxin gene; death on curing protein (Doc) and its antitoxin; prevent host death protein (Phd) maintain stability of the plasmid by post-segregational killing of plasmid-free daughter cells. Thus imparting a survival advantage to the host cell. When the Doc toxin is active the protein binds to the 30S ribosomal subunit preventing translation (Unterholzner et al., 2013). The antitoxin binds forming a protein-protein complex with the toxin resulting in its

neutralization. As the toxin is more stable than the antitoxin, if the plasmid is lost the antitoxin will be rapidly degraded no longer counteracting the toxin. Consequently, the toxin becomes activated and can act on its cellular targets (Unterholzner et al., 2013).

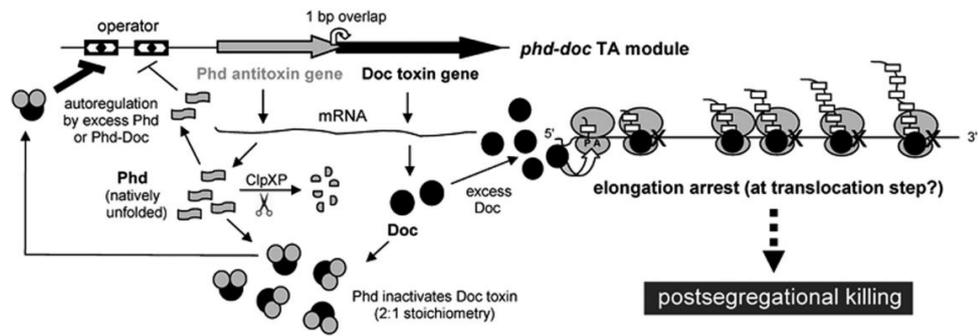


Figure 3.27. Depiction of Phd-Doc function resulting in postsegregational killing (Liu et al., 2008). Toxicity occurs when free Doc toxin is able to arrest translation elongation by binding to the 30S subunit, leading to bacterial cell death.

TA systems and their components can be utilised as tools for biotechnology and molecular biology. Various systems are frequently used for selection in cloning.

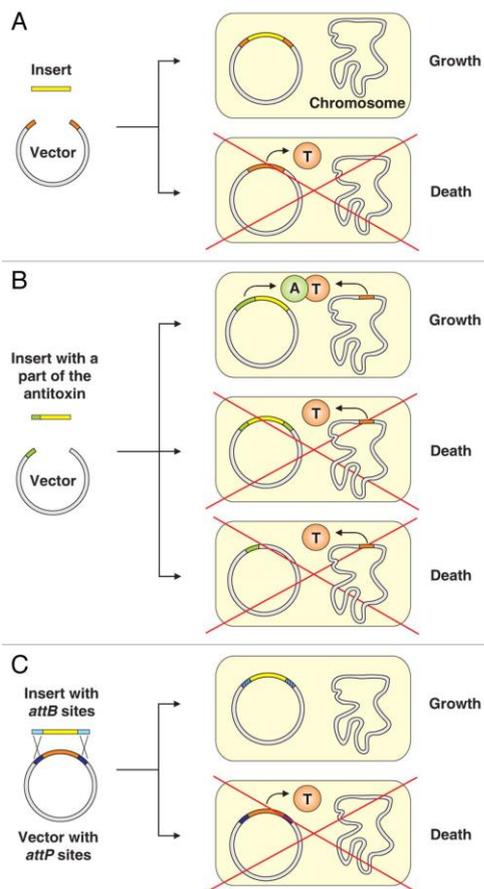


Figure 3.28. Application of TA systems for DNA cloning (Unterholzner et al., 2013). A. Insertion of a gene of interest into a toxin gene inactivates the toxin gene and allows the bacterium to grow, enabling positive selection of inserts. B. Principle of the StabyCloning system. C. Principle of the selection used in the Gateway cloning system.

TA systems can be used as an alternative to antibiotic selection for maintaining a plasmid in culture. For enhanced plasmid stability the toxin gene can be integrated into the genome and the antitoxin gene placed on the expression vector. This separated component stabilization ensures maintenance of the plasmid, increasing the heterologous gene expression levels and recombinant protein yields.

The Phd-Doc TA system identified here offers various approaches which could be used to develop molecular tools for *Geobacillus*. The system is

demonstrated to be compatible and functional in *Geobacillus* therefore the proteins must be thermostable. The high stability of pNCI001 demonstrates efficacy of the TA system, as no other genes on the plasmid appear to be essential.

The pNCI001 knock out strain generated, as described previously, could be used as a host to integrate the *doc* toxin gene into the genome. With the antitoxin present on an expression plasmid. This would enable highly stable plasmid based gene expression, without the need for antibiotic supplemented media.

3.10 Discussion

G. thermoglucosidasius is a promising organism for industrial application. Development to date has been hindered by a lack of genetic tools. The work presented here has increased the tools and parts available, with the aim of facilitating future strain engineering. Much further work is needed to enable ease of genetic modification. As a novel organism many approaches are yet to be explored.

Initial work to develop reporter assays was first needed for the quantification and characterisation of parts and systems. In this work four reporter assays are used, three of which can be used simultaneously. This provides for extensive analysis of a variety of gene expression within a cell. Qualitative and quantitative assays were developed. A novel highly thermotolerant fluorescent protein is presented. For use in *Geobacillus* robust reporter assays are required. The protein must be temperature and pH stable. This limits the applications of many reporters. A potential luminescent reporter protein for future development is NanoLuc (Promega). NanoLuc is a luciferase reported to be stable at 55°C (Hall et al., 2012). This could be a potential candidate for an additional thermophilic reporter.

The control of gene expression is necessary for optimization of metabolic pathways. In order to manipulate gene expression in a microbial host, characterised promoters should be available. Here a range of promoters were screened. Although the strength of a range of promoters has been determined, to date only two promoters driving strong expression in *Geobacillus* have been identified. Error prone PCR is commonly used to generate component libraries however this approach may not be feasible for use with *Geobacillus* as similar sequences recombine readily and so promoters with distinct sequences are needed.

Inducible promoters are available as an alternative to the constitutive promoters tested here. However for industrial application constitutive expression is required. Induction would add an additional step to the production process and would require additional reagents or fermentation supplementation. An alternative approach is the use of a temperature sensitive promoter. Novotny et al. (2008) found the *sgsE* gene coding for a surface layer protein in *Geobacillus stearothermophilus* was strongly induced when the culture was shifted from 55°C to 67°C. The group found that the P_{sgsE} promoter had very low basal activity at 28°C, with intermediate and high levels at 37°C and 45°C, respectively. This *sgsE* promoter could provide the potential to adjust expression levels of genes by temperature adjustments.

Use of an orthogonal expression system was investigated. The use of the TcdB/TcdR system to drive gene expression demonstrated little advantage. It was thought the use of dedicated machinery solely for the purpose of specific gene expression may have resulted in higher levels of expression, however this was not found. Alternative orthogonal expression systems offer great potential for synthetic biology systems and should be investigated further. Techniques such as logic functions, feed-forward and feed-back loops can be used to provide decision-making processes and tailored dynamics in gene expression (An & Chin, 2009).

The pMTL60000 vector series was recently introduced by Sheng et al. (2017). Here new component parts extending the modular vector system are presented. Providing multiple options for each part enabled the demonstration of co-transformation of two plasmids into *G. thermoglucosidasius*.

Identification and characterisation of Gram-positive replicons resulted in identification of GEOTH02, a high copy number and highly stable replicon which can be used on a gene expression plasmid. Also BST1, a high copy number and highly unstable replicon which can be used on a delivery vector. Other replicons which could be investigated include; pGTD7 and pGTG5, two cryptic low molecular weight, RC plasmids isolated from *Geobacillus* spp. 610 and 1121 respectively. Both strains were identified by Kananavičiūtė et al. (2014). Plasmid pGTG5 at 1.5 kbp is, to date, the smallest plasmid isolated from *Geobacillus* bacteria. *Geobacillus stearothermophilus* strain TK05 harbours a plasmid pSTK1. Plasmids from other genus of thermophilic bacteria could also be investigated.

Horizontal gene transfer between different species has been recognized as a common and major evolutionary process. Here all replicons tested were functional in both *G. thermoglucosidasius* and *B. subtilis*. Some broad host range plasmids can transfer across bacterial phyla and even across domains of life (Heinemann & Sprague, 1989; Waters, 2001). Klümper et al. (2014) showed replicons may not be designated Gram-positive or negative. Their results indicated that maintenance and transfer is possible across the Gram border. If a universal replicon, functional in both *E. coli* and *Geobacillus* was identified, plasmids could be streamlined requiring less components and reducing metabolic burden on the cells. It may be the case plasmids are able to cross the Gram border to deliver genes but not replicate. The transient presence of a plasmid can provide the host with additional genes. Accessory genes on plasmids are mostly arranged in transposons flanked by insertion sequence (IS)

elements, which can recombine with the recipient bacterial chromosomes delivering DNA without the need for plasmid replication.

The marker genes available for use with *Geobacillus* are; kanamycin and spectinomycin resistance genes. All of which confer resistance to an aminoglycoside antibiotic. All marker genes available rely on the same mechanism of aminoglycoside modification for resistance. Alternative marker systems with differing resistance mechanisms could be explored to increase the diversity of options available.

The two megaplasmids present in *G. thermoglucosidasius* provide an array of options for modification and utilisation. Here they are knocked out, resulting in a reduced cellular burden. The replicons are used on other plasmids. The megaplasmids also offer a stable location for gene integration, with a higher copy number than the chromosome.

The Phd-Doc toxin/antitoxin system identified on pNCI001 is likely responsible for the highly stable maintenance of this plasmid. This mechanism could be utilised for other purposes. For example as an alternative selection mechanism, ensuring plasmid maintenance without the requirement of antibiotic supplementation in the growth media.

In addition to the work presented here, there are an array of additional tools which can be developed to aid engineering of *Geobacillus*. For example the CRISPR-Cas9 system was briefly used in this work (work done by Lau, M.). The development of this precise, targeted genome editing method offers a powerful tool for genetic modification.

With the tools developed to date and the vast array of further advancements in genetic engineering methodologies on the horizon, there is much potential

for future development to exploit the potential of *Geobacillus* and other thermophilic organisms.

CHAPTER 4

CoA DEPENDENT *n*-BUTANOL
PRODUCTION

4.1 Introduction

Finite fossil fuel resources and the negative environmental effects of fossil fuel usage has resulted in the need for alternative, renewable energy sources such as biofuels. Biofuel development to date has focused on bio-ethanol, here *n*-butanol is the target molecule. *n*-butanol is desirable for its superior fuel properties. The thermophilic bacterium; *G. thermoglucosidasius* has been selected as the host organism. Various advantages are associated with use of a thermophilic production organism. *G. thermoglucosidasius* was considered a promising organism for industrial application. The molecular tools developed, as described in the previous chapter, will be utilised in the following work in attempt to engineer the *G. thermoglucosidasius* host strain to produce *n*-butanol.

To date, *n*-butanol production in a non-native, thermophilic and aerobic organism, has not been demonstrated. Natural *n*-butanol producers, such as *C. acetobutylicum*, *C. beijerinckii* and *C. pasteurianum* are exclusively strict anaerobes and are most commonly mesophilic bacteria. In order to develop thermophilic and oxygen tolerant *n*-butanol production, such a process must address growth and metabolite production in an aerobic and thermophilic environment.

4.2 A synthetic *n*-butanol metabolic pathway

The aim of this work is to produce *n*-butanol using *Geobacillus* as the production strain. As *n*-butanol is not a product native to *Geobacillus*, the first approach considered was replication of the natural production process. Acetone, butanol and ethanol (ABE) fermentation is a natural biosynthetic process, native to *Clostridium* strains. ABE fermentation by *C. acetobutylicum* is a well-established production, being one of the largest industrial fermentation processes in the early twentieth century (Liao et al., 2016). By replicating this known functional pathway, the reactions required for *n*-butanol production have previously been characterised, albeit in different organisms.

The introduction of a heterologous *n*-butanol pathway based on clostridial genes will be introduced into *Geobacillus*. The introduction of genes encoding enzymes in the *n*-butanol biosynthetic pathway should result in a recombinant bacterium capable of producing *n*-butanol. This approach has been adopted previously, with successful introduction of the ABE pathway into a variety of alternative organisms, described in section 1.5.4.2. Most commonly in the well characterised *E. coli* and *S. cerevisiae*. The use of alternative hosts aims to increase yields and improve processing conditions.

The native solvent-producing clostridia use the coenzyme A (CoA)-dependent pathway to produce *n*-butanol. The biosynthesis of acetone, butanol and ethanol share the same metabolic pathway from glucose to acetyl-CoA, branching into different pathways thereafter, as shown in Figure 4.1. A typical ABE fermentation using *C. acetobutylicum* yields acetone, butanol and ethanol in the ratio of 3:6:1. The *n*-butanol pathway starts with the condensation of two acetyl-CoA to form acetoacetyl-CoA catalysed by thiolase. Acetoacetyl-CoA then undergoes a series of dehydration and reduction steps to form butyryl-CoA. These reaction steps are essentially the reversal of β -oxidation used in fatty acid decomposition. The reduction of crotonyl-CoA to butyryl-CoA is catalysed by the butyryl-CoA dehydrogenase (Bcd) and electron transferring flavoprotein (Etf) complex. This protein complex is an electron-bifurcating enzyme that couples the exergonic reduction of crotonyl-CoA to butyryl-CoA to the endergonic reduction of ferredoxin both with NADH (Li et al., 2008). Butyryl-CoA is reduced to butyraldehyde, then to *n*-butanol using a bifunctional alcohol/aldehyde dehydrogenase. Butyryl-CoA can also be hydrolysed to butyrate, with concomitant generation of ATP. The butyrate pathway is essential for the energy metabolism of clostridia. Native clostridia generate acids in the growth phase, for energetic advantage. The acids are re-assimilated to form the corresponding acyl-CoA and then the alcohols when NADH is abundant.

CoA DEPENDENT n-BUTANOL PRODUCTION

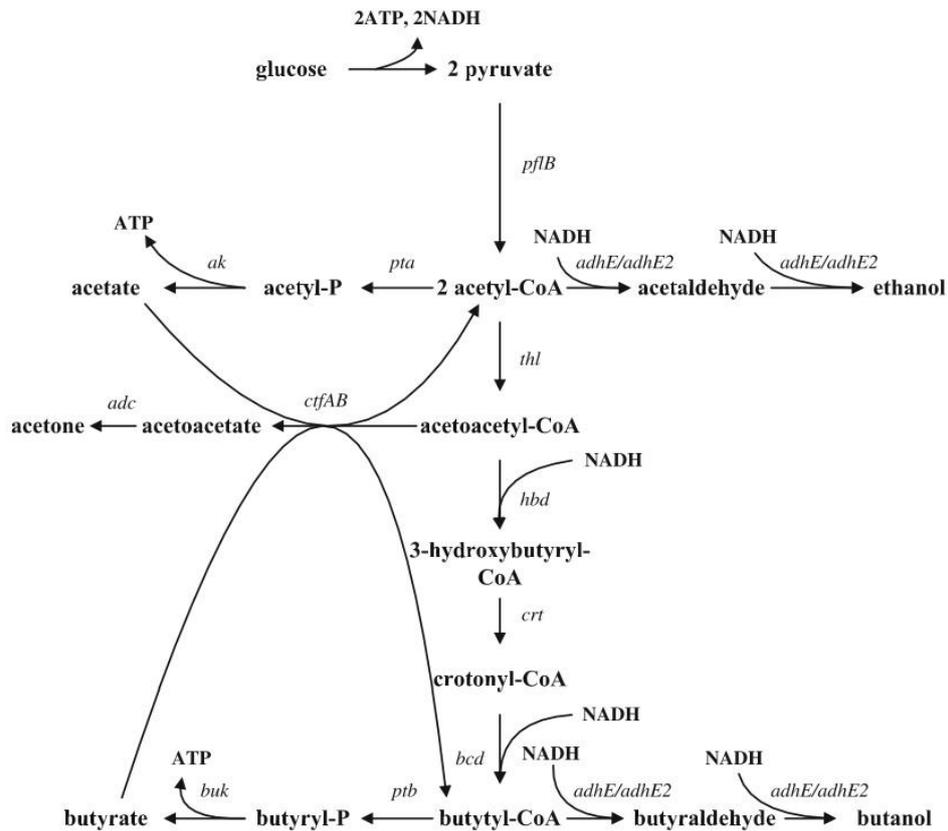


Figure 4.1. Diagram showing the ABE fermentation pathway reactions as present in *Clostridia*.

4.2.1 A thermophilic gene source

Thermoanaerobacterium thermosaccharolyticum, previously known as *Clostridium thermosaccharolyticum*, is a natural thermophilic butanol producer. Wild type *T. thermosaccharolyticum* has an *n*-butanol titre of 1.8 mM (Bhandiwad et al., 2013). However, *T. thermosaccharolyticum* is not an ideal process organism due to low butanol tolerance. In addition, the organism is a strict anaerobe making strain handling and genetic manipulation difficult. *T. thermosaccharolyticum* provides a thermophilic source of the *n*-butanol pathway genes. When heterologously expressed in a thermophilic host strain, the genes should be stable. Protein folding and enzyme activity should be compatible with high growth temperatures, such as the *Geobacillus* optimal growth temperature.

In *T. thermosaccharolyticum* the genes; enoyl-CoA hydratase (*crt*), acyl-CoA dehydrogenase (*bcd*), electron transfer flavoprotein alpha/beta subunits (*etfA/B*), 3-hydroxyacyl-CoA dehydrogenase (*hbd*) and acetyl-CoA acetyltransferase (*thl*) are present in the bacterial cellulose biosynthesis (*bcs*) operon, for formation of butyryl CoA (Bhandiwad et al., 2013). Six alcohol dehydrogenase genes have been identified on the *T. thermosaccharolyticum* genome, however only one has sequence similarity to the conserved domains of the bi-functional aldehyde-alcohol dehydrogenase thought to be used for catabolic reduction of acetyl-CoA to ethanol (Bhandiwad et al., 2013). This bi-functional *adhE* gene is also assumed to be responsible for butyryl-CoA reduction to *n*-butanol (Bhandiwad et al., 2013).

4.2.2 Selection of heterologous genes

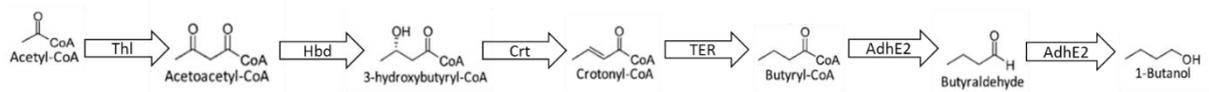


Figure 4.2. The metabolic ABE pathway for the production of *n*-butanol.

The metabolic pathway based on the ABE fermentation route, as natively found in clostridia, requires six reactions for the conversion of acetyl-CoA to the final product *n*-butanol, shown in Figure 4.2. It was hypothesised that the introduction and successful expression of compatible genes encoding all enzymes in the *n*-butanol biosynthetic pathway would result in a recombinant bacterial strain capable of producing *n*-butanol.

Table 4.1. The five enzymes selected for conversion of acetyl-CoA to *n*-butanol.

Enzyme commission number	Gene	Enzyme	Gene Source	Size (bp)	Protein Size (Kda)	Reaction
EC 2.3.1.9	<i>thl</i>	acetoacetyl-CoA thiolase/ acetyltransferase	<i>T. thermo-saccharolyticum</i>	1223	45	$2 \text{ acetyl-CoA} \rightleftharpoons \text{CoA} + \text{acetoacetyl-CoA}$
EC 1.1.1.157	<i>hbd</i>	3-hydroxybutyryl-CoA dehydrogenase	<i>T. thermo-saccharolyticum</i>	914	33	$3\text{-acetoacetyl-CoA} + \text{NADPH} + \text{H}^+ \rightleftharpoons 3\text{-hydroxybutanoyl-CoA} + \text{NADP}$
EC 4.2.1.55	<i>crt</i>	crotonase	<i>T. thermo-saccharolyticum</i>	824	30	$3\text{-hydroxybutanoyl-CoA} \rightleftharpoons \text{crotonyl-CoA} + \text{H}_2\text{O}$
EC 1.3.1.44	<i>ter</i> ₇₃₃₄	<i>trans</i> -2-enoyl-CoA reductase	<i>Treponema caldaria</i> DSM 7334	1271	47.8	$\text{acyl-CoA} + \text{NAD}^+ = \text{trans-2,3-didehydroacyl-CoA} + \text{NADH} + \text{H}^+$ (acts on crotonyl-CoA)
EC 1.3.1.44	<i>ter</i> ₆₁₉₂	<i>trans</i> -2-enoyl-CoA reductase	<i>Spirochaeta thermophila</i> DSM 6192	1223	46	$\text{acyl-CoA} + \text{NAD}^+ = \text{trans-2,3-didehydroacyl-CoA} + \text{NADH} + \text{H}^+$ (acts on crotonyl-CoA)
EC 1.1.1.1	<i>adhE</i>	Bi-functional aldehyde/alcohol dehydrogenase	<i>T. thermo-saccharolyticum</i>	2624	97.4	$\text{An alcohol} + \text{NAD}^+ \rightleftharpoons \text{an aldehyde} + \text{NADH} + \text{H}^+ / \text{an aldehyde} + \text{NAD}^+ + \text{H}_2\text{O} \rightleftharpoons \text{an acid} + \text{NADH} + \text{H}^+$

The enzymes presented in Table 4.1 were selected for use in this work. The thiolase enzyme (Thl) converts two units of acetyl-CoA to acetoacetyl CoA and CoA. The thiolase superfamily enzymes catalyse the carbon–carbon bond formation via a thioester-dependent Claisen condensation reaction mechanism. 3-hydroxybutyryl CoA dehydrogenase (Hbd) is an oxidoreductase catalysing the reduction of 3-acetoacetyl-CoA to 3-hydroxybutanoyl-CoA requiring NADPH. Crotonase (Crt) is an enoyl coenzyme A hydratase catalysing the reaction 3-hydroxybutanoyl-CoA to crotonyl-CoA and H₂O.

Trans-2-enoyl-reductase (Ter)

An enzyme in the clostridial *n*-butanol pathway is the butyryl-CoA dehydrogenase electron transferring protein complex. The reaction from crotonyl-CoA to butyryl-CoA is catalysed by butyryl-CoA dehydrogenase (Bcd). Bcd requires an additional enzyme complex, Etf, to couple ferredoxin reduction with NADH oxidation (Atsumi et al., 2008). In addition Bcd is reported to be oxygen sensitive (Chen & Liao, 2016), deeming the gene incompatible with the aerobic host strain used here.

To avoid complications involved with ferredoxin reduction and oxygen sensitivity, the Bcd-Etf complex is replaced with trans-2-enoyl-reductase (Ter). Ter couples the same reaction directly with NAD(P)H. This approach should provide an alternative route replacing use of Bcd. This strategy has been demonstrated in both *E. coli* (Shen et al., 2011) and *Synechococcus elongates* (Lan & Liao 2011). Ter directly utilizes NADH as the reducing co-factor for the reduction of crotonyl-CoA to butyryl-CoA with a large negative change in free energy, making it irreversible and thus serving as an effective driving force for the pathway. The use of an NADH-dependent enzyme also improves the overall redox balance of the pathway. NADH can be regenerated through numerous metabolic reactions.

Here two Ter homologues from thermophilic sources are used; *Spirochaeta thermophila* DSM 6192 and *Treponema caldaria* DSM 7334. Both Ter enzymes have previously shown expression in *G. thermoglucosidasius* (Sheng, 2014), demonstrating compatibility for the host.

Aldehyde/alcohol dehydrogenase (AdhE)

The dehydrogenase enzyme (AdhE) is reported to be a bifunctional protein with an N-terminal acetaldehyde-CoA dehydrogenase domain and a C-terminal alcohol dehydrogenase domain. The AdhE gene catalyses the two final steps in the pathway converting butyryl-CoA, through butyraldehyde intermediate to *n*-butanol. This reaction is coupled to the oxidation of two NADH molecules to maintain the NAD(+) pool during fermentative metabolism (Extance et al., 2013).

C. acetobutylicum possesses two homologous *adhE* genes, *adhE1* and *adhE2*, responsible for butanol production in solventogenic and acidogenic/alcohologenic cultures, respectively (Yoo et al., 2016). Solventogenic cultures produce; acetone, butanol, and ethanol when grown at low pH with glucose. Acidogenic cultures produce; acetate and butyrate when grown at neutral pH with glucose. Alcohologenic cultures form butanol and ethanol when grown at neutral pH under conditions of high NAD(P)H availability (Yoo et al., 2016).

There is only one potentially bifunctional *adhE* gene identified in the *T. thermosaccharolyticum* genome. This enzyme is yet to be fully characterised. Therefore its suitability and properties in differing culture conditions for optimal *n*-butanol production are yet to be elucidated.

4.2.3 Codon optimisation

Effective and accurate expression of heterologous genes requires codon optimisation specific to the host. As the genetic code is made up of four

different bases and three of these bases make up a codon, the maximum codon variation is 64. This far exceeds the twenty naturally occurring amino acids.

It is thought that variation in codon usage may provide evolutionary advantages such as compensation of point mutations. Codon usage may also be driven by natural selection for translation accuracy and efficiency and can therefore impact expression levels. This has resulted in a wide spectrum of codon bias across different species. In thermophilic bacteria, high temperature may drive codon selection. Bias within thermophilic prokaryotes was found to differ from mesophiles (Lynn et al., 2002). This could be the result of mRNA stability or the thermodynamic properties of base pairings.

Here, the synthetic genes were codon optimised for *G. thermoglucosidasius* based on the codon usage table in Appendix A8.

4.2.4 Pathway construction and expression

To build the metabolic pathway, construction of multiple genes into an operon, was carried out. The genes listed in Table 4.1 were synthesised with codon optimisation to reflect the codon bias of *G. thermoglucosidasius*. The BioBrick prefix was added to the 5' end and a BioBrick suffix was added to the 3' end. The synthetic RBS sequence was added. Two variants were synthesised; with and without a flag-tag. The Biobrick cloning method was used to assemble the genes into operons. Expression of the operon was driven by the strong P_{ldh} . Operon assembly was carried out in the Biobrick compatible cloning vector pJ201 before being transferred into the expression vector pMTL61122. The vectors were then transformed into wild type *G. thermoglucosidasius*, for plasmid based expression of the operon.

CoA DEPENDENT n-BUTANOL PRODUCTION

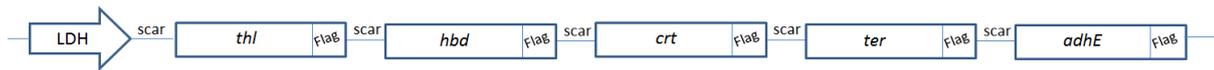


Figure 4.3. The desired operon construction for the *n*-butanol pathway. The gene order replicates that of the native *bcs* operon and also the sequential order in which the intermediate compounds are produced and subsequently converted towards the final product. Here the Flag-tagged variants are shown. An 8 bp scar between each fragment results from the Biobrick assembly methodology used.

Difficulties in cloning the operon were encountered. In order to determine if the genes were toxic to *E. coli* cells, each was cloned individually into reporter plasmid pTMO372. All genes were successfully cloned and transformed into both *E. coli* and *G. thermoglucosidasius*. Correct gene sequences from recovered plasmid were confirmed by Sanger sequencing. This showed, individually, the genes are not toxic to either *E. coli* or *G. thermoglucosidasius*.

The *G. thermoglucosidasius* cells harbouring the individual genes were then used to assay gene expression. Western blot was carried out to detect the flag tagged proteins in order to determine expression of the genes.

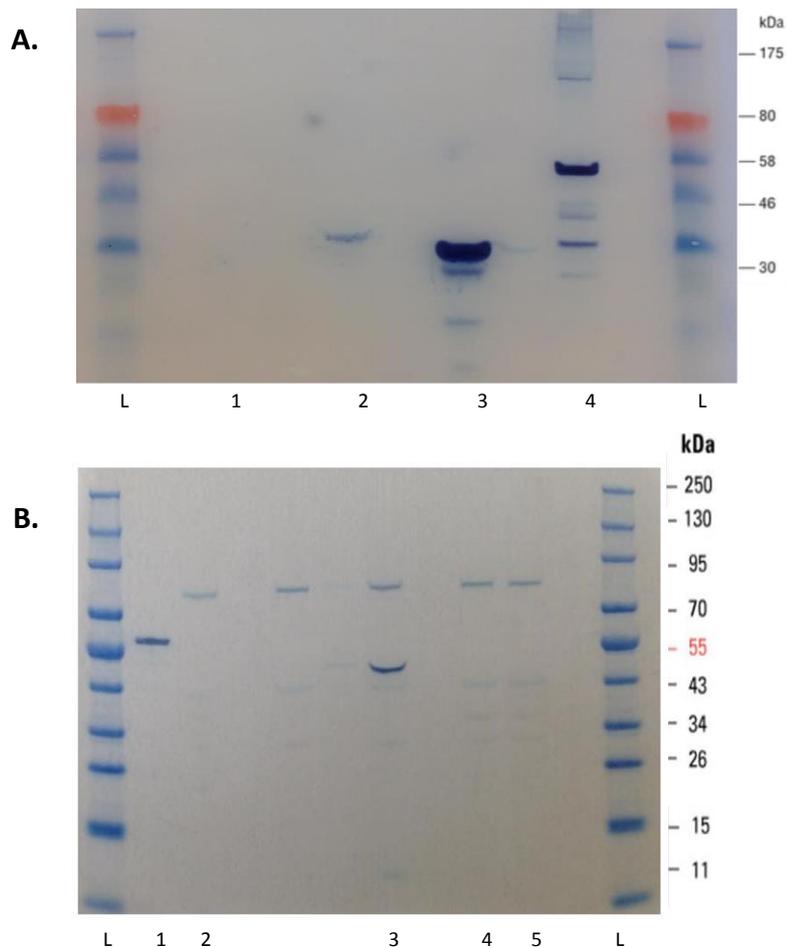


Figure 4.4. Western blot of *Geobacillus* expressing synthetic *n*-butanol pathway genes. *G. thermoglucosidasius* harbouring pTMO372 with the genes as indicated. Samples were prepared for Western blot as described in section 2.7.2. **A.** L. PageRuler prestained protein ladder (Thermo Fisher); 1. Wild type *G. thermoglucosidasius* negative control; 2. Hbd (33 KDa); 3. Crt (30 KDa); 4. Ter (48 KDa). **B.** L. PageRuler prestained NIR protein ladder (Thermo Fisher); 1. Positive control protein; 2. Wild type *G. thermoglucosidasius* negative control; 3. Thl (45 KDa); 4 and 5. AdhE (97.4 KDa).

The protein visualisation by Western blotting shown in Figure 4.4 confirms plasmid based expression of genes; *hbd*, *crt*, *ter* and *thl* in *Geobacillus*. Analysis of the soluble fraction showed expression, with the correct sized bands visible in lanes 2, 3 and 4 Figure 4.4 A. and lane 3 of Figure 4.4 B. This result indicates

these enzymes are being expressed and are functional in *Geobacillus*. No expression of the *adhE* gene was detected. This suggests the gene is not compatible with a *Geobacillus* host. This enzyme may be problematic as it is a larger protein, expression of which will introduce a greater burden on the host cells. The protein may not fold into its functional form in the conditions used here.

Additional bands in lanes 3 and 4 of Figure 4.4 A. show break down products of Crt and Ter. The smaller bands, less than 30 and 47.8 KDa respectively, suggest the proteins are degraded with the partial proteins detected here. In lane 4 of Figure 4.4 A., larger bands greater than the protein size of the Ter gene, are present. These additional bands are approximately three and four times the size of the protein and could show proteins aggregating into complexes. In Figure 4.4 B. background bands are present in all *Geobacillus* samples. This non-specific binding could be the result of sequences similar to the flag tag sequence, to which the antibodies are able to bind. SDS presence can cause non-specific antibody binding. Optimisation of the protocol to improve; gel washing, blocking and antibody concentration could minimise this in future experiments.

Following observation of expression four of the required *n*-butanol pathway genes in *Geobacillus*, an operon consolidating the genes into one plasmid was constructed, as shown in Figure 4.5. This operon was transformed into *Geobacillus* and assayed for gene expression.



Figure 4.5. Diagram of a four gene operon, a partial *n*-butanol pathway. The Biobrick cloning method was used to assemble four flag tagged genes; *thl*, *hbd*, *crt* and *ter* into an operon. With P_{ldh} cloned in front of the *n*-butanol genes to drive operon expression. Biobrick assembly methodology was used, resulting

in an 8bp scar between each fragment. Assembly was carried out in the Biobrick compatible cloning vector pJ201 before being transferred into the expression vector pMTL61122. The vector was then transformed into wild type *G. thermoglucosidasius*, for plasmid based expression of the operon.

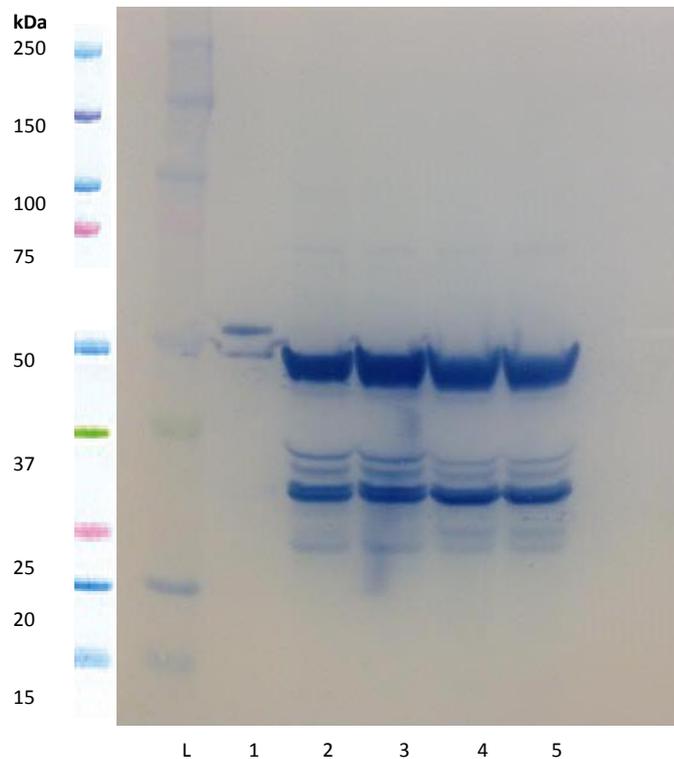


Figure 4.6. Western blot of *Geobacillus* expressing an operon of four synthetic *n*-butanol pathway genes. *G. thermoglucosidasius* harbouring pMTL61122 with the four genes; *thl*, *hbd*, *crt* and *ter* in an operon driven by P_{ldh} . Samples were prepared for Western blot as described in section 2.7.2. L. Precision Plus Kaleidoscope protein ladder (Bio-Rad); 1. Positive control protein; 2, 3, 4, 5. Replicates of *G. thermoglucosidasius* with pMTL61122 + Thl (45 KDa), Hbd (33 KDa), Crt (30 KDa) and Ter (48 KDa).

Four genes each with a flag-tag, were expressed as an operon in *Geobacillus*. The result of the Western blot expression assay are shown in Figure 4.6. The blot shows a large band of approximately 48 kDa in size. This could show

expression of both Thl and Ter, which are 45 and 48 kDa, respectively. As the two proteins are of a similar size the bands may have merged to form one larger band here. The two remaining proteins Hbd and Crt can be seen here as bands of 33 and 30 kDa size, respectively. Although these proteins are also of a similar size, separation of the bands can be seen due to apparent weaker expression levels. In addition better resolution of the lower molecular weight proteins is expected. Various weak bands are visible on the blot from approximately 25 to 36 KDa. This could be due to background noise from insufficient blocking or degradation of the proteins. Strong bands of the expected size for each of the four genes in the operon, confirms expression of the operon in *G. thermoglucosidasius*.

The operon, shown in Figure 4.5, was constructed and expressed without the *adhE* gene. This forms only a partial *n*-butanol pathway. Alcohol/aldehyde dehydrogenase (AdhE) enzymes are found in many organisms. AdhE catalyses the conversion of an acyl-CoA to an alcohol via an aldehyde intermediate. Wild type *Geobacillus* are native ethanol producers. In order to produce ethanol, it is hypothesised, *Geobacillus* will possess at least one functional *adhE* gene. If a native alcohol/aldehyde dehydrogenase accepts butyryl-CoA, in addition to acetyl-CoA, as a substrate this four gene operon generating butyryl-CoA will be sufficient to introduce *n*-butanol production.

4.2.5 Utilisation of a modified host strain

In addition to testing the synthetic *n*-butanol operon in the wild type NCIMB 11955 strain, strain LS242 (Δdh , pdh^{up} , Δpfl) (Sheng et al., 2017) was also used as a potential production strain. *G. thermoglucosidasius* natively produces a mixture of lactate, acetate, formate and ethanol (Cripps et al., 2009). In this work, the production of a single, high yielding product is desired. Deletion of competing pathways could increase the carbon flux towards *n*-butanol.

Modifications made to strain LS242; the *ldh* gene encoding L-lactate dehydrogenase is deleted to reduce lactic acid production. Deletion of the gene *pfl* encoding pyruvate formate lyase should prevent formate production. Upregulation of the gene encoding pyruvate dehydrogenase using the strong P_{ldh} from *G. stearothermophilus*. Deletion of *ldh* and *pfl* results in reduction of by-product biosynthesis. Increased pyruvate levels are available to channel into the *n*-butanol biosynthetic pathway. Re-directing carbon to the desired product. Pyruvate dehydrogenase (encoded by the *pdh* gene) provides a route to acetyl-CoA under aerobic conditions. The Pdh route provides reducing equivalents to balance the production of *n*-butanol product from glucose. To increase the availability of acetyl-CoA and NADH made through Pdh, the encoding gene is upregulated. The native promoter is replaced with a strong promoter. This modification should also prevent pyruvate accumulation following deletion of the *ldh* pathway. *ter* integration was used for greater expression levels, as will be shown subsequently in section 4.3.2.

4.2.6 Culture for *n*-butanol production

In order to determine if the addition of the synthetic *n*-butanol pathway genes resulted in *n*-butanol production, various *G. thermoglucosidasius* strains were cultured for product analysis. The screened strains are listed in Table 4.2. Each host strain with and without plasmid pMTL61122 + *thl-hbd-crt-ter* or pMTL61122 + *thl-hbd-crt*, as indicated. Host strain modifications were generated by Sheng, L (work not described here).

Table 4.2 Strains screened for *n*-butanol production.

	<i>G. thermoglucosidasius</i> host strain modifications	Plasmid
A	NCIMB 11955 wild type	None
B	NCIMB 11955 wild type	pMTL61122 + <i>thl-hbd-crt-ter</i>
C	+ <i>ter</i> Integration	None
D	+ <i>ter</i> Integration	pMTL61122 + <i>thl-hbd-crt</i>
E	+ <i>ter</i> Integration + Δ <i>ldh</i>	None
F	+ <i>ter</i> Integration + Δ <i>ldh</i>	pMTL61122 + <i>thl-hbd-crt</i>
G	+ <i>ter</i> Integration + Δ <i>ldh</i> + Δ <i>pfl</i>	None
H	+ <i>ter</i> Integration + Δ <i>ldh</i> + Δ <i>pfl</i>	pMTL61122 + <i>thl-hbd-crt</i>
I	+ <i>ter</i> Integration + Δ <i>ldh</i> + Δ <i>pfl</i> + <i>pdh</i> ^{UP}	None
J	+ <i>ter</i> Integration + Δ <i>ldh</i> + Δ <i>pfl</i> + <i>pdh</i> ^{UP}	pMTL61122 + <i>thl-hbd-crt</i>

The strains in Table 4.2 were cultured in aerobic, micro aerobic and fully anaerobic conditions. Cultures were incubated at 52°C and 60°C with shaking at 250 rpm for 24 hours. Growth media; AYSE, UYSE and M9 minimal medium with 20 g/l glucose, were screened. Samples were analysed by GC-FID analysis for *n*-butanol production.

CoA DEPENDENT n-BUTANOL PRODUCTION

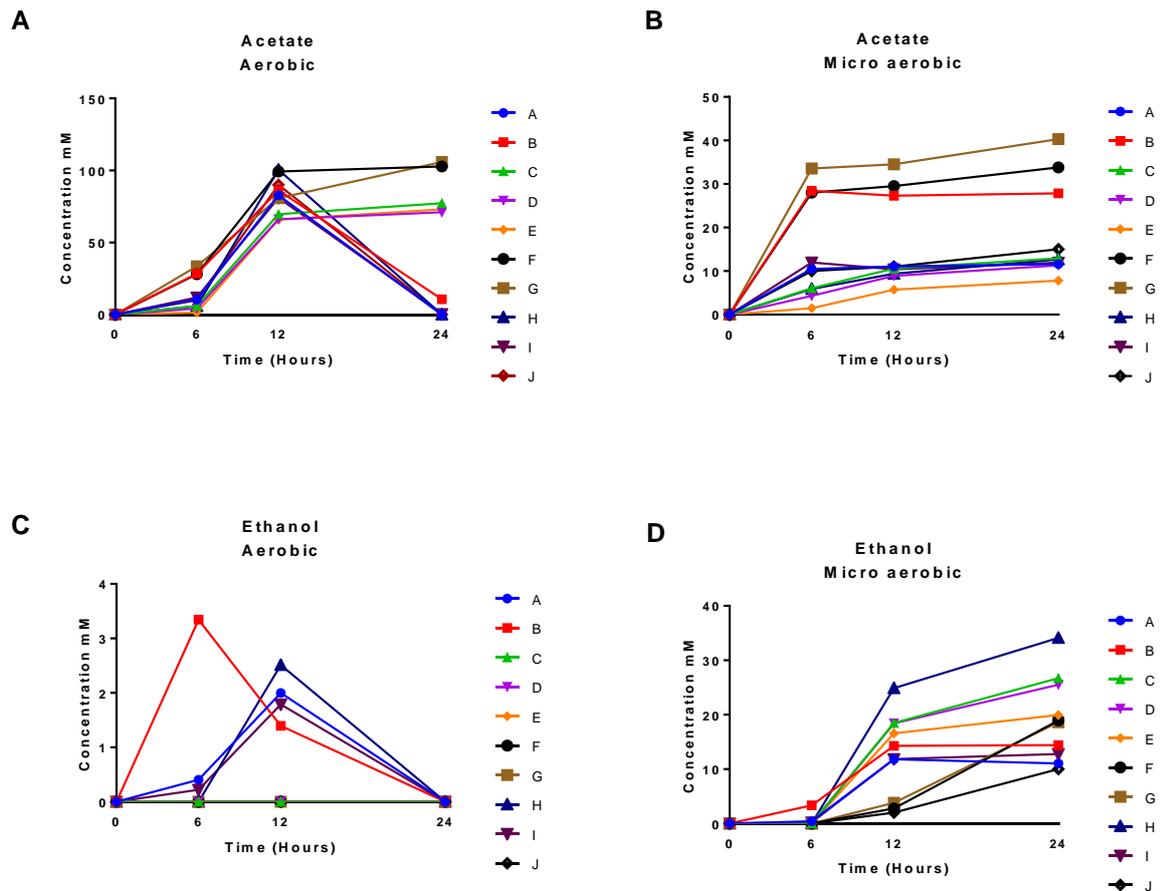


Figure 4.7. Acetate and ethanol production of strains cultured in aerobic and micro aerobic conditions. Data here shows the strains, genotype detailed in Table 4.2, cultured for 24 hours at 52°C in UYSE medium with 20 g/l glucose. Samples were analysed by GC-FID, with acetate and ethanol concentrations presented.

Results of this experiment showed no *n*-butanol production in any of the strains screened. Analysis of the cultures grown in minimal medium showed low OD (data not presented). This indicates some yeast extract is required for growth of *G. thermoglucosidasius*. The cultures grown in fully anaerobic environment were not able to grow. This confirmed *Geobacillus* requires oxygen for growth and suggests *G. thermoglucosidasius* is an obligate aerobe. Of the cultures grown in AYSE or UYSE in aerobic and micro aerobic conditions, Figure 4.7 shows production of acetate and ethanol. No *n*-butanol was detected.

Troubleshooting

In attempt to understand the cause of a lack of *n*-butanol production resulting from the previous screening experiment, aspects of the strains and culture methodology were investigated further.

ter integration

First confirmation of *ter* gene integration was sought. 3 µl samples of culture after 72 hours incubation were subject to PCR amplification of the genomic integration site, to determine presence of the integrated *ter* gene.

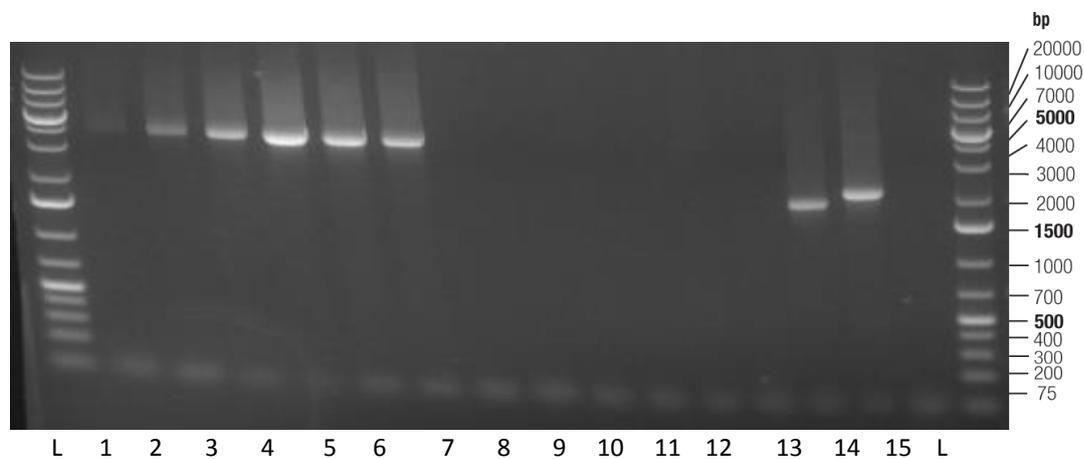


Figure 4.8. Visualisation of PCR amplification to confirm *ter* genomic integration. PCR protocol as previously described, with 55°C annealing temperature and 3 minute extension time. Primers detailed in appendix A1. L indicates GeneRuler 1 kb Plus DNA ladder. 1. C. + *ter* Integration, 2. E. + *ter* Integration + Δ *ldh*, 3. G. + *ter* Integration + Δ *ldh* + Δ *pfl*, 4. I. + *ter* Integration + Δ *ldh* + Δ *pfl* + *pdh*^{UP} cultures grown in AYSE medium aerobically at 52°C. 5. C., 6. E., 7. G., 8. I. cultures grown in UYSE medium micro aerobically at 52°C. 9. C., 10. E., 11. G., 12. I. cultures grown in UYSE medium micro aerobically at 60°C. 13. *pyrE* mutant control, 14. A. *G. thermoglucosidasius* wild type, 15. negative control. Correct *ter* integration resulting in a band size of 3500 bp. *pyrE* mutant with no gene integration results in an 1876 bp band and wild type *G. thermoglucosidasius* results in a 2101 bp band.

The PCR analysis of the cultured strains indicates the genomic integration of the *ter* gene may not be stable. After culture of the strains for 72 hours some of the strains have lost the gene in certain conditions, such as reduced oxygen and higher temperature. However further investigation is required in order to confirm this. As there is no amplification seen in lanes 7 to 12 of Figure 4.8, this indicates the whole *pyrE* region is lost or has been re-arranged. As the primers flanking the *pyrE* gene region gave no amplification. If the integration had not been present or a mixed culture, bands equivalent to wild type of the *pyrE* mutant, as shown in lanes 13 and 14, would be amplified.

To determine if temperature is a determining factor of the integrational stability of *ter*, strains were re-streaked on TSA three times with incubation at 60°C and subject to colony PCR for amplification of the genomic integration locus.

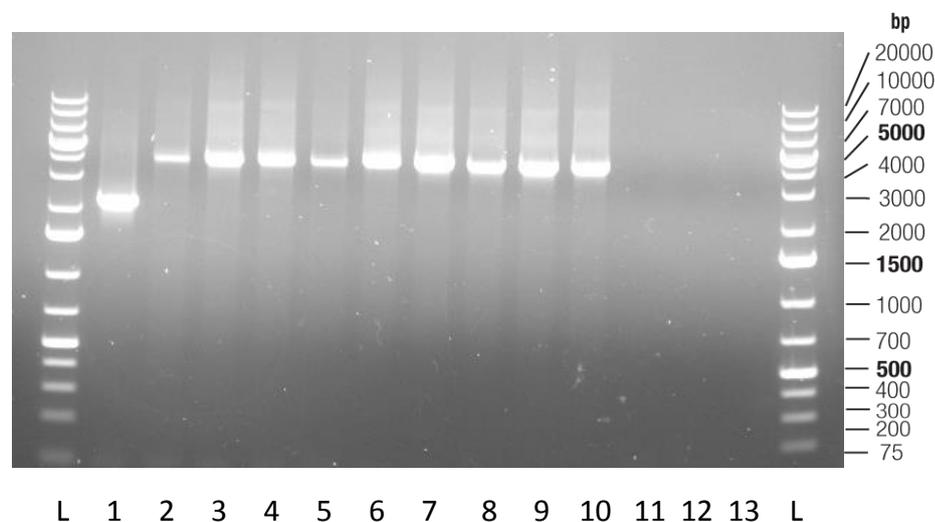


Figure 4.9. Visualisation of PCR amplification to confirm *ter* genomic integration stability at 60°C. PCR protocol as previously described, with 55°C annealing temperature and 3 minute extension time. Primers detailed in appendix A1. L indicates GeneRuler 1 kb Plus DNA ladder. 1. *ter* 3 re-streaks followed by colony PCR. 1. Wild type *G. thermoglucosidasius* 2101 bp band, 2. to 5. colonies of strains C., G., E. and I. of Table 4.2, after 1st re-streak, 6. to 9.

colonies of strains C., G., E. and I. of Table 4.2, after 2nd re-streak, 10. to 13. colonies of strains C., G., E. and I. of Table 4.2, after 3rd re-streak.

The colony PCR results shown in Figure 4.9 shows only the *ter* integration strain C. in Table 4.2, is detected after 3 re-streaks and incubation at 60°C. This instability of the gene integration could be due to repetition of the P_{ldh} promoter. Here up to four copies of the P_{ldh} promoter are present in the cell; its native location, driving the *ter* integration, up regulation of the *pdh* gene and on the pMTL61122 plasmid driving operon expression. These regions of homology could enable loss of the genes by recombination.

Butanol evaporation

As *n*-butanol is a volatile compound with a low boiling point, it was considered any butanol produced could be lost before detection as a result of the incubation conditions used.

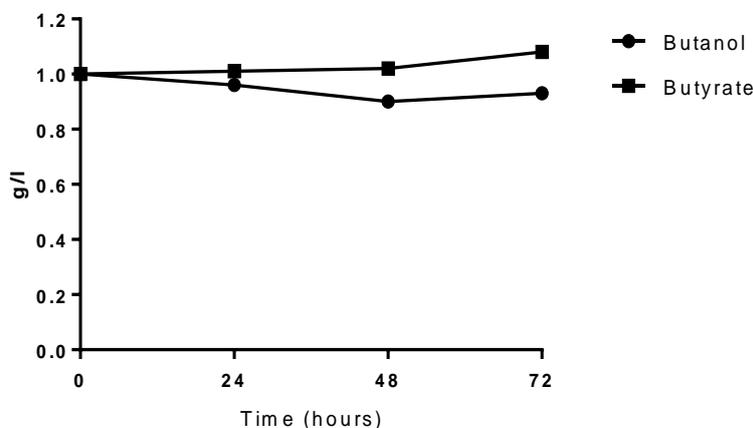


Figure 4.10. Butanol and butyrate evaporation. Concentrations of butanol and butyrate were quantified by HPLC sample analysis. 50 ml water inoculated to 1 g/l was incubated at 52°C with shaking at 250 rpm for 72 hours. Samples were taken for HPLC analysis at 24 hour intervals.

After 72 hours incubation in conditions mimicking *Geobacillus* culture conditions, 0.93 g/l of *n*-butanol was detected by HPLC from a starting concentration of 1 g/l. This is a 7.39% loss of *n*-butanol. No loss of butyrate was seen. This result confirms any butanol or butyrate produced will be detected and not lost during the culture process despite the incubation temperature and aeration at 250 rpm.

4.2.7 Discussion

Initial introduction of synthetic *n*-butanol pathway genes has yet to yield any *n*-butanol. In the initial cultures a synthetic *adhE* gene was not introduced. The synthetic *adhE* gene identified was not shown to express by Western blot, in addition difficulties due to plasmid recombination meant an operon containing all five of the synthetic genes was not constructed and therefore could not be included in the screening. An alternative, functional, *adhE* gene is required which is compatible with the host. Observations indicating the integration of the *ter* gene into the *G. thermoglucosidasius* genome is not stable, presents further challenges. Further work is needed to determine if the genes are functional or to identify an alternative approach.

4.3 Upregulation of *Geobacillus* native genes for *n*-butanol production

Problems associated with the compatibility and stability of heterologous genes could be overcome by overexpression of native genes. Genes native to the host are inherently compatible and the proteins thermostable and oxygen tolerant.

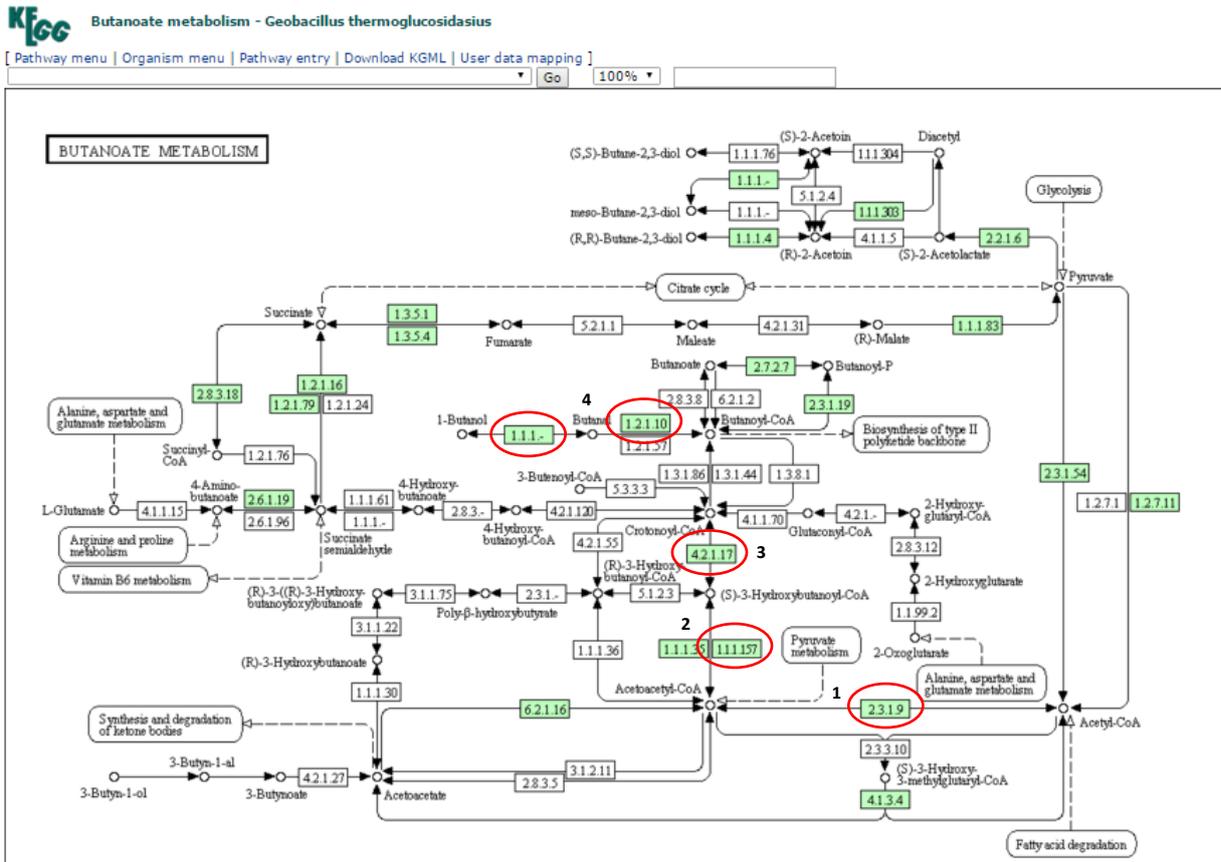


Figure 4.11. Pathway diagram showing butanoate metabolism in *G. thermoglucosidarius*. Taken from the KEGG pathway database. Genes present in the genome are shown in green. The genes relevant for *n*-butanol production are circled in red. **1.** acetyl-CoA C-acetyltransferase [EC:2.3.1.9], **2.** 3-hydroxybutyryl-CoA dehydrogenase [EC:1.1.1.157], **3.** enoyl-CoA hydratase [EC:4.2.1.17], **4.** bifunctional acetaldehyde dehydrogenase / alcohol dehydrogenase [EC:1.2.1.10/1.1.1.1].

Four of the five genes required for a full *n*-butanol production pathway were identified to be present in the *G. thermoglucosidasius* genome, as highlighted in Figure 4.11. Although *G. thermoglucosidasius* is not a native butanol producer, the genes are present as the reactions are required for a range of other pathways in the cell. The primary function of the genes identified here is as part of the beta (β) oxidation pathway for fatty acid degradation, forming an essential function in the cell's basic metabolism. The ubiquitous nature of β -oxidation, aldehyde/alcohol dehydrogenase and thioesterase enzymes has the potential to enable the efficient synthesis of these products in other industrial organisms.

4.3.1 Overexpression of native genes

Here four native genes, catalysing the reactions identified in Figure 4.11, were cloned into a plasmid based operon for overexpression in *G. thermoglucosidasius*.

Of the four reactions identified in Figure 4.11 multiple gene options were available in the *G. thermoglucosidasius* genome. (1. *thl* x2 genes, 2. *hbd* x4 genes plus one 3-hydroxyacyl-CoA dehydrogenase, 3. *crt* x1 gene, 4. *adhE* x1 acetaldehyde/alcohol dehydrogenase plus two acetaldehyde dehydrogenases). Where multiple options were available the identity of each was compared to the genes found in two *n*-butanol producing organisms; *C. acetobutylicum* and *T. thermosaccharolyticum*. In each case the closest match was selected for overexpression. The similarity to butanol producing organisms was used as the criteria for selection as this was considered to improve the potential utility of the gene for this desired purpose.

Table 4.3. Blastp (protein-protein BLAST) comparison of protein to protein sequences of the native genes compared to the genome of two *n*-butanol producing organisms.

Enzyme commission number	Enzyme	<i>Clostridia acetobutylicum</i>			<i>Thermoanaerobacterium thermosaccharolyticum</i>		
		Query cover	E value	Identity	Query cover	E value	Identity
EC:2.3.1.9	Thl	99%	4e-131	51%	99%	2e-135	52%
EC:1.1.1.157	Hbd	98%	2e-120	58%	98%	2e-120	58%
EC:4.2.1.17	Crt	96%	2e-63	40%	98%	1e-72	41%
EC:1.2.1.10/ 1.1.1.1	AdhE	98%	0.0	52%	98%	0.0	70%

Table 4.3 shows the similarity of the native genes selected to both *C. acetobutylicum* and *T. thermosaccharolyticum*. The Blastp (protein-protein BLAST) algorithm was used to search non-redundant protein sequences in the organism *C. acetobutylicum* or *T. thermosaccharolyticum*, as indicated. Query cover shows the percentage of the query sequence aligned. E value shows the number of matches with same score expected by chance and therefore the probability of a random alignment. $E < 0.05$ is considered significant. Identity shows the extent to which nucleotide or protein sequences are related, with similarity between the two sequences expressed as percent sequence identity.

The genes selected for overexpression:

thl; encoding acetyl-CoA acetyltransferase from the thiolase N superfamily. Catalyses the carbon-carbon-bond formation required for fatty acid and polyketide synthesis, via a thioester-dependent Claisen condensation reaction mechanism.

hbd; codes for 3-hydroxybutyryl-CoA dehydrogenase. An NAD binding fatty acid oxidation protein involved in the degradation of long-chain fatty acids via the β -oxidation cycle.

crt; encoding crotonase/enoyl-CoA hydratase from the crotonase-like superfamily. Involved in lipid transport and metabolism.

adhE; codes for a bifunctional acetaldehyde-CoA/alcohol dehydrogenase from the PRK13805 superfamily.

Following selection of the genes, AlignX software (Invitrogen) based on the Clustal W algorithm (Thompson et al., 1994), was used to compare the amino acid sequences of the native genes to be overexpressed with the corresponding synthetic gene sequence. This was done to ensure the two gene sequences were distinct. The results of the alignments showed the two sequences were distinct with no large regions of homology, greater than 14 bases, in any case (data not shown).

The selected genes were amplified from *G. thermoglucosidasius* wild type genomic DNA using primers found in appendix A1. NEB Builder HiFi DNA assembly was used to construct two plasmids; pMTL61222 + Native *thl*, *hbd* and *crt* and pMTL61222+ *ter* and Native *AdhE*, as shown in Figure 4.13. The vector backbone was linearised by restriction enzyme digestion. The synthetic RBS sequence and restriction enzyme sites were built into the spacer regions. Restriction enzyme sites were included for possible future alterations or sub-cloning of the genes using restriction enzyme cloning.

CoA DEPENDENT n-BUTANOL PRODUCTION

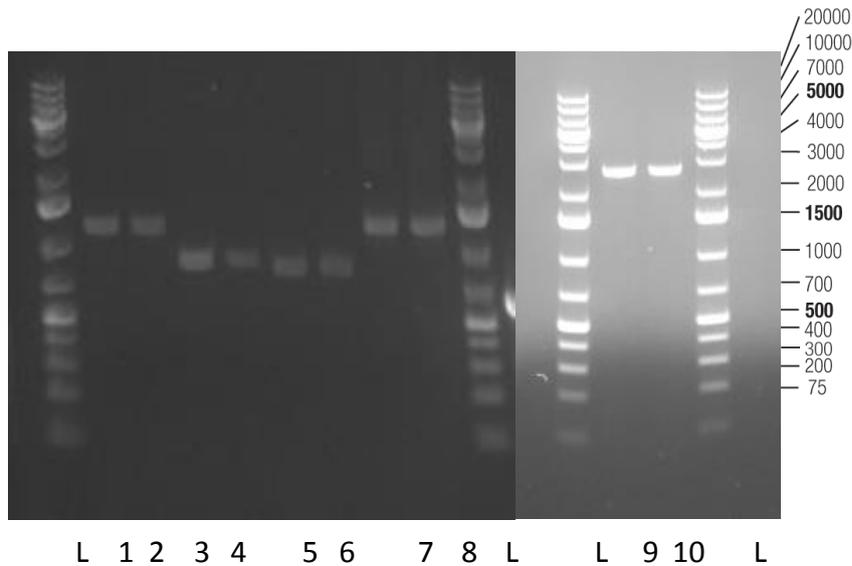


Figure 4.12. Visualisation of native gene amplification from genomic DNA.

PCR was carried out, as previously described, using *G. thermoglucosidasius* WT DNA as the template. Primers detailed in appendix A1 were used. The non-native *ter* gene was amplified from a plasmid source to complete the pathway. L; 1kb Plus GeneRuler DNA Ladder (Thermo Scientific), 1 and 2; native *thl* 1185 bp, 3 and 4; native *hbd* 855 bp, 5 and 6; native *crt* 777 bp, 7 and 8; *ter6192* 1188 bp, 9 and 10; native *adhE* 2604 bp.

The DNA fragments visualised in Figure 4.12 were then excised from the gel and the DNA purified before being assembled using HiFi DNA assembly method, and transformed into *E. coli*, as described previously. Correct assembly of the plasmids was confirmed by colony PCR and restriction enzyme digestion of recovered plasmid. The plasmids were then transformed into *G. thermoglucosidasius*, as described previously. With PCR confirmation of gene presence in *Geobacillus*.

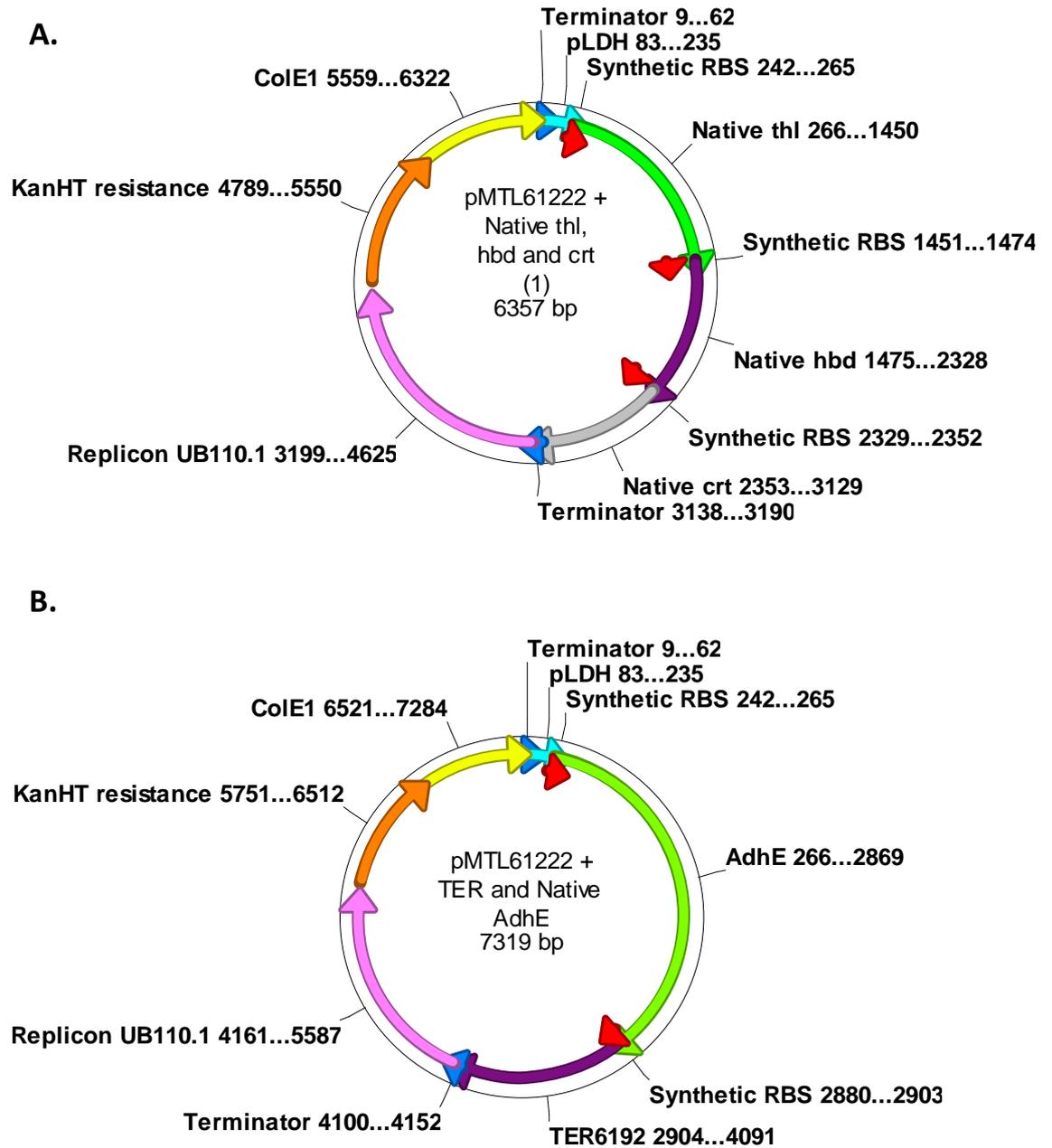


Figure 4.13. Plasmid maps of the pMTL61222 vector containing native genes.
A. pMTL61222 with native *thl*, *hbd* and *crt* genes. B. pMTL61222 with *ter6192* and native *adhE*. The two/three gene operons are driven by P_{ldh} and include the synthetic RBS to initiate protein translation.

Following construction of the two plasmids, further cloning steps were carried out in attempt to generate a single vector containing all five genes. Initially the

Biobrick cloning method was used to insert a fragment containing *ldh*, *ter* and native *adhE* 5' of the native *crt* gene shown in Figure 4.13 A. After attempts using the Biobrick methodology failed a PCR method was also attempted. This again proved unsuccessful. The problems encountered could be due to the large size of the final vector (10.4 kbp) or the repeat sequences present in the vector. Two P_{ldh} promoters and 5 synthetic RBS sequences are present in the final construct. This gives two homologous regions of 141 bp and five homologous regions of 18 bp, respectively. Such regions of homology enable recombination of the plasmid. Any recombination results in loss of the pathway genes or other required features of the plasmid.

4.3.2 Enzyme activity assays

In order to determine heterologous expression and activity of the *n*-butanol pathway genes, enzyme assays were carried out. The enzyme assays were used to confirm functional activity in the *G. thermoglucosidasius* host. Demonstration of enzyme activity suggests protein translation and correct folding of the proteins. Thus confirming compatibility of the enzymes in the host. Both native and synthetic enzyme activities were assayed for comparison of the genes.

CoA DEPENDENT n-BUTANOL PRODUCTION

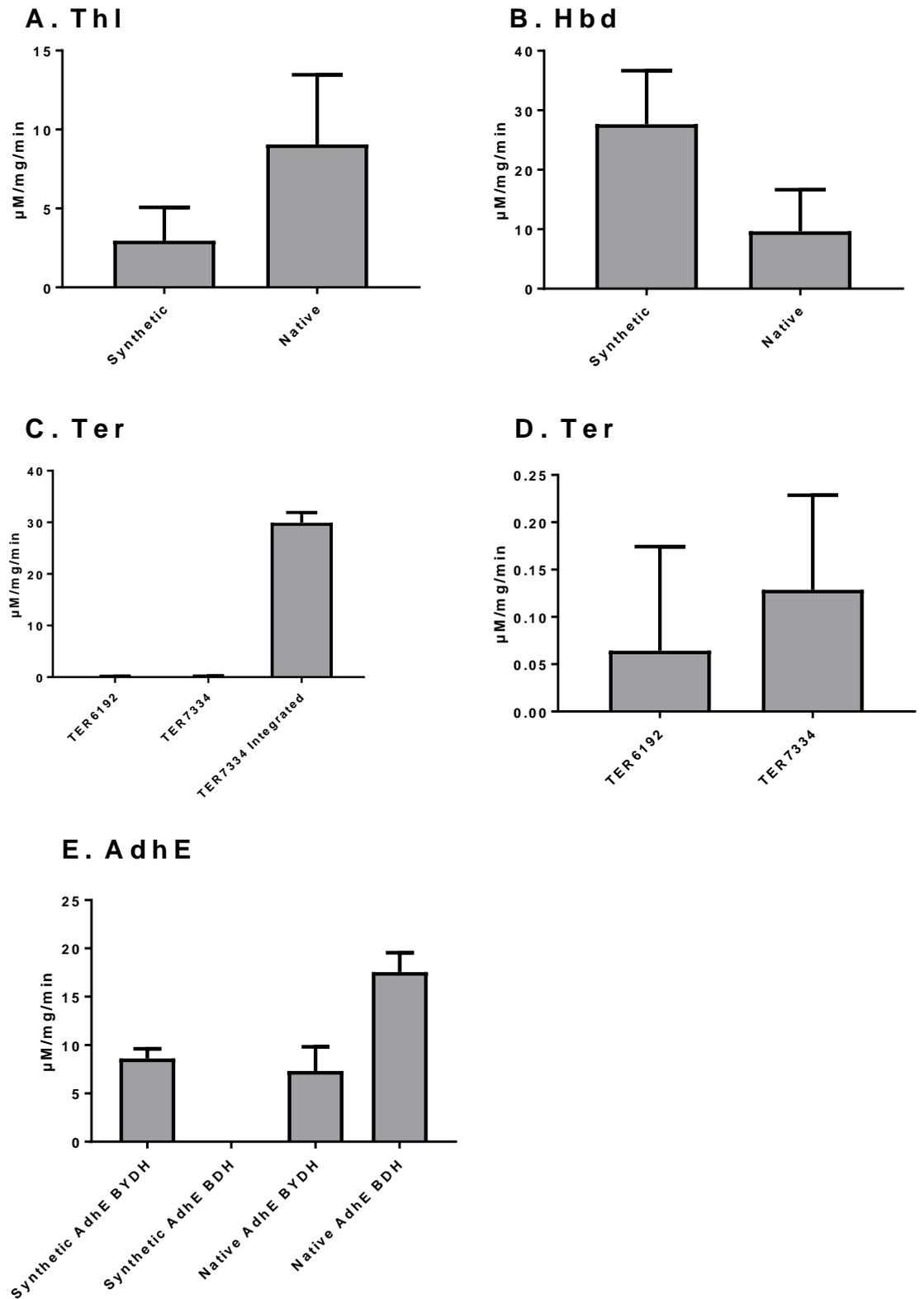


Figure 4.14. Enzyme activities of synthetic and native *n*-butanol pathway genes. Cultures of *G. thermoglucosidasius* containing the gene on plasmid,

CoA DEPENDENT n-BUTANOL PRODUCTION

unless indicated integrated, were cultured in 2SPYNG prior to cell lysis and protein quantification. Wild type *G. thermoglucosidasius* culture was included as a negative control. All assays were performed using a SpectraMax M3 microplate reader (Molecular Devices) with the reaction monitored over one minute. Gradient of the Slope, $R^2 > 0.99$ in each case. One unit of enzyme activity was defined as the amount of enzyme catalyzing the cleavage of 1 μmol of substrate per minute ($\mu\text{M}/\text{mg}/\text{min}$). Average and standard deviation of triplicate data is shown. The data was analysed by two tailed t test to determine significance. The result was deemed significant when $p < 0.05$. **A.** Thiolase synthetic and native gene activity was determined by monitoring the disappearance of acetoacetyl-CoA by decrease in absorbance at 303nm. The molar absorption coefficient used was 16.9 mM^{-1} . This result is significant as $p = 0.0028$. **B.** *hbd* synthetic and native gene activity was determined by monitoring the decrease of absorption at 340 nm, corresponding to consumption of NADH. This result is significant as $p = 0.000697$. **C.** TER activity was determined by monitoring the forward direction reaction by a decrease in NADH measured at 340 nm with crotonyl-CoA as the substrate. Two versions of the gene were assayed from plasmid based expression and one gene was integrated into the host strain genome. **D.** The activity of the two plasmid based *ter* genes. The p-value is 0.2716. The result is not significant at $p < 0.05$. **E.** Synthetic and native AdhE activity was determined by monitoring the decrease of absorbance at 340 nm corresponding to the consumption of NADH or NADPH. Butyryl-CoA was the substrate for the butyraldehyde dehydrogenase reaction (BYDH) and butyraldehyde was the substrate for the butanol dehydrogenase reaction (BDH). The molar extinction coefficient for consumption of NADH at 340 nm was taken as $6,220 \text{ M}^{-1}\text{cm}^{-1}$.

The enzyme activity results, as shown in Figure 4.14 A., demonstrate the native *thl* gene results in greater Thl activity than the corresponding synthetic gene. Figure 4.14 B. shows cells containing the synthetic *hbd* gene produce higher activity than cells carrying the native gene. The two *ter* gene variants tested,

both synthetic, showed similar activity with no significant difference between the two (Figure 4.14 D.). However the activity greatly increased when the *ter* gene was integrated into the genome, as shown in Figure 4.14 C. The AdhE assay monitored both aldehyde and alcohol dehydrogenase activity. The results, shown in Figure 4.14 E. show that lysates of cells carrying the synthetic *adhE* gene are only capable of using butyryl-CoA as a substrate and exhibit no alcohol dehydrogenase activity. The synthetic *adhE* gene was reported to encode a protein with bi-functional activity (Bhandiwad et al., 2013), however this result indicates in a *Geobacillus* host, this is not the case. The previous study was carried out in an anaerobic environment, therefore oxygen exposure is likely to be the influencing factor affecting compatibility of the gene. The native AdhE showed activity with both butyryl-CoA and butyraldehyde substrates, demonstrating aldehyde and alcohol dehydrogenase activity.

4.3.3 A fully native system

As the previous results indicate the native genes are functional, a replacement for the heterologous *ter* gene was desired in order to obtain an all native system. The *Geobacillus* genome was investigated for alternative genes to carry out the reaction catalysing the conversion of crotonyl-CoA to butyryl-CoA.

CoA DEPENDENT n-BUTANOL PRODUCTION

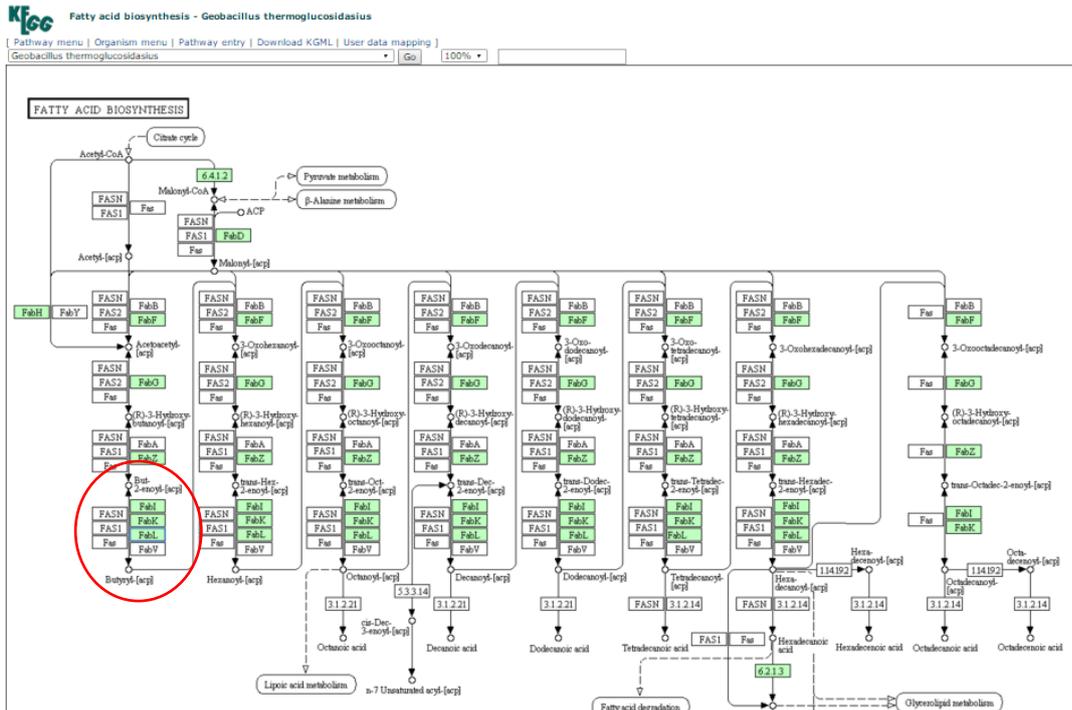


Figure 4.15. Pathway diagram showing fatty acid biosynthesis in *G. thermoglucosidarius*. Taken from the KEGG pathway database. Genes present in the genome are shown in green. An enoyl reductase step is circled in red.

Highlighted in Figure 4.15, the *fabI* gene catalyses the reaction; But-2-enoyl-ACP → butyryl-ACP. *FabI* is the essential enoyl-ACP reductase from the fatty acid biosynthesis pathway (Vick et al., 2015) and has been reported to have the ability to reduce crotonyl-CoA to butyryl-CoA. Bergler et al. (1996) reported utilisation of both NADH or NADPH as cofactor, with the enzyme accepting both 2,3-unsaturated acyl-ACP and acyl-CoA derivatives when activity is NADH-dependent. Vick et al. (2015) describe enoyl-CoA reductase activity from *FabI* supporting β-oxidation reversal in *E. coli*. They found overproduction of *FabI* to be as effective as *ter* for the production of butyrate and longer-chain carboxylic acids in *E. coli*. Despite *FabI* being less kinetically efficient than *Ter* for the NADH-dependent reduction of crotonyl-CoA and with reduced specific activity (Vick et al., 2015). This analysis was carried out in aerobic conditions.

The *Geobacillus* native enoyl reductase enzyme; FabI, provides a potential native gene as a potential replacement for the synthetic *ter* gene. If the findings from previous *E. coli* studies (Vick et al., 2015; Bergler et al., 1996) are found to be comparable in *Geobacillus*. In addition, five different native RBS sequences, previously characterised, were selected to replace the synthetic RBS on all native pathway genes. This offers an all native system for the potential generation of *n*-butanol in an organism where it is not naturally occurring product.

4.3.4 Culture for *n*-butanol production

In order to determine if overexpression of the native system selected was capable of generating *n*-butanol, two plasmids compatible for co-transformation were constructed. The HiFi cloning method was used to assemble the two vectors, before transformation into *G. thermoglucosidasius*.

CoA DEPENDENT n-BUTANOL PRODUCTION

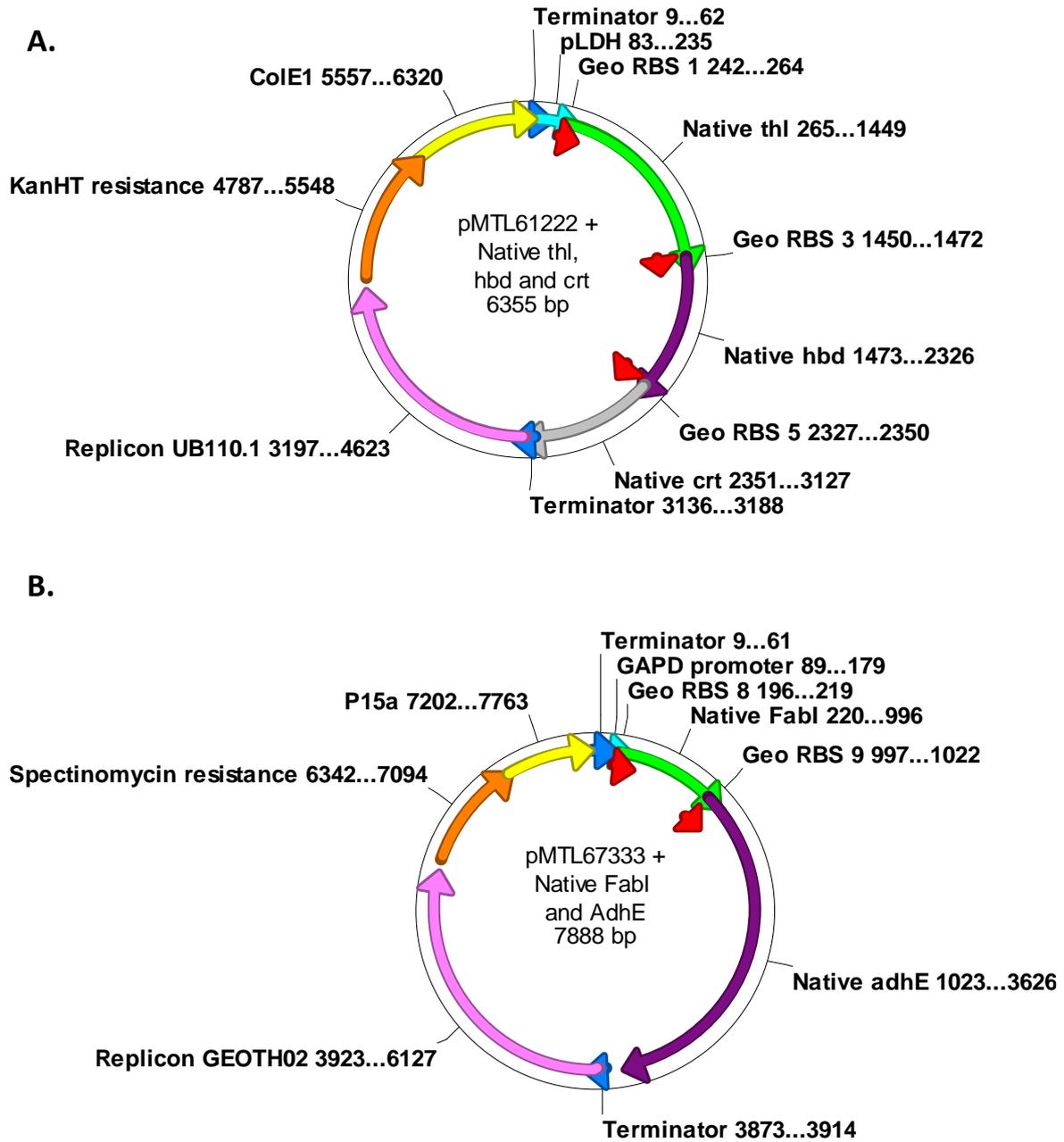


Figure 4.16. Maps of the native gene plasmids co-transformed in *G. thermoglucosidasius*. A. pMTL61222+ native *thl*, *hbd* and *crt* driven by the P_{ldh} promoter, B. pMTL67333+ native *fabI* and *adhE* driven by P_{GAPD} . All five genes included a *Geobacillus* native RBS sequence for initiation of protein translation.

CoA DEPENDENT n-BUTANOL PRODUCTION

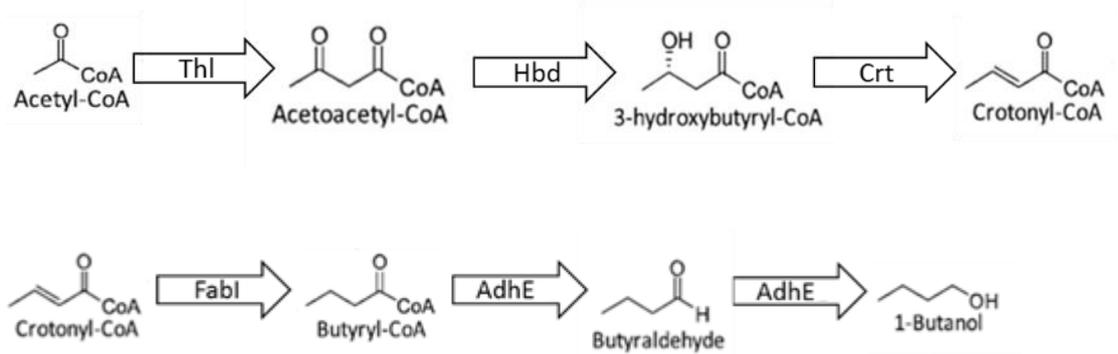


Figure 4.17. The native gene pathway. Overexpression of the five *G. thermoglucosidasius* native genes shown in arrows, has the potential to drive the pathway from Acetyl-CoA to *n*-butanol production.

The two plasmids, as shown in Figure 4.16, were co-transformed into *G. thermoglucosidasius* strains; wild type, *acc* integrated strain and LS242+ Δ *adhE*. The wild type strain includes no further modifications. The *acc* integrated strain includes an acetyl-CoA carboxylase gene (*acc*) integrated into the genome to increase levels of malonyl-CoA. Malonyl-CoA is converted to Acetyl-CoA and therefore increases the substrate pool for Thl activity. Strain LS242+ Δ *adhE* is *G. thermoglucosidasius* with the following genome modifications; Δ *ldh*, *pdh* upregulated, Δ *pfI* and Δ *adhE*. This strain was generated by Sheng, L (work not described here). The LS242+ Δ *adhE* strain was used here as the LS242 strain was designed for high ethanol production. The modifications made are relevant to optimise this pathway also. *Geobacillus* uses the *adhE* gene for ethanol production. The chromosomal *adhE* deletion included in this host strain may disrupt the ethanol production. Here it is attempted to modify the cellular processes into producing butanol instead of ethanol. Overexpression of genes in operons may assist to channel the flow of metabolites through the pathway.

The strains were cultured in both aerobic and micro aerobic conditions in UYSE medium with 20 g/l glucose. Cultures were incubated at 52°C for 49 hours. All

CoA DEPENDENT n-BUTANOL PRODUCTION

experiments were carried out in triplicate. Samples were analysed by GC for *n*-butanol production.

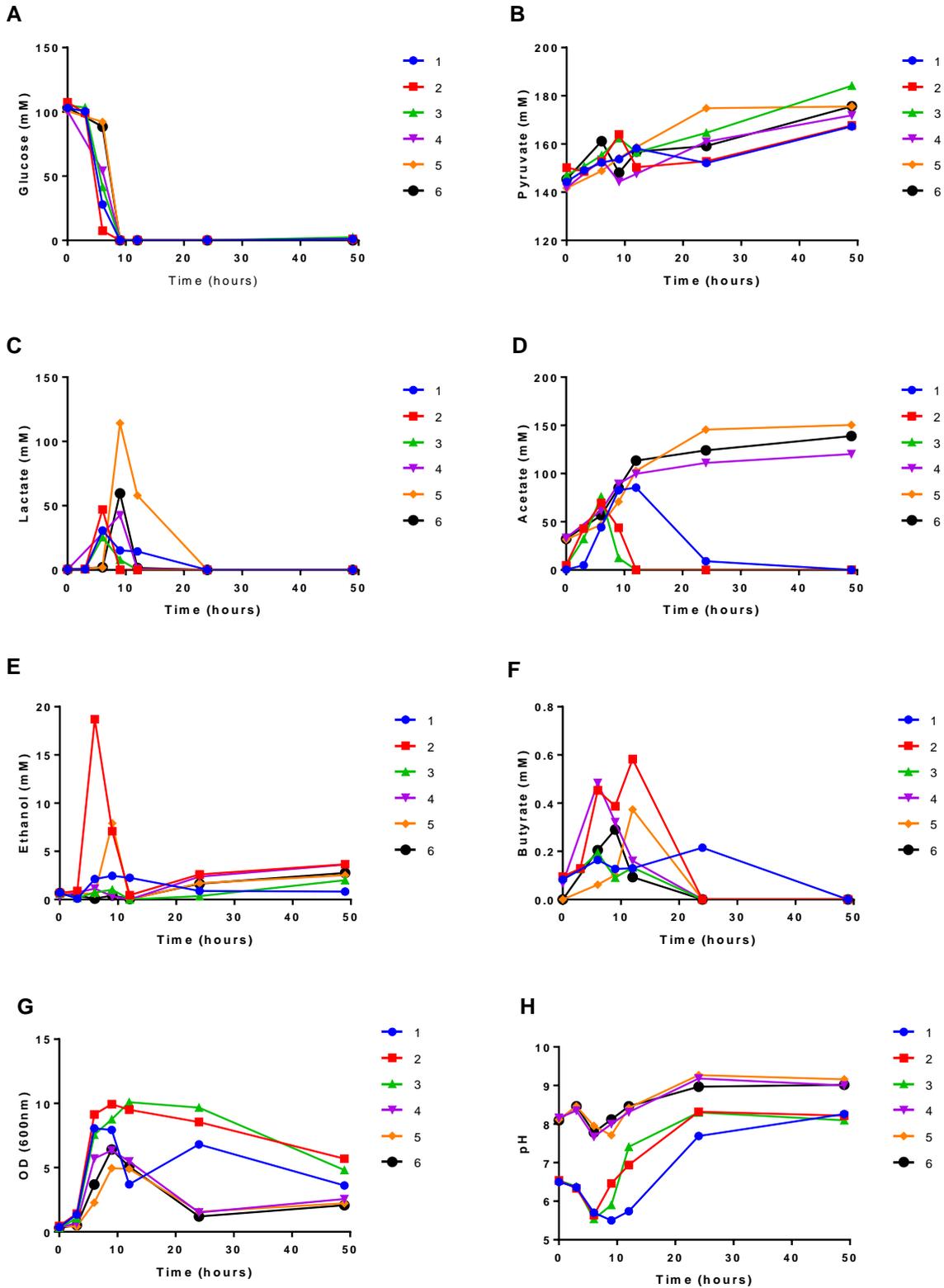


Figure 4.18. Glucose consumption, metabolite production, growth and pH of three *G. thermoglucosidasius* strains expressing the native gene *n*-butanol pathway. Cultures were grown in UYSE medium with 111 mM initial glucose at 52°C for 49 hours. Samples were taken throughout and analysed by GC for metabolite concentrations. A. Glucose, B. Pyruvate, C. Lactate, D. Acetate, E. Ethanol, F. Butyrate, G. Growth determined by absorbance at 600 nm, H. pH. The plasmids expressed in; 1. Wild type *G. thermoglucosidasius*, 2. *acc* integrated strain, 3. Strain LS242 $\Delta adhE$ grown aerobically and 4. Wild type *G. thermoglucosidasius*, 5. *acc* integrated strain, 6. Strain LS242 $\Delta adhE$ grown micro aerobically.

4.3.5 Discussion

The native system did not generate *n*-butanol

No *n*-butanol was detected from cultures expressing the all native gene pathway. This result could be due to a number of factors. The enzyme assay data, presented in Figure 4.14, demonstrated activity of three of the five genes used. The *crt* and *fabI* genes may not be functional. Here no enzyme activity data for the *crt* genes is shown. The Crt enzyme assay requires further method development to provide a conclusive result. For the Crt assay used, activity is detected by monitoring the reaction in the UV range. In order to detect changing absorbance in the UV range the assay was performed in UV quartz cuvettes. The quartz cuvettes have a volume of 3.5 ml. This reaction volume requires greater quantities of reagent, therefore assay development was more limited.

Geobacillus is a native ethanol producer. The native AdhE may therefore be ethanol specific. There is, to date, one bi-functional AdhE encoded by the *G. thermoglucosidasius* genome. The function of this enzyme is to act upon ethanol precursors; acetyl-CoA and acetaldehyde, rather than the *n*-butanol precursors; butyryl-CoA and butyraldehyde. It may be the case the enzyme does not function on butyryl-CoA and butyraldehyde substrates. The enzyme

activity monitored shows a reduction of NADH. This reaction is not specific to the reduction of butyryl-CoA and butyraldehyde. Therefore the activity observed could show an alternative undesired reaction. Alternatively AdhE may show preference for ethanol production, reducing the activity for *n*-butanol production to below detectable levels.

Overproduction of AdhE may result in competition for the acetyl-CoA pool. AdhE converts acetyl-CoA to ethanol via acetaldehyde, shown in Figure 4.1. In addition, acetyl-CoA is a precursor for many other reactions in the cell. AdhE and a range of other enzymes could be diverting the acetyl-CoA precursor, reducing availability for the thiolase reaction. The route to *n*-butanol is a much longer pathway than ethanol, and a pathway which would not naturally occur in the cell. For these reasons the cell is more likely to bypass the Thl enzyme in favour of other reactions. In this case little or no *n*-butanol would be produced.

Here the native FabI enzyme was selected as a potential native gene replacement for the synthetic *ter*. The results of this study do not determine if the *fabI* gene is functional in reducing the CoA, rather than ACP, form of the enoyl substrate. *fabI* alternative genes present natively in *Geobacillus* include *fabK* and *fabL*, which code for enoyl-ACP reductase II and enoyl-ACP reductase III genes, respectively.

Overexpression of this five gene pathway, as shown in Figure 4.17, will create a metabolic burden on the cell. In order to reduce this burden the cell may lose the plasmid or genes. As the genes present on the plasmid are native to the host, introduction results in regions of homology with the chromosome. At these homologous regions recombination events could occur to incorporate the antibiotic resistance gene into the chromosome. Allowing for loss of the remaining plasmid, whilst still ensuring survival under selective pressure. Any recombination event resulting in disruption of the operon or loss of genes would account for the lack of *n*-butanol production.

The enzyme assay results, shown in Figure 4.14, found the integrated *ter* had higher expression levels than plasmid based gene expression. A 231-fold increase was observed. This observation was unexpected. It was considered that the increase in copy number of the plasmid based gene would result in higher expression and therefore greater activity of the enzyme reaction. This was not found to be the case. Higher activity of the integrated gene could be due to greater genetic stability of chromosomal integration in comparison to plasmid based gene expression. The introduction of an additional plasmid could be creating a burden, affecting cellular function. Alternatively inhibitors or enhancing elements could be acting on the gene expression of the plasmid or chromosome respectively. The higher expression level, seen here with the *ter* gene, should be investigated further by integrating the other genes and comparing if a change in activity is observed.

In this work the synthetic and native gene variants have been expressed and evaluated in separate systems. A combination of the synthetic and native genes may offer a pathway with increased activity. According to the enzyme assay data, as shown in Figure 4.14, the native Thl and AdhE enzymes in combination with synthetic Ter and Hbd enzymes, would give a pathway with the greatest activity for each reaction.

Regulatory mechanisms in the cell control expression of the native pathway genes. Further work is required in order to better understand regulation in *Geobacillus* and the impact this may have on product formation.

Reverse β -oxidation

The reactions of the β -oxidation pathway are reversible. Therefore, there is potential to reverse the direction of the pathway. Exploitation of the pathway to operate in the synthetic, rather than degradative direction. Reversing the β -oxidation cycle can be used as a metabolic platform for the synthesis of alcohols. This approach has been demonstrated previously. Reversal of the β -

oxidation cycle was engineered in *E. coli* and used in combination with endogenous dehydrogenases and thioesterases to synthesize *n*-alcohols, fatty acids and 3-hydroxy-, 3-keto- and trans-D2-carboxylic acids (Dellomonaco et al., 2011).

The reverse β -oxidation pathway described by Dellomonaco et al. (2011) involves simplification of the native pathway described here, in a modified host strain. Thiolase and aldehyde/alcohol dehydrogenase activity are required. Utilisation of this approach negates the need for the other three genes (*hbd*, *crt* and *ter /fabI*) used in this pathway.

In order to run the β -oxidation cycle in reverse, there must be an absence of the natural substrate, fatty acids, and presence of an alternative carbon source such as glucose. However in such conditions the cell's regulatory systems would repress expression of the β -oxidation pathway. Therefore mutation or removal of the cells regulatory components is required. Dellomonaco et al. (2011) created a strain with; *fadR*, *atoC* and *crp* mutations and *arcA* gene deletion. This enabled constitutive expression of the β -oxidation system in *E. coli*, in the absence of fatty acids and also prevented carbon catabolite repression in the presence of glucose.

n-butanol can be generated from one turn of the reverse β -oxidation cycle, with carbon chain elongation generating butyryl-CoA from acetyl-CoA (+2 carbons). This carbon chain elongation can be carried out by a thiolase. Further conversion of the butyryl-CoA to *n*-butanol requires expression of a butyraldehyde/butanol dehydrogenase.

A functional one turn reversal of the β -oxidation cycle was demonstrated to produce *n*-butanol in *E. coli* (Dellomonaco et al., 2011). In this work *E. coli* endogenous genes; *fucO* and *yqeF* were overexpressed. Resulting in a reported flux exceeding that of native or engineered fermentative pathways

(Dellomonaco et al., 2011). However no *n*-butanol was synthesized upon overexpression of *fucO* and *yqeF* in a wild-type *E. coli* host. Demonstrating the requirement of host modifications are essential to result in an active reversal of the β -oxidation cycle.

4.4 Modification of the synthetic pathway

Following a lack of *n*-butanol resulting from use of both the synthetic and native *n*-butanol pathway genes, as described previously, here changes to the synthetic pathway are made in attempt to construct a functional pathway.

4.4.1 Creating a driving force

The enzymes used in this pathway are reversible, catalysing both forward and reverse reactions. This presents a challenge in generating flux through the pathway toward the desired final product. To create a metabolic driving force through this pathway enzyme replacements were made. Selection of enzymes catalysing irreversible reactions were selected to create a driving force.

Acetoacetyl-CoA synthase (NphT7)

The initial step of *n*-butanol biosynthesis is the condensation of two molecules of acetyl-CoA forming acetoacetyl-CoA. This reaction is catalysed by thiolase and is fully reversible in the cell. Acetyl-CoA condensation catalyzed by thiolase is thermodynamically unfavourable with K_{eq} of 10^{-5} .

To circumvent this unfavourable reaction, an alternative route using malonyl-CoA was used. Acetoacetyl-CoA synthase (NphT7) catalyzes a single condensation of acetyl-CoA and malonyl-CoA to give acetoacetyl-CoA and CoA. This alternative route through malonyl-CoA synthesis is effectively irreversible due to ATP hydrolysis and irreversible decarboxylation. ATP hydrolysis serves as an energy input to convert the thermodynamically unfavourable acetyl-CoA condensation into a favourable reaction (Lai & Lan, 2015). ATP is invested to drive the reaction.

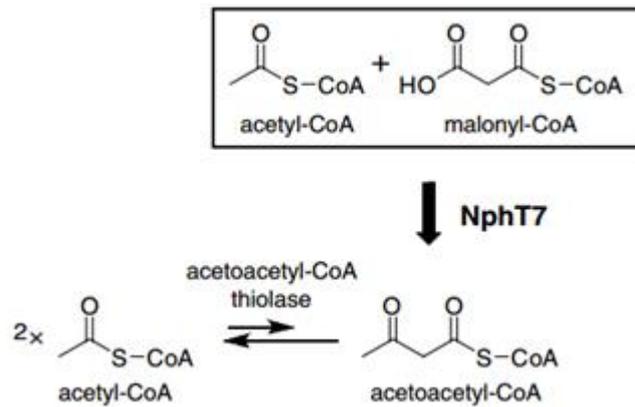


Figure 4.19. Diagram depicting the reaction catalysed by both the thiolase and acetoacetyl-CoA synthase genes (Okamura et al., 2010). Two acetyl-CoA molecules are condensed to acetoacetyl-CoA by thiolase, this reaction favours the reverse direction. Alternatively one acetyl-CoA molecule is condensed with malonyl-CoA to acetoacetyl-CoA in an essentially irreversible reaction.

In the following strategy, overexpression of an *nphT7* (EC 2.3.1.194; acetoacetyl-CoA synthase) gene was used as an alternative to the *thl* gene, to create a driving force through the pathway. An Acyltransferase catalysing the reaction; acetyl-CoA + malonyl-CoA = acetoacetyl-CoA + CoA + CO₂. From the thiolase superfamily involved in the mevalonate pathway. Sourced from a soil isolated *Streptomyces* sp. strain. Here the gene is synthesised, codon optimised for *G. thermoglucosidasius*, according to the codon usage table in Appendix A8. The synthetic gene was designed to include the RBS from *Geobacillus* glucose-6-phosphate isomerase and a C-terminal flag tag.

CoA DEPENDENT n-BUTANOL PRODUCTION

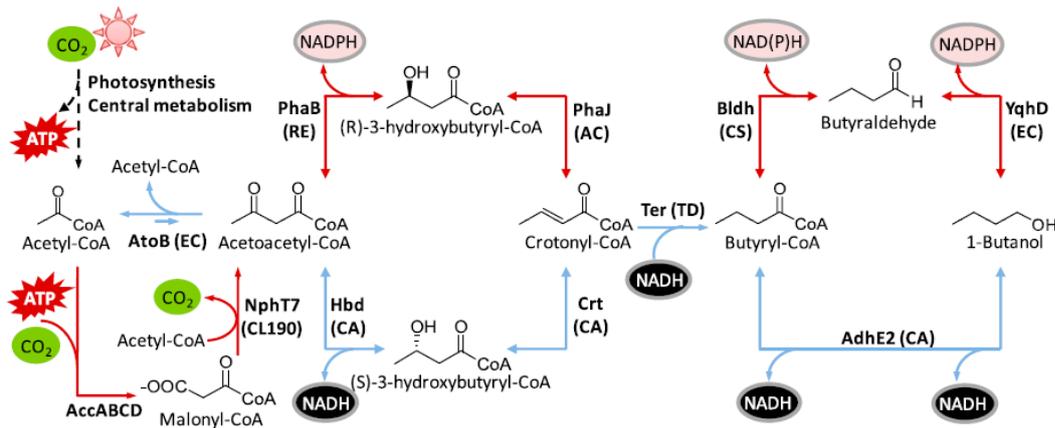


Figure 4.20. Variations in the CoA-dependent *n*-butanol pathway (Lan & Liao, 2012). The original synthetic pathway is shown in blue. Alternative routes are shown in red.

Okamura et al. (2010) suggest that use of NphT7 can significantly increase the concentration of acetoacetyl-CoA in cells. This approach has previously been demonstrated in the cyanobacterium *Synechococcus elongatus*, improving *n*-butanol production (Lan & Liao, 2012).

Acetyl-CoA carboxylase (Acc)

Exploitation of NphT7 requires both acetyl-CoA and malonyl-CoA substrates to generate acetoacetyl-CoA. In this modified synthetic pathway, the strategy was to overproduce acetyl-CoA carboxylase (Acc) in order to increase the malonyl-CoA pool in the cell through the integration of the encoding gene into the genome. Acc is taken from the fatty acid synthetic pathway where it forms malonyl-CoA from acetyl-CoA. Cultures overexpressing the *acc* gene have been shown to overproduced malonyl-CoA (Davis et al., 2000). Accordingly, the strategy was to overproduce Acc in combination with NphT7 to increase the substrate pools available and thereby increase flux through the *n*-butanol pathway.

The Acc enzyme is a complex made up of four subunits. The subunits; carboxyltransferase α , biotin carboxyl carrier protein (BCCP), biotin carboxylase

and carboxyltransferase β , are encoded by genes *accA*, *accB*, *accC*, and *accD*, respectively.

The Acc reaction consists of two discrete half-reactions, as shown in Figure 4.21. In the first reaction, biotin is carboxylated by bicarbonate in an ATP-dependent reaction to form carboxybiotin. In the second reaction, the carboxyl group is transferred from carboxybiotin to acetyl-CoA to form malonyl-CoA (Davis et al., 2000). *accC* and *accB* are responsible for the first reaction. *accA* and *accD* are responsible for the second.

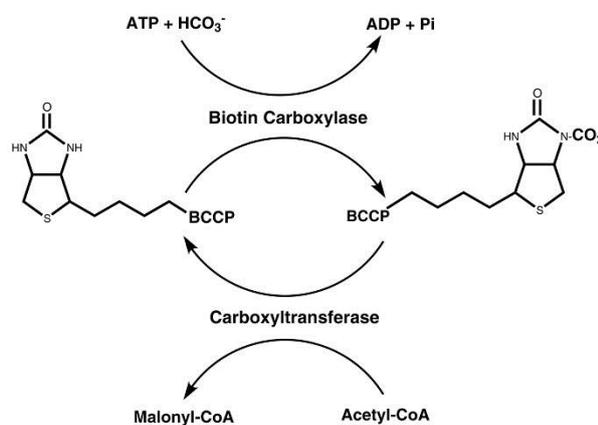


Figure 4.21. The reaction mechanism of acetyl-CoA carboxylase (Davis et al., 2000). A four sub-unit enzyme complex is required for the conversion of Acetyl-CoA to Malonyl-CoA.

Attempts to clone the gene into an expression vector containing other *n*-butanol genes were not successful. This could be due to the large size of the complex, a total size of 4,390 bp. Alternatively it could have been due to repeat regions. Each of the four subunits was synthesised with the synthetic RBS, resulting in four homologous sequences. Instead of plasmid based expression of the *acc* gene, a genome integrated strain was used. The *acc* integration work was done by Bashir, Z. and not described here. The *acc* integrated gene was found to be stable in the host.

Trans-enoyl-CoA reductase (Ter)

Ter was used in the synthetic pathway to replace the Bcd–Etf complex for reduction of crotonyl-CoA. Ter directly utilizes NADH as the reducing cofactor for the reduction of crotonyl-CoA to butyryl-CoA with a large negative change in free energy, making it irreversible (Lai & Lan, 2015). Ter, therefore creates a further driving force for this pathway, to direct the flux towards *n*-butanol.

Following stability concerns encountered previously, when using a genome integrated *ter* gene, here the *ter* gene is expressed on plasmid.

4.4.2 AdhE alternatives

To obtain *n*-butanol production the final steps in the pathway require conversion of butyryl-CoA to butyraldehyde and subsequent conversion to *n*-butanol. The synthetic AdhE previously considered, has been shown not to be suitable for this application. Although the enzyme is reported to be bi-functional (Bhandiwad et al., 2013), enzyme assay data, presented in Figure 4.14 and discussed previously, found this not to be the case. Only butyraldehyde activity was observed. The AdhE protein is also reported to be highly oxygen sensitive (Lan et al., 2013). A lack of expression detected by Western blot, section 4.2.4, suggests the enzyme is not compatible with a *Geobacillus* host. Therefore an alternative gene/genes are required as a replacement for the synthetic *adhE* previously selected. The replacement gene must be oxygen tolerant and thermostable for compatibility. The replacement gene must also carry out the C4 aldehyde and alcohol dehydrogenase activity required, specific for *n*-butanol production.

Table 4.4. Aldehyde and alcohol dehydrogenase enzymes screened.

	Enzyme	Source	Bi-functional	Activity	Thermo-tolerant	Oxygen tolerant	<i>n</i> -butanol specific
1	AdhE2	<i>T. thermosaccharolyticum</i>	No	No	Yes	No	No
2	AdhE2	<i>C. acetobutylicum</i>	Yes	No	No	No	Yes
3	AdhE2	<i>G. thermoglucosidasius</i>	Yes	Yes	Yes	Yes	No
4	AdhE1	<i>C. acetobutylicum</i>	No	No	No	No	Yes
5	Bldh	<i>C. saccharoperbutyl acetonicum</i>	No	No	Unknown	No	Yes
	Yqhd	<i>E. coli</i>	No	Yes	Yes	Yes	Yes
6	PduP	<i>Salmonella enterica</i> serovar <i>Typhimurium</i>	No	Yes	Yes	Yes	Yes
	YqhD	<i>E. coli</i>	No	Yes	Yes	Yes	Yes

Selection of alternative genes for screening

In addition to the synthetic *adhE2* gene from *T. thermosaccharolyticum* and the *G. thermoglucosidasius* native *adhE2*, *adhE1* and *adhE2* from *C. acetobutylicum* were also screened. These two enzymes were selected as they are responsible for butanol production in the native production organism *C. acetobutylicum*. *C. acetobutylicum* produces 0.28–0.33 g *n*-butanol/g glucose (Kataoka et al., 2015). The two aldehyde dehydrogenases were screened to determine if they were compatible with a *Geobacillus* host. In addition to the reported bi-functional enzymes, two sets of gene pairs were screened. The pairs each encode separate aldehyde and alcohol dehydrogenase genes. By separating each step in the pathway, it should be possible to screen for both butyraldehyde and *n*-butanol production to determine functionality and

eliminate any blocks in the pathway. Butyraldehyde dehydrogenase encoded by *bldh* from *C. saccharoperbutylacetonicum* was selected as a candidate as Kataoka et al. (2015) reported aerobic *n*-butanol production using an engineered *E. coli* strain expressing a synthetic pathway including the *bldh* gene from *C. saccharoperbutylacetonicum*. The latter is a butanol producing bacterium. An *n*-butanol titre of 4.07 g/L has been reported with inclusion of this enzyme (Kataoka et al., 2015). A second CoA-acylating aldehyde dehydrogenase gene has been reported to be oxygen tolerant; *pduP* from the 1,2-propanediol degradation pathway of *S. enterica* (Lan et al., 2013). This enzyme was selected for investigation here as it has been demonstrated functional for butanol production in an aerobic cyanobacterial system (Lan et al., 2013). PduP from *S. enterica* has a catalytic efficiency (kcat/Km) of 292 s⁻¹ mM⁻¹ for butyryl-CoA which is 7 times higher than that for acetyl-CoA (Lan et al., 2013). With oxygen tolerance and specificity towards the substrate butyryl-CoA, *pduP* is a promising candidate gene. YqhD is a broad substrate range aldehyde reductase with various applications in production of biofuels and chemicals (Jarboe, 2011). Here the NADPH-dependent YqhD from *E. coli* was screened. This enzyme has also previously been demonstrated in an oxygen tolerant system (Lan et al., 2013).

The genes listed in Table 4.4 were amplified from the source organisms listed, except 1. *adhE2* and 6. *pduP* encoding genes were synthesised with codon optimisation for *G. thermoglucosidasius*. The genes were cloned into vector pMTL61122. The genes all included the synthetic RBS and were cloned behind P_{ldh}. To determine functionality and potential application in the pathway for *n*-butanol production, enzyme assay and growth in supplemented culture were carried out.

CoA DEPENDENT n-BUTANOL PRODUCTION

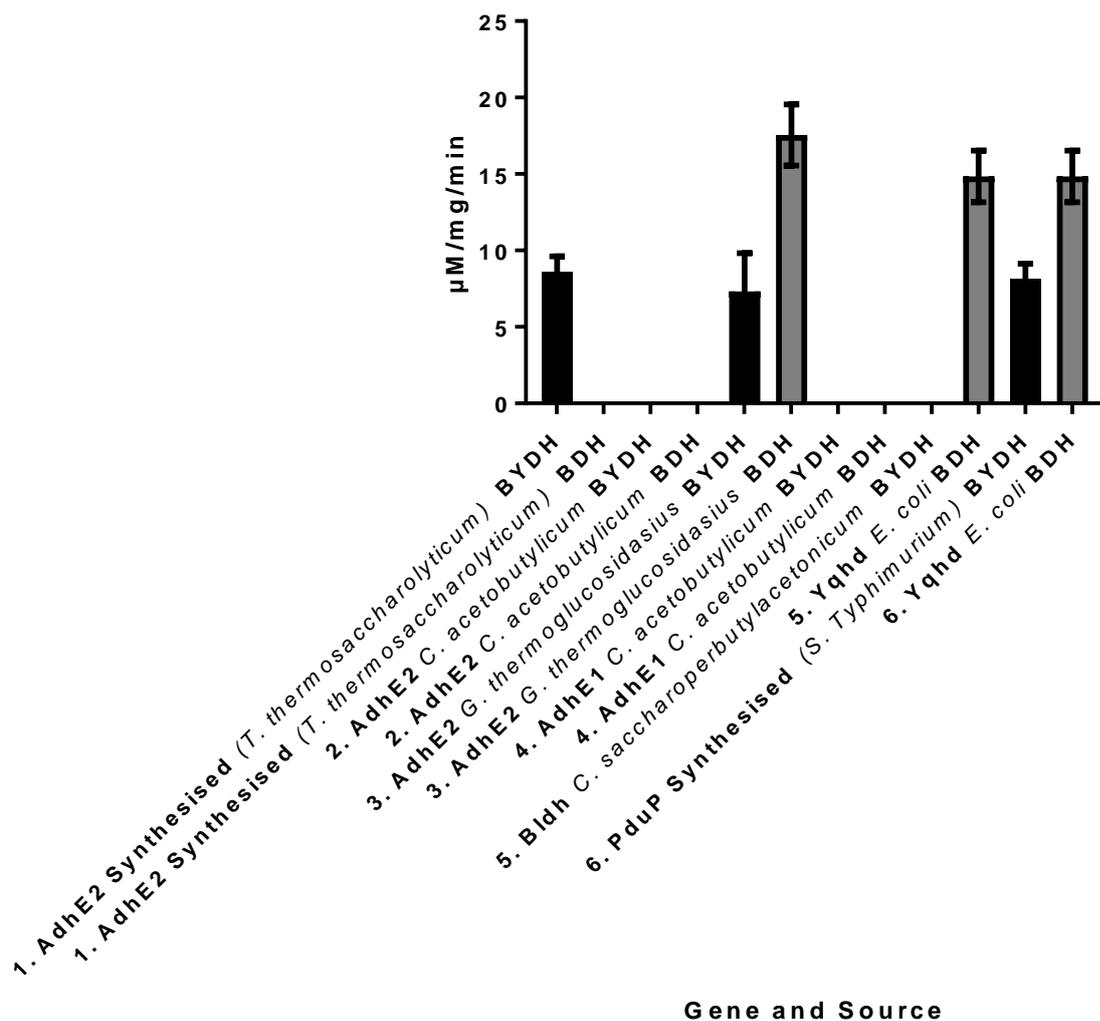


Figure 4.22. Enzyme activities of potential AdhE replacements. Activity of the enzymes listed in Table 4.4 was determined by monitoring a decrease in absorbance at 340 nm corresponding to the consumption of NADH or NADPH at 52°C. Butyryl-CoA was the substrate for the butyraldehyde dehydrogenase reaction (BYDH) and butyraldehyde was the substrate for the butanol dehydrogenase reaction (BDH). The molar extinction coefficient for consumption of NADH at 340 nm was taken as 6,220 M⁻¹cm⁻¹.

Table 4.5. Concentration of product formation from supplemented culture.

2SPYNG medium supplemented with 5mM butyryl-CoA for butyraldehyde detection and 5mM butyraldehyde for n-butanol detection.

	Butyraldehyde (mM)	<i>n</i> -butanol (mM)
1. AdhE2 Synthesised (<i>T. thermosaccharolyticum</i>)	1.64 (+/-0.46)	0
2. AdhE2 <i>C. acetobutylicum</i>	0	0
3. AdhE2 <i>G. thermoglucosidasius</i>	0	0
4. AdhE1 <i>C. acetobutylicum</i>	0	0
5. Bldh and Yqhd	0	0
6. PduP and Yqhd	2.38 (+/-0.67)	3.49 (+/-0.88)

Enzyme assay results, shown in Figure 4.22, demonstrate the combination PduP and Yqhd is the only solution functional in both BYDH and BDH reactions. The native *G. thermoglucosidasius* AdhE enzyme also demonstrated activity here, however this enzyme was previously tested and found not to generate *n*-butanol (section 4.3.4). The synthetic AdhE2 enzyme from *T. thermosaccharolyticum* did not demonstrate activity with butyraldehyde substrate indicating the enzyme is compatible but not bi-functional. This enzyme could be used in combination with a functional alcohol dehydrogenase enzyme. Both the AdhE1 and AdhE2 enzymes from *C. acetobutylicum* demonstrated no activity for either reaction. This indicates the enzymes are not compatible with *Geobacillus*. The enzymes are from an anaerobic and mesophilic source and so either or both of these factors could be the cause of incompatibility here. The Bldh enzyme demonstrated no activity, indicating this enzyme is also incompatible with the host organism or conditions tested here.

G. thermoglucosidasius harbouring plasmid containing the gene/genes listed in Table 4.4 were grown in 2SPYNG medium supplemented with 5 mM butyryl-CoA for butyraldehyde detection and 5 mM butyraldehyde for *n*-butanol detection. Cultures were grown for 12 hours incubated at 52°C with shaking at

250 rpm. Samples were taken and analysed by HPLC. The results, shown in Table 4.5, concur with the findings from enzyme assay. Only the combination of PduP and YqhD demonstrate conversion of butyraldehyde to *n*-butanol. Therefore this combination was selected for inclusion in the *n*-butanol pathway going forward, replacing the *adhE* gene.

4.4.3 The *n*-butanol production strain

The modifications, detailed previously, resulted in the engineered pathway shown in Figure 4.23. The plasmid based pathway was expressed in a *G. thermoglucosidasius* strain with the *acc* gene integrated into the genome. This pathway provides a driving force at the initial committed step using the NphT7 enzyme and by Ter through the pathway. The *acc* gene is expected to increase the malonyl-CoA pool for the initial substrate. Finally PduP and YqhD have been demonstrated able to carry out the two final steps in the pathway converting butyryl-CoA to *n*-butanol.

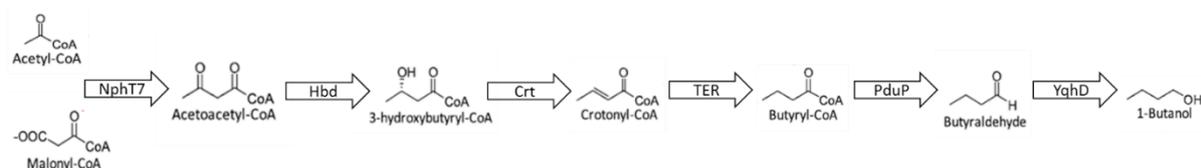


Figure 4.23. A modified CoA dependent *n*-butanol pathway. This improved pathway includes two irreversible reactions to create driving force and a functional alcohol dehydrogenase.

CoA DEPENDENT n-BUTANOL PRODUCTION

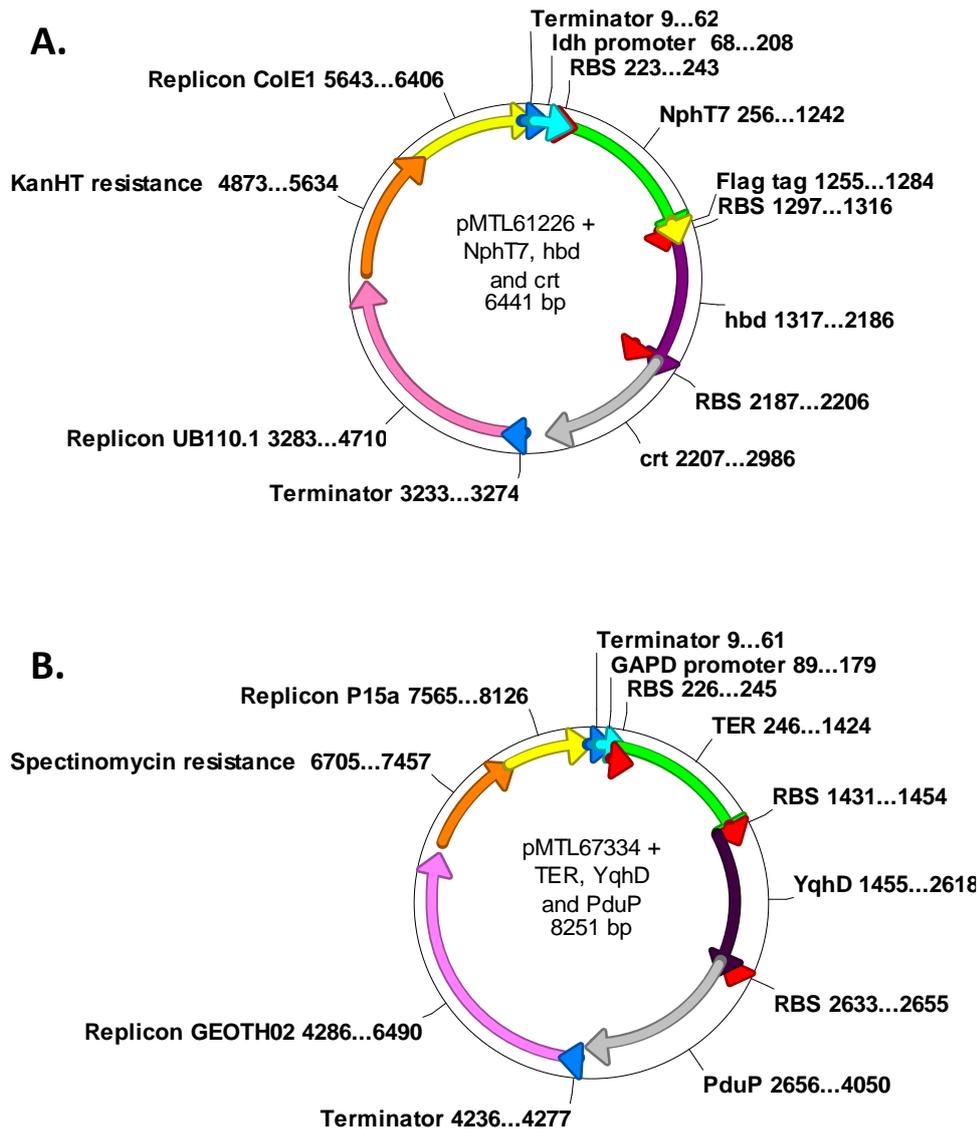


Figure 4.24. Maps of the two plasmids containing the modified CoA dependent *n*-butanol pathway. A. pMTL61226+ *nphT7* with the RBS from *Geobacillus*' glucose-6-phosphate isomerase, *hbd* with RBS 6 from *C. sporogenes*, *crt* with RBS 5 from *C. sporogenes*. The three gene operon was cloned behind P_{Idh} . **B.** pMTL67334+ *ter* with RBS 9 from *C. sporogenes*, *yqhD* with RBS 5 from *Geobacillus*, *pduP* with the RBS from the *Geobacillus* fructose-bisphosphate aldolase (class II) gene. The three gene operon was cloned behind the P_{GAPD} . The two vectors, A. and B. were co-transformed into a *G. thermoglucosidasius* strain with the *acc* gene integrated into its genome.

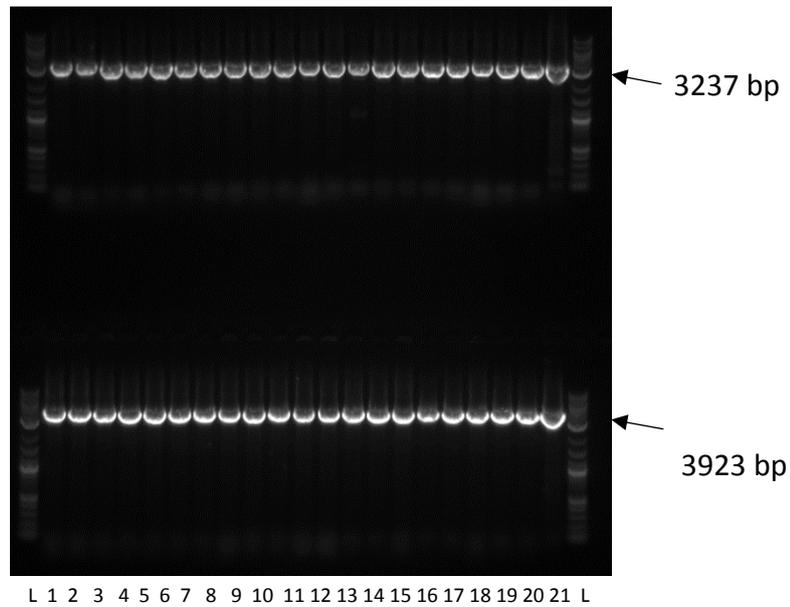


Figure 4.25. Co-transformation colony PCR. Visualisation of colony PCR of 20 *G. thermoglucosidasius* colonies post co-transformation with both pMTL61226+ *nphT7*, *hbd* and *crt* and pMTL67334+ *ter*, *yqhd* and *pduP*. Lanes 1 to 20 are test colonies, lane 21 is a positive control. L represents 2 log ladder. The top bands of 3237bp show presence of pMTL61226+ *nphT7*, *hbd* and *crt*. The lower bands of 3923 bp show presence of pMTL67334+ *ter*, *yqhd* and *pduP*.

CoA DEPENDENT n-BUTANOL PRODUCTION

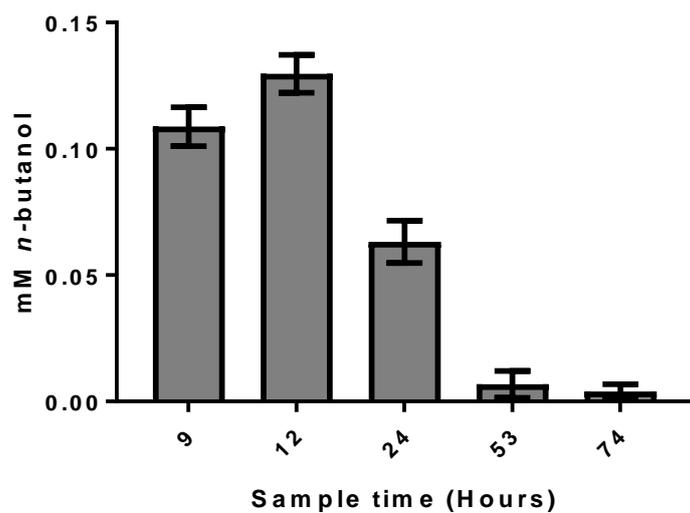


Figure 4.26. *n*-butanol production results under aerobic growth conditions.

The modified production strain was cultured for *n*-butanol production. The *Geobacillus* strain described was grown in UYSE medium + 20 g/l glucose at 55°C with 250 rpm shaking. A starting inoculum of OD 0.3 was used. Culture was performed in triplicate. Samples were taken at the time points indicated and analysed by GC.

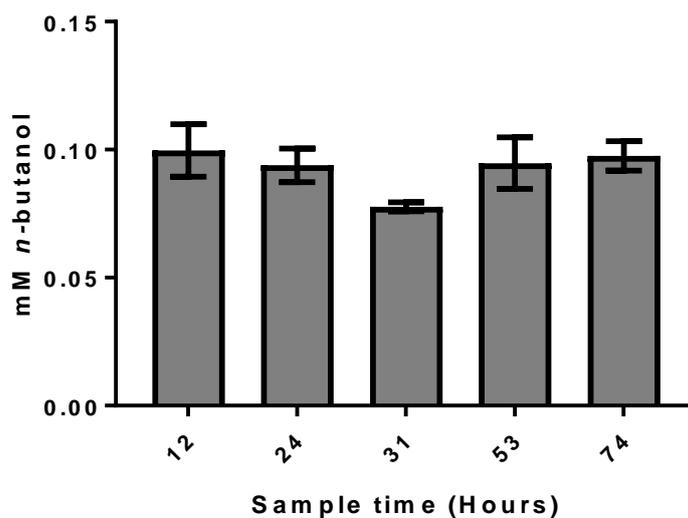


Figure 4.27. *n*-butanol production results under micro aerobic growth conditions. The modified production strain was cultured for *n*-butanol production. The *Geobacillus* strain described was grown in UYSE medium + 20 g/l glucose at 55°C with 250 rpm shaking. A starting inoculum of OD 0.3 was used. After 9h the culture reached OD 3 and was subsequently switched to micro aerobic growth conditions. 1ml of 20% glucose was added to the culture medium at the switch point. The culture was performed in triplicate. Samples were taken at the time points indicated and analysed by GC.

CoA DEPENDENT n-BUTANOL PRODUCTION

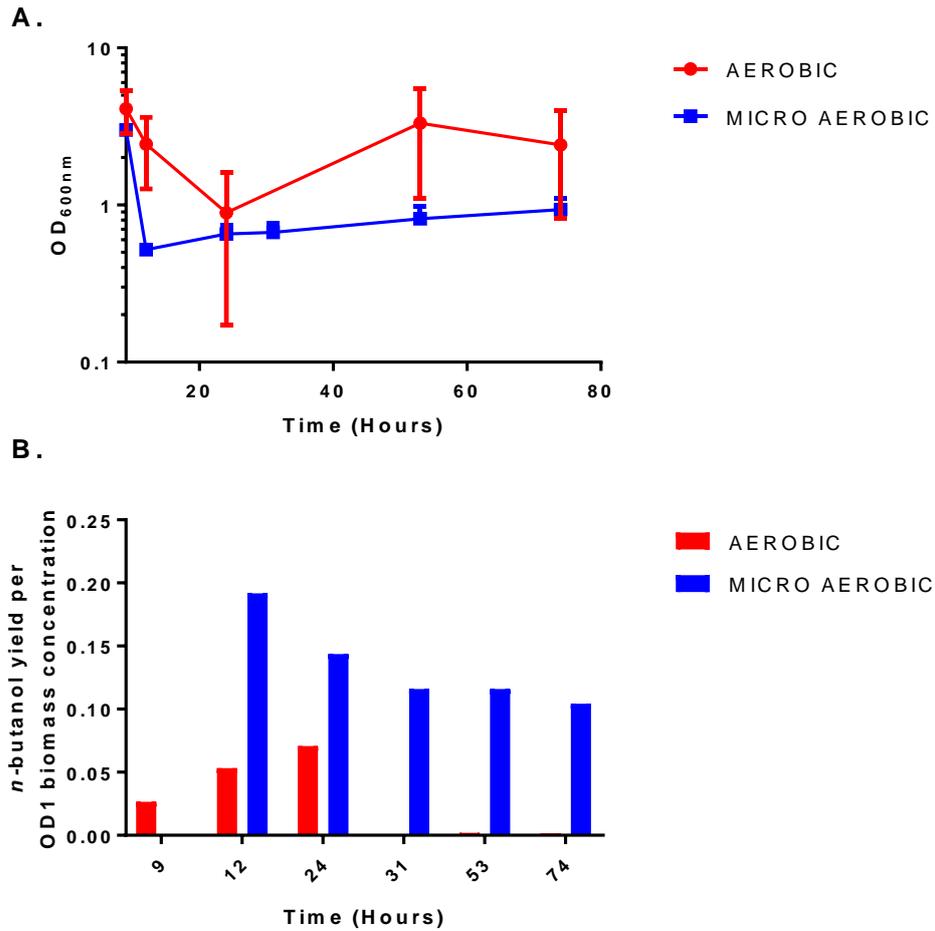


Figure 4.28. Growth and relative yield of the modified *G. thermoglucosidasius* *n*-butanol production strain. A. The OD_{600nm} of aerobic culture samples as described in figure 4.26 shown as red circles. The OD_{600nm} of micro aerobic culture samples as described in figure 4.27 shown as blue squares. B. The *n*-butanol yield per culture biomass concentration of OD1. The average *n*-butanol concentration and average biomass concentrations were used.

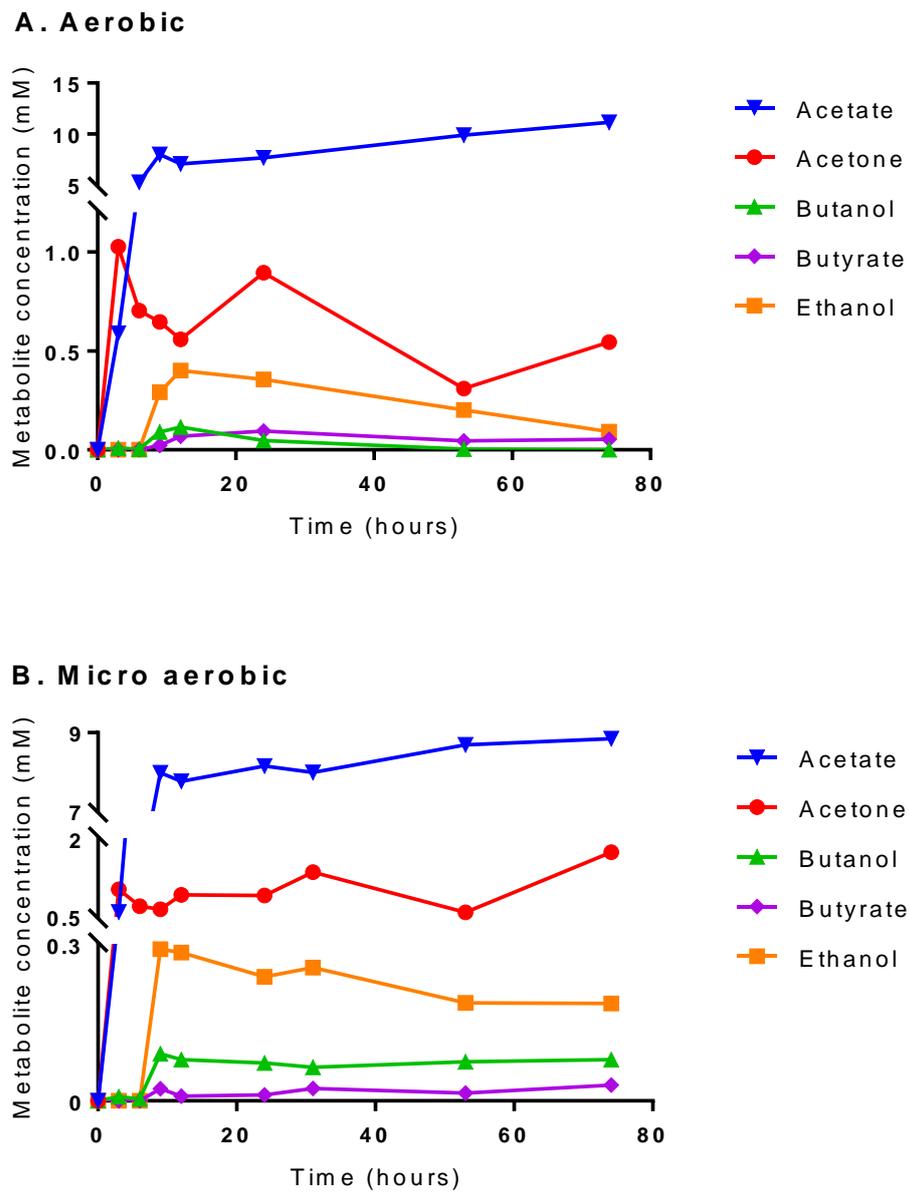


Figure 4.29. Metabolite production of the modified *G. thermoglucosidasius* *n*-butanol production strain. The strain was cultured, as described previously, with samples analysed for metabolite concentrations at the time points indicated.

This *n*-butanol production demonstrates, to our knowledge, the first reported *n*-butanol produced in aerobic and thermophilic conditions. Although at low concentration. The highest concentration observed was 0.137 mM or 10.166

mg/l after 12 hours aerobic culture. This result, nevertheless, demonstrates proof of concept.

Figure 4.28 A. shows the OD_{600nm} of samples cultured both aerobic and micro aerobically. When cultures are switched to micro aerobic conditions the cell density reduces. The aerobic cultures show increased variability due to clumping of the cells as discussed in section 6.5.1. When *n*-butanol production is normalised by biomass concentration, as shown in Figure 4.28 B., the yields produced from cultures in a micro aerobic environment are higher than an aerobic environment. The reduction in *n*-butanol concentration over time under aerobic conditions, shown in Figure 4.26, could be due to evaporation of the volatile compound. Alternatively the *n*-butanol produced could be re-assimilated by the cells. This reduction in concentration over time is not seen in the cultures grown micro aerobically. This difference could be due to the closed system preventing evaporation of *n*-butanol. Alternatively, it could be due to the supplementation of the medium with additional glucose after 9 hours. The cells are able to utilise this carbon source preferentially and therefore do not require the *n*-butanol.

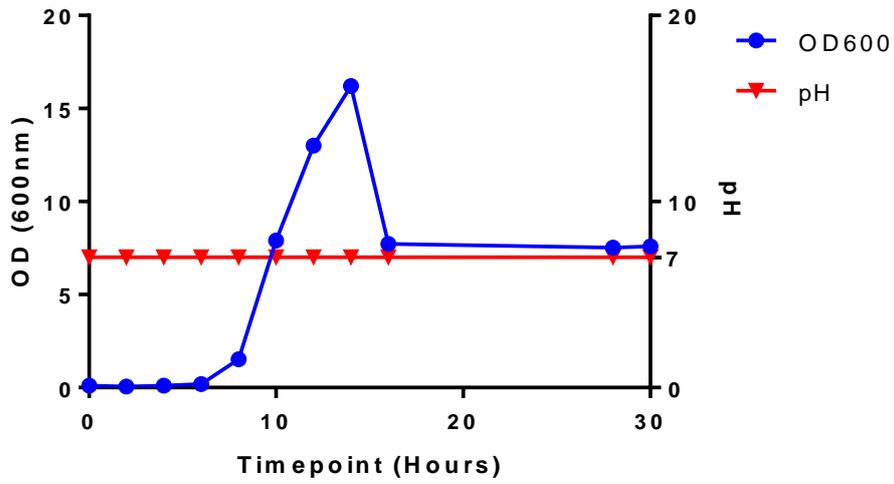
Following this demonstration of *n*-butanol production using the modified CoA dependent pathway, the same plasmids, as shown in Figure 4.24, were co-transformed into a *G. thermoglucosidasius* strain with the *adhE* deletion. This *adhE* deletion strain was created by Sheng, L. work not described here. The *n*-butanol pathway was tested in this strain as the *adhE* deletion knock out reduces ethanol production. *G. thermoglucosidasius* culture results in mixed product. By diverting the cellular resources and substrate pools to *n*-butanol rather than ethanol production, the concentration of *n*-butanol could be increased. Following culture of the $\Delta adhE$ strain with the full *n*-butanol pathway in both aerobic and micro aerobic conditions in UYSE medium with 20 g/l glucose incubated at 55°C. No *n*-butanol production was detected. This result indicates either the *acc* gene is required to increase the substrate pool

to enable flux through the pathway. Or deletion of the *adhE* gene could be having a detrimental effect on the host strain.

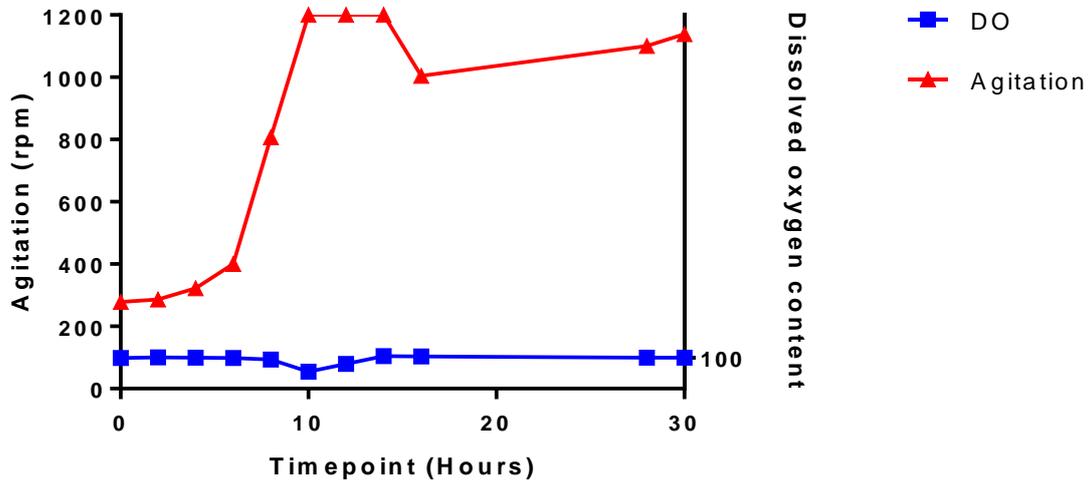
In an attempt to increase the *n*-butanol concentration, the final production strain was cultured on a larger scale in a 1 litre fermentation vessel. The *n*-butanol production strain; *G. thermoglucosidasius* with *acc* integration harbouring both pMTL61226+ *nphT7*, *hbd*, *crt* and pMTL67334+ *ter*, *yqhd*, *pduP* plasmids, was grown in UYSE medium with 20 g/l glucose, equivalent to an initial starting concentration of 111 mM glucose. The incubation was maintained at 55°C. pH 7 was maintained with 3M NaOH. Gas flow rate of 1 vessel volume per minute was used and the gas used was air. Antifoam was added when required. Samples were taken throughout and analysed by HPLC.

CoA DEPENDENT n-BUTANOL PRODUCTION

A



B



C

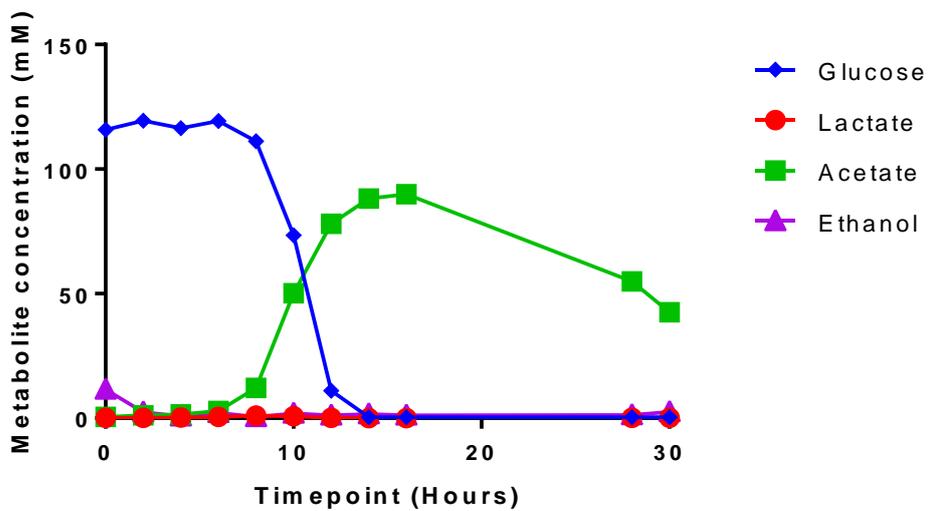


Figure 4.30 Scaled up culture of the *G. thermoglucosidasius* *n*-butanol production strain. Conditions monitored during culture of the production strain in a 1 l vessel. A. Growth and pH, B. Agitation and dissolved oxygen content, C. Metabolite production.

After 6 hours lag time, OD and glucose consumption increased. A maximum OD of 16.2 was reached after 14 hours. As the OD increased the agitation required to maintain dissolved oxygen content at 100 also increased up to the maximum 1200 rpm. The major metabolite was acetate, producing 90 mM. No *n*-butanol was detected. This indicates further optimisation of the growth conditions is required to ascertain suitable protocol for *n*-butanol production.

4.4.5 Discussion

n-butanol production

By engineering the CoA dependent pathway, production of *n*-butanol has been demonstrated here. An *n*-butanol concentration of 0.137 mM (10.166 mg/l), although small, shows proof of concept. This is the first demonstration of *n*-butanol production from a *Geobacillus* host and also from any thermophilic and aerobic production. These results demonstrate the importance of driving forces for the production of non-native products. Further modifications to the host strain, the *n*-butanol pathway, media used and culture conditions used are required to increase production further.

Driving force

By using a synthetic pathway design that incorporates an effectively irreversible step at the initial committed step in the pathway, the substrates are drawn through the pathway toward product formation despite subsequent reversible steps. *n*-butanol production was not observed with the thiolase enzyme, indicating a driving force against the unfavourable thermodynamic

gradient is required. However intermediates can be lost from the pathway to alternative products.

AdhE alternatives

As a by-product of aerobic respiration, reactive oxygen species are formed within prokaryotic and eukaryotic cells. These oxides and radicals can oxidize membrane lipids in cell walls, which leads to the formation of lipid peroxides (Pérez et al., 2008). The cellular degradation of these lipid peroxides results in the formation of short-chain aldehydes, including butyraldehyde. Short chain aldehydes are toxic to the cell and so aldehyde dehydrogenases reduce aldehyde substrates to alcohols.

Alcohol dehydrogenases catalyse reversible oxidation of ethanol or butanol to acetaldehyde or butyraldehyde, respectively. Adh enzymes can be grouped based on their specific co-factors; NADP dependent enzymes, pyrroloquinolinequinone-, haem-group and F₄₂₀-dependent Adhs and flavin adenine dinucleotide-dependent isozymes. NADP-dependent alcohol dehydrogenases are most commonly used in heterologous pathways for metabolic engineering purposes. NADP-dependent alcohol dehydrogenases can be categorized into three structurally and catalytically, different groups; zinc-containing long chain Adhs, metal-free short chain Adhs and metal-containing Adhs (Elleuche & Antranikian, 2013). The alcohol dehydrogenase enzyme used here, Yqhd, is metal containing with a Zn²⁺ ion in the active site.

Lan et al. (2013) describe the oxygen sensitivity of CoA-acylating aldehyde dehydrogenase as the key limiting factor to produce alcohols through the CoA-dependent route, in their work with cyanobacteria. This appears to also be the case here. With use of *Geobacillus* as the host strain, oxygen tolerance is essential for all reactions in the pathway. In order to identify a functional *n*-butanol pathway, six options for the alcohol/aldehyde reduction steps were screened. This screening identified a compatible pair of enzymes which have

now been shown capable of conversion of butyryl-CoA to *n*-butanol in an aerobic environment.

4.5 Discussion

4.5.1 The CoA dependent *n*-butanol pathway

In recent years the CoA dependent pathway for the generation of *n*-butanol has been well studied. Based on the native clostridial production, the pathway has been expressed in a range of non-native production organisms such as *E. coli* (Atsumi et al., 2008), *S. cerevisiae* (Schadeweg & Boles, 2016) and cyanobacteria (Lan & Liao, 2013). Initial attempts in heterologous expression of the CoA dependent pathway in other organisms resulted in much lower titres than clostridia. Much optimisation of NADH co-factor utilization was required to achieve 30 g/l *n*-butanol titre in fed-batch fermentation of *E. coli* (Shen et al., 2011).

In this work several variations of the CoA dependent pathway were tested. The pathway required redesign for functionality in the desired host strain. With successful *n*-butanol production resulting from a heterologous pathway containing genes from a variety of sources.

Initial work to introduce *n*-butanol production using heterologous expression of a synthetic pathway based on native production in the thermophilic *T. thermosaccharolyticum* was unsuccessful. This result highlighted problems with the original gene/enzyme source selected. Even in the native strain, *T. thermosaccharolyticum*, consistent *n*-butanol formation has not been reported (Bhandiwad et al., 2013). Following this an alternative approach was considered. The use of *G. thermoglucosidasius* native genes was explored. The native genes are present in the *Geobacillus* genome to carry out alternative reactions, however, it was hypothesised these genes could be overexpressed and exploited for the purpose of *n*-butanol production. The native genes offered an alternative source of compatible enzymes which would be

functional under thermophilic and aerobic conditions. This approach did not generate any final product here and so finally alterations were made to the CoA dependent synthetic pathway. A driving force was introduced in attempt to channel the metabolic flux of the pathway to produce the desirable product. Substrate pools were increased. Various functional genes encoding aldehyde and alcohol dehydrogenase genes were identified to replace the faulty *adhE*. These modifications resulted in 0.137 mM *n*-butanol production. Although a low concentration, this is the first demonstration of *n*-butanol production in *G. thermoglucosidasius* and also in any thermophilic and aerobic conditions. Further work is needed to increase the titre, however this work has demonstrated proof of concept.

4.5.2 Problems encountered

Anaerobic growth

n-butanol production in non-native organisms has been difficult due to oxygen sensitivity of the enzymes employed. Native *n*-butanol production occurs solely in anaerobic organisms, therefore the enzymes lack the inherent oxygen tolerance required here. Despite *G. thermoglucosidasius* reportedly having a functional fermentation pathway, growth studies throughout this work have shown oxygen is required for growth. When cultured under anaerobic conditions no growth or *n*-butanol production occurred. Despite being described in the literature as facultatively anaerobic. All other *Geobacillus* species are obligate aerobes which suggests *Geobacillus* cell processes require oxygen. The genome sequence of *Geobacillus* shows that the production of thiamine, an essential co-factor, involves an oxygen requiring route glycine to iminoglycine rather than the tyrosine to iminoglycine route typically found in obligate and facultative anaerobes. Supplementation of oxygen limited cultures of *G. thermoglucosidasius* with thiamine was shown to enhance growth rate (Hussein et al., 2015). The inability of *Geobacillus* to grow in oxygen limited conditions restricts the available genes which can be utilised for its engineering.

Recombination

Initially, attempts to clone the synthetic *n*-butanol pathway genes into multiple gene operons was problematic. This led to consideration that the genes may be toxic to the cells. This was shown not to be the case, as individual genes were cloned into *Geobacillus*, with four having shown expression, shown in Figure 4.4. Further investigation revealed recombination of the cloning vectors was occurring at sites of low level homology in the plasmid sequence.

Repeat sequences in the synthetic gene structure were enabling the host cells to remove sections of the plasmid, resulting in loss of partial or whole genes. Initially the attempted operon construction contained homologous regions; the synthetic RBS, the regions upstream and downstream of the RBS, the Biobrick scar and flag tags. These repetitive sequences were resulting in unwanted recombination and shuffling of the DNA, when assembled into operons.

A. 

CCGGCTTAGTCGAAAAAGTGCTTGTCAAAGCGGGACAAGGCGTCAAAAAGGG

CGATATTTGGTAGTAATTAATAA TGATGATTAAGGGAGTTGTTTGGAAATATG

ACAGCGACATTTG**AAAAAC**ATGGCAATATTCCACTTTAAAAAATATATTGG

AGGTTGTTTCTAGATTGCAAATTGTCAAAGTTCGATCGTGTGACGTCCACACAAG

ATTTTGC GGAAGCGGTCTCCGAAATGATTGATGGAGATTTTGTGGTTCGTCGCGA

AAGAACAACGAAAGCGAGGGGGCGCTATGGCAGAGAATGGTATAGCCCTCG

TGGCGGACTTTGGGTGACGTATGTGGTCAAAAATTTAATGTCGAAGAAATTG

GCATTAGCACGCTTAAAGTGGCGTTGGCGATTGCAGCATTCTTTCCAAAATGG

TAGATGCGAAAATTCGCTGGCCTAATGATATTGTTAATCATAAAAAAGTAT

CGGGCGTGTTACTTGAAGCGATTACAGTAGGGGAACGCTCGACAGGCTTTATT

GGCTTGGAGTCAATACGAATGTGGTCTCGTTTCCTCAAGGCATTTCCG

B. Subunit 3 sequence:

atgacagcgacatttg***aaaaac***cggatatgtcctaacttgcgaaga.....tgcttaaaata

agcgtgaatttcgctatccta***aaaaac***atggcaatattccacttta**AAAAATATATTGGAGGTTGT**

TTCTAGA

Figure 4.31. Sequencing of recovered plasmid containing the four sub-unit *acc* gene. A. Red; the end of the second subunit, Yellow; RBS' from subunits 3 and

4, Green; partial sequence of subunit 3, Blue; the start of the fourth subunit. **B.** The correct subunit 3 gene sequence with the homologous region where recombination has occurred highlighted red. Full sequences can be found in appendix A9.1.

- A.** CCAAGGCTCTTAAGGTTAATCCTCGAAATATTGGATTGGACGTA CTGTTTAGCA
 AGTAAAAAAAAAGGTTTTGTACGACCCTTATGTCAATTGGATTGAGAAAAGTACA
 TCCAGTGACAAGCTACTTGTTTGGAAATTCAGGAAGTGAACCGATCGTCGTCGA
 TACGTCAGAACGGGTAAAATCAGAGCTCGTATATCCAGATTGATCACGCTGTGG
 AGGAGAAGCGCGATGACGCGCTAAAGAATGTGGTGTTTACGATATCGTATACG
 AGTTCGAACGTTGCTTAAATTCGCGACATTCGCAACGCGCTTCAAATTCCTGA
 ATGGGCAAAAAAGCGTTGTTTCATGGCTATTCCGACAACATCAGCACAGGCTTCG
 GAAGTGACAGCGTTCGCGGTAATCACAGATAAGAAACGCAACATCAAATACCC
 ATTAGCTGATTACGAATTAACACCTGATATCGCTATCATCGACCCTGACTTGACA
 AAAACGGTACCTCCTTCGGTTACAGCGGACACAGGCATGGACGTCTTAACGCAT
 GCTATCGAAGCTTATGTGTCGGTGATGGCGTCGGATTACACAGATGCTTTAGCG
 GAAAAAGCTATCAAATTGGTGTTCGAATACTTACCAAAAAGCTTACAAAAACGGA
 AACGATGAAGTGGCGCGCAAAAAATGCATAATGCGTCGTGCATGGCGGGCA
 TGGCGTTTACAAATGCTTTCTTAGGCATCAACCATAGCATGGCTCATATCTTGGG
 AGGCAAATTCATATCCCGCATGGCCGCGCTAACGCGATCTTGTTACCGTACGT
 GATTAATACAATGCGGAAAAACCGACGAAATTTGTGGCGTATCCACAATACG
 AATATCCTAAAG
- B.** GACAGCGTTCGCGGTAATCACAGATAAAACGCAACATCAAATACCCATTAG
 CTGATTACGA

Figure 4.32. Sequencing of recovered plasmid containing the synthetic *adhE* gene. **A.** Sequencing data of recovered plasmid. **B.** The expected *adhE* gene sequence. Yellow; unknown sequence, Green; partial sequence of the synthetic *adhE* gene. Full sequences can be found in appendix A9.1.

These two examples of anecdotal evidence show recombination occurring at sites of low level homology. Figure 4.31 shows where 6 bp of homology occur, surrounded by asterisk, a recombination event has resulted in loss of 1537 bp of subunit 3. In these two examples poly A sequences are the culprit, with

recombination occurring at sites of 5 A's and 6 A's. The unknown sequence highlighted yellow in Figure 4.32 returned no match from BLAST search. The identity of the inserted sequence cannot be determined. These recombination events are not always seen, in other cloning experiments longer regions of homology, up to 13 bp, do not recombine.

Homologous recombination is a major DNA repair process in bacteria. It is also important for producing genetic diversity in bacterial populations. The *G. thermoglucosidasius* NCIMB 11955 genome has eight recombinase genes annotated; recombination protein RecR, recombination regulator RecX, recombinase, tyrosine recombinase XerC, recombinase RecA, site specific tyrosine recombinase XerD, recombinase RarA and recombinase XerD. Problems encountered during cloning of the plasmids and instability of the *ter* integration may be attributed to high recombination activity of *G. thermoglucosidasius*. Observations show as little as 6 bp homology region is sufficient to become a site of homologous recombination.

To overcome this problem the operon was re-designed. All regions of homology were removed from the plasmid. The synthetic RBS was replaced with alternative RBS' from the screening work, chapter 3.4. Un-tagged gene variants were used. A single promoter was included. Spacer regions were removed. Alternative cloning methods were employed to prevent introduction of the BioBrick scar region.

4.5.3 Further work

Alternative gene candidates

Modifications made to the genes used in the pathway have resulted in improvements. Many other genes candidates are described which could be used to explore further alternatives. By screening other gene options further improvements to the *n*-butanol production could be made. Examples of some of the possible promising alternative genes which could be screened include;

atoB acetyl-CoA acetyltransferase. This gene has high specificity for short-chain acyl-CoA molecules. *YqeF* is a predicted acyltransferase with high sequence similarity to *AtoB*. *fucO* encodes L-1,2-propanediol oxidoreductase, an aldehyde/alcohol dehydrogenase from *E. coli* with high sequence and structural similarity to clostridial butyraldehyde/butanol dehydrogenase. Overexpression of *fucO* in *E. coli* led to higher concentration of *n*-butanol and lower concentration of ethanol (Dellomonaco et al., 2011). Simultaneous overexpression of *yqeF* and *fucO* in *E. coli* has been shown to yield high *n*-butanol titre of 1.9 g/l (Dellomonaco et al., 2011).

fabB is a (S)-3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase gene from an active fatty acid degradation operon of *Cupriavidus necator*. FadB is an enzyme with two catalytic domains exhibiting a single monomeric structure. The C-terminal harbours enoyl-CoA hydratase activity and is able to convert trans-crotonyl-CoA to (S)-3-hydroxybutyryl-CoA. The N-terminal comprises an NAD(+) binding site and is responsible for (S)-3-hydroxyacyl-CoA dehydrogenase activity converting (S)-3-hydroxybutyryl-CoA to acetoacetyl-CoA. FadB is strictly stereospecific to (S)-3-hydroxybutyryl-CoA (Volodina and Steinbüchel, 2014). FabB (87 kDa) is able to complete two reaction steps in the *n*-butanol pathway. This enzyme is a potential replacement for both Hbd and Crt (total 63 kDa).

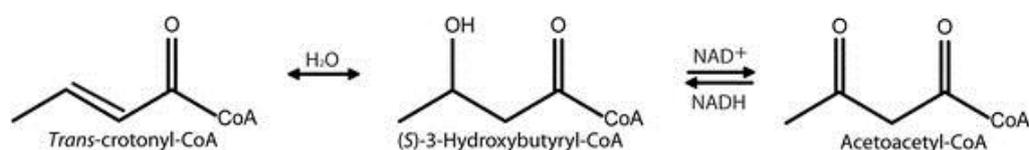


Figure 4.33. The enoyl-CoA hydratase and (S)-3-hydroxyacyl-CoA dehydrogenase reactions of the fatty acid degradation pathway, catalyzed by FadB (Volodina & Steinbüchel, 2014).

Stereospecificity

The *n*-butanol pathway intermediate 3-hydroxybutyryl-CoA, is an enantiomeric molecule. The molecule can be one of two forms; (R)-isomer or (S)-isomer, dependent on its stereochemistry. The engineered pathway described in this work utilises the S form as an *n*-butanol intermediate. In addition to the desired reaction, the S form of the molecule is also used in other cellular reactions. The intermediate will be utilised by the cell for other metabolic pathways, reducing the substrate pool available for crotonase. The R form, however, does not naturally occur in the cell. It is expected that the R form would only be utilised for the desired purpose. The *phaB* gene, from *Cupriavidus necator* (formerly known as *Ralstonia eutropha*), encodes acetoacetyl-CoA reductase which catalyses the chiral NADPH-dependent reduction of acetoacetyl-CoA to (R)-3-hydroxybutyryl-CoA (Kim et al., 2014).

CoA DEPENDENT n-BUTANOL PRODUCTION

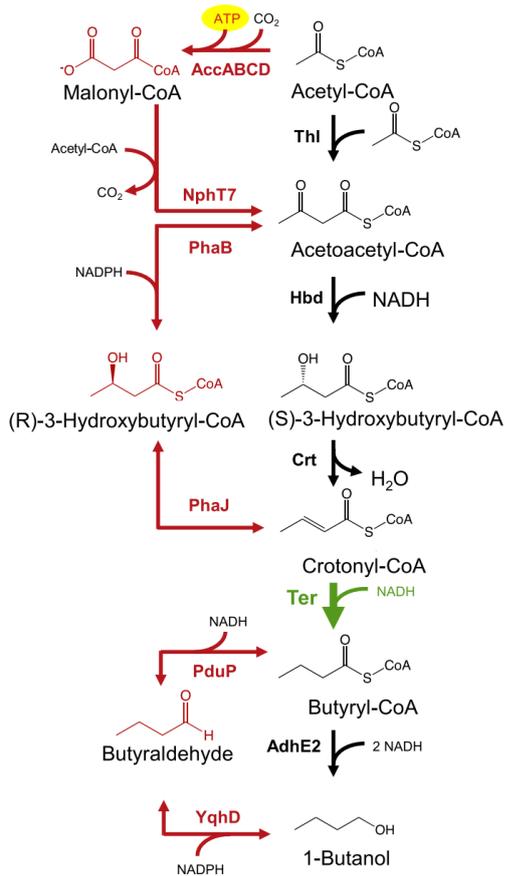


Figure 4.34. Future engineering the CoA-dependent *n*-butanol production pathway for possible higher efficiency (Chen & Liao, 2016). The pathway could be further adapted for *n*-butanol production by replacing *hbd* and *crt* with *phaB* and *phaJ*. Use of the *phaB* gene instead of *hbd* to generate the R form rather than the S form of 3-hydroxybutyryl-CoA. *phaJ* is the corresponding R specific crotonase.

Knock out of competing pathways

For feasible biofuel production extremely high volumes and a low cost are needed. This means high yield and high productivity are required. In attempt to increase yield, titre and productivity competing pathways that drain precursors, products or cofactors should be eliminated or reduced a minimum. In addition precursor supply should be increased. In the *n*-butanol pathway used here, the intermediate molecules are used by the cell for other reactions. Diverting the pathway away from the desired product. Transcriptional down

regulation of such by pass pathways or traditional gene knockout methods could focus the metabolic flux towards *n*-butanol biosynthesis (Na et al., 2013).

Geobacillus strain LS242, described previously in section 4.2.5, with ΔIdh , Δpfl and pdh^{UP} should be used in combination with the final production pathway described here. Micro aerobic growth of *G. thermoglucosidasius* results in production of lactate, acetate, ethanol and formate. Deletion of lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*) will reduce the competing pathways. Enhancement of *pdh* expression using the strong P_{ldh} , will increase the conversion of pyruvate to acetyl CoA. Resulting in decreased pyruvate accumulation and an increase in the substrate pool. In addition a further deletion, $\Delta adhE$, offers further improvement to the host strain by also removing ethanol production. Future strain modifications should target disruption of acetic acid. The putative genes responsible for acetic acid production; phosphate acetyltransferase (*pta*) and acetate kinase (*ack*) have been identified as targets for deletion. By knocking out the competing pathways carbon flux will be redirected. *n*-butanol production should increase and mixed product production will be minimised. In addition to these knock-outs the final strain should include the *acc* integration and *ter* integration, as described previously.

CoA DEPENDENT n-BUTANOL PRODUCTION

CHAPTER 5

CoA INDEPENDENT *n*-BUTANOL PRODUCTION

5.1 Introduction

In the previous chapter *n*-butanol production was demonstrated via a series of CoA dependent reactions. Here an alternative route is considered. In the forthcoming work a CoA independent pathway is explored. This pathway offers an oxygen tolerant route to *n*-butanol.

The CoA dependent pathway is reliant on the O₂ sensitive AdhE enzyme to catalyse the sequential reduction of both butyryl-CoA and butyraldehyde. High yields of butanol have, to date, only been demonstrated in strictly anaerobic fermentations. As the host strain in this work, *Geobacillus*, is an aerobic organism demonstrating poor growth in oxygen limited environments, an alternative oxygen tolerant *n*-butanol pathway is desired.

The proposed O₂ tolerant *n*-butanol pathway is based on the activity of an ACP-thioesterase, acting on butyryl-ACP in the native fatty acid biosynthesis pathway. This approach will utilise a chain length specific thioesterase to interrupt the host's native fatty acid biosynthesis pathway, producing butyric acid as an intermediate. Butyric acid can be further converted to *n*-butanol. This approach has been demonstrated in *E. coli*, where 300 mg/l *n*-butanol was produced in 24 h from batch growth conditions in shaking flasks (Pasztor et al., 2014). To our knowledge, to date, this approach has not been investigated in a thermophilic organism.

5.2 Fatty acid metabolism

Fatty acids are an integral part of all living organisms, and are generally composed of a hydrophobic hydrocarbon chain ending in one hydrophilic carboxylic acid functional group. The metabolic pathway of fatty acid metabolism in organisms is well-studied (Figure 5.1). Fatty acids are commonly built via de novo synthesis and elongation. Figure 5.1 shows that the de novo fatty acid synthesis starts from the primer acetyl-CoA and the extender malonyl-CoA through a cyclic series of reactions catalyzed by fatty acid

synthases. The synthesized fatty acids are almost entirely composed of even-length and straight carbon chains that have various numbers of carbon atoms (<6, short chain; 6–12, medium chain; >14, long chain) with differing degrees of unsaturation (saturated, mono unsaturated, and polyunsaturated).

Fatty acid derivatives can be converted to desirable value-added chemicals through metabolic engineering. Hydroxy fatty acids, fatty alcohols, fatty acid methyl/ethyl esters and fatty alka(e)nes have a wide range of industrial applications including plastics, lubricants and fuels. Currently, these chemicals are mainly obtained through chemical synthesis, which is complex and costly. Their availability from natural biological sources is extremely limited. Metabolic engineering of microorganisms has provided a platform for effective production of these valuable biochemicals.

CoA INDEPENDENT n-BUTANOL PRODUCTION

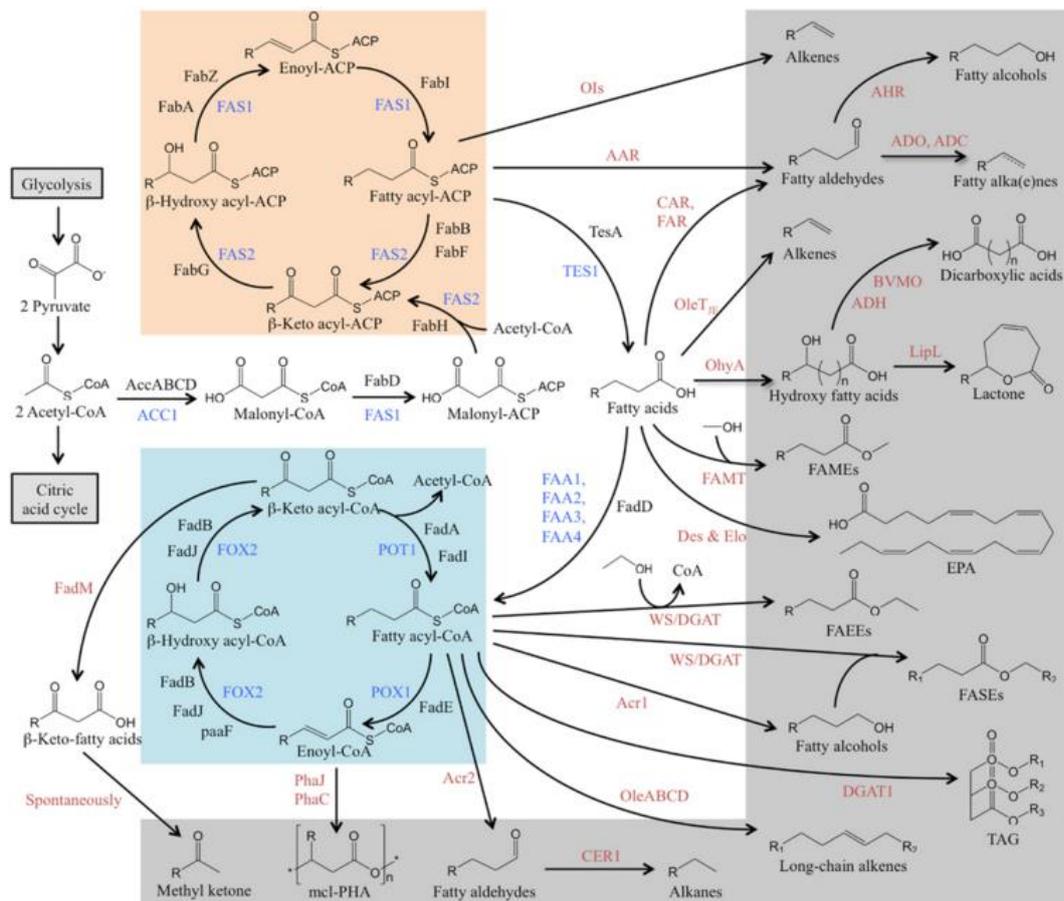


Figure 5.1. Overview of metabolic pathways that lead to the production of fatty acids and fatty acid-derived chemicals (Yu et al., 2014). The fatty acid biosynthesis (orange), β -oxidation cycle (blue) and the biosynthesis pathway of fatty acid-derived chemicals (grey) are presented. The enzymes of fatty acid metabolism in *S. cerevisiae* are shown in blue, in *E. coli* are shown in black and the enzymes for conversion of fatty acids to their derivatives from other organisms is in red.

5.3 Production of Short Chain Fatty Acids

Short-chain fatty acids (SCFAs) have a broad range of applications in chemical and fuel industries. Butyric acid is the SCFA precursor to *n*-butanol. Currently butyric acid is produced by chemical synthesis from crude oil by oxidation of butyraldehyde. The chemical production of butyric acid is costly and non-ecofriendly (Jawed et al., 2016). There is, therefore, a need to produce butyric acid from renewable carbon sources using a microbial platform to replace

chemical synthesis. As is the case with *n*-butanol, native producers of butyric acid, including *C. tyrobutyricum*, *C. beijerinckii* and *C. thermobutyricum*, are all strictly anaerobic organisms. A production approach in non-native microorganisms is to exploit the native and aerobically active fatty acid biosynthesis cycle, shown in Figure 5.1 in orange.

5.3.1 Thioesterase selection

For production of SCFAs from the native fatty acid synthesis pathway, a thioesterase enzyme is required. *De novo* fatty acid biosynthesis is an iterative process of chain extension, with acetyl-CoA reacting with malonyl-ACP. Acyl-acyl carrier protein thioesterases catalyze the hydrolysis of the thioester bond that links the acyl chain to the sulfhydryl group of ACP. This reaction terminates acyl chain elongation of fatty acid biosynthesis, releasing free fatty acid and ACP. Thus the thioesterase determines fatty acid chain length. By identification and selection of a chain length specific thioesterase the fatty acids can be generated to a desired chain length. A thioesterase with C4 specificity will act on butyryl-ACP to generate butyric acid. This prevents butyryl-ACP entering the next turn of an elongation cycle where the chain length would be increased by 2 carbons.

Jing et al. (2011) screened 31 thioesterase enzymes from a range of different organisms in order to understand the diversity in enzymatic specificity and activity. The study determined the *in vivo* substrate specificity of the enzymes in an *E. coli* K-27 host strain lacking FadD (fatty acyl-CoA synthase). Table 5.1 shows the fatty acid production profiles of the thioesterases from bacterial sources. The screening showed the enzymes varied greatly in both total fatty acid production and in the concentrations of fatty acids of differing chain lengths. Many of the thioesterases have broad-range substrate specificity, capable of acting on short and medium carbon chain length acyl-ACPs producing the corresponding fatty acids. This specificity for different chain

length fatty acids can be utilized in attempt to engineer production of a desired molecule.

For further investigation in *G. thermoglucosidasius*, the thioesterase from *Bacteroides thetaiotaomicron* (TesBT) was selected. Highlighted yellow in Table 5.1. *tesBT* was selected as the best candidate gene for butyric acid production as it produced by far the greatest percentage of C4 fatty acids. With 27.4% of the total fatty acids being 4:0 (Jing et al., 2011).

A *Geobacillus* thioesterase was analyzed in the screening work carried out by Jing et al. (2011). The *Geobacillus* thioesterase gene was not selected for overexpression in this work as it produced a low percentage of C4 fatty acids. The *Geobacillus* thioesterase generated primarily longer chain 12:1 and 14:1 unsaturated fatty acids. This indicates the *Geobacillus* thioesterase screened, may not be an optimal choice for production of butyric acid.

Table 5.1. Molar percentages and total concentrations of fatty acids produced by different thioesterases in *E. coli* (Jing et al., 2011).

ACC NO.	Organism	Total FA (nmol/ml)	Percentage of individual FA (mol %)										
			4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1
ACL08376	<i>Desulfovibrio vulgaris</i>	330±9	0.25±0.06	1.56±0.20	28.8±1.2	3.47±0.18	0.41±0.03	7.88±0.28	24.2±1.4	5.95±0.37	23.6±1.1	1.19±0.35	2.6±0.3
CAH09236	<i>Bacteroides fragilis</i>	215±6	13.1±0.8	2.72±0.23	20.2±1.5	2.66±0.22	0.32±0.07	3.64±0.38	19.2±1.4	5.05±0.30	25.4±1.2	2.23±0.34	5.44±0.40
ABR43801	<i>Parabacteroides distasonis</i>	70.3±4.4	0.40±0.25	1.29±0.74	18.0±4.5	6.28±0.39	nd	16.3±1.1	9.32±0.79	21.3±2.0	27.1±2.1	nd	nd
AAO77182	<i>Bacteroides thetaiotaomicron</i>	60.4±2.9	27.4±3.2	1.64±0.37	13.4±0.8	2.09±0.15	nd	4.59±0.68	16.7±0.9	6.06±1.08	25.6±1.4	nd	2.59±0.25
ABG82470	<i>Clostridium perfringens</i>	72.0±9.5	3.02±0.67	14.0±2.5	70.3±4.4	3.03±0.51	0.10±0.03	nd	1.05±0.19	nd	8.45±2.18	nd	nd
EEG55387	<i>Clostridium asparagiform</i>	25.9±4.2	1.92±0.69	4.45±1.52	26.0±5.7	5.51±1.41	nd	6.69±2.15	1.61±0.82	35.0±9.2	17.5±5.8	nd	1.27±0.87
EET61113	<i>Bryantella formatexigens</i>	381±3	15.0±0.3	20.4±0.2	31.8±0.3	5.08±0.13	0.70±0.02	4.30±0.15	8.88±0.46	1.85±0.19	10.5±0.2	0.40±0.24	1.17±0.11
EDV77528	<i>Geobacillus sp.</i>	64.9±12.0	2.35±1.70	1.09±0.57	8.58±4.44	2.42±1.08	2.57±1.43	6.76±3.13	30.8±10.6	11.0±3.4	31.7±8.2	0.30±1.11	2.37±1.22
BAH81730	<i>Streptococcus dysgalactiae</i>	624±14	3.92±0.18	13.2±0.6	29.9±1.1	5.02±0.18	0.80±0.02	5.73±0.33	13.5±0.7	4.44±0.25	20.0±1.1	0.25±0.14	3.30±0.14
ABJ63754	<i>Lactobacillus brevis</i>	710±10	7.05±0.29	13.7±0.3	55.5±0.7	2.58±0.06	0.73±0.01	3.75±0.07	7.94±0.19	1.85±0.14	6.27±0.19	nd	0.68±0.14
CAD63310	<i>Lactobacillus plantarum</i>	436±10	3.09±0.28	11.0±0.4	68.0±0.8	1.24±0.06	0.08±0.01	2.81±0.18	4.63±0.21	1.87±0.11	6.87±0.42	nd	0.45±0.06
EEI82564	<i>Anaerococcus tetradius</i>	1381±146	0.53±0.07	1.35±0.19	86.7±1.5	2.18±0.30	0.46±0.08	1.13±0.18	2.79±0.48	1.18±0.23	3.01±0.54	0.05±0.09	0.66±0.11
CAE80300	<i>Bdellovibrio bacteriovorus</i>	333±18	0.10±0.09	0.86±0.14	36.9±3.1	3.26±0.45	0.44±0.07	6.65±0.66	7.56±0.64	8.20±0.73	27.8±2.1	1.60±0.22	6.57±0.49
ABN54268	<i>Clostridium thermocellum</i>	97.7±3.2	0.59±0.14	0.75±0.24	8.36±0.37	4.50±0.16	0.27±0.01	2.66±0.23	7.94±0.38	9.84±0.77	59.5±1.4	0.83±0.67	4.74±0.42

Table 5.1 shows mean \pm standard error ($n = 4$), nd = not detected. In the work to produce this data the genes were codon-optimized for expression in *E. coli*.

5.3.2 Expression in *Geobacillus*

Gene cloning

The *tesBT* gene was obtained with thanks from the Yazdani lab, ICGEB, India. The DNA was codon optimised for optimal expression in *E. coli*. The gene included a C-terminal 6xHis-tag. The gene was sub-cloned into the *Geobacillus* shuttle vector pMTL61122, driven by P_{ldh} . PCR was used to introduce the synthetic RBS and the BioBrick prefix restriction enzyme sites to the 5' and the BioBrick suffix restriction enzyme sites to the 3' of the gene. The addition of BioBrick restriction enzymes enabled ligation of the *tesBT* gene into the vector backbone.

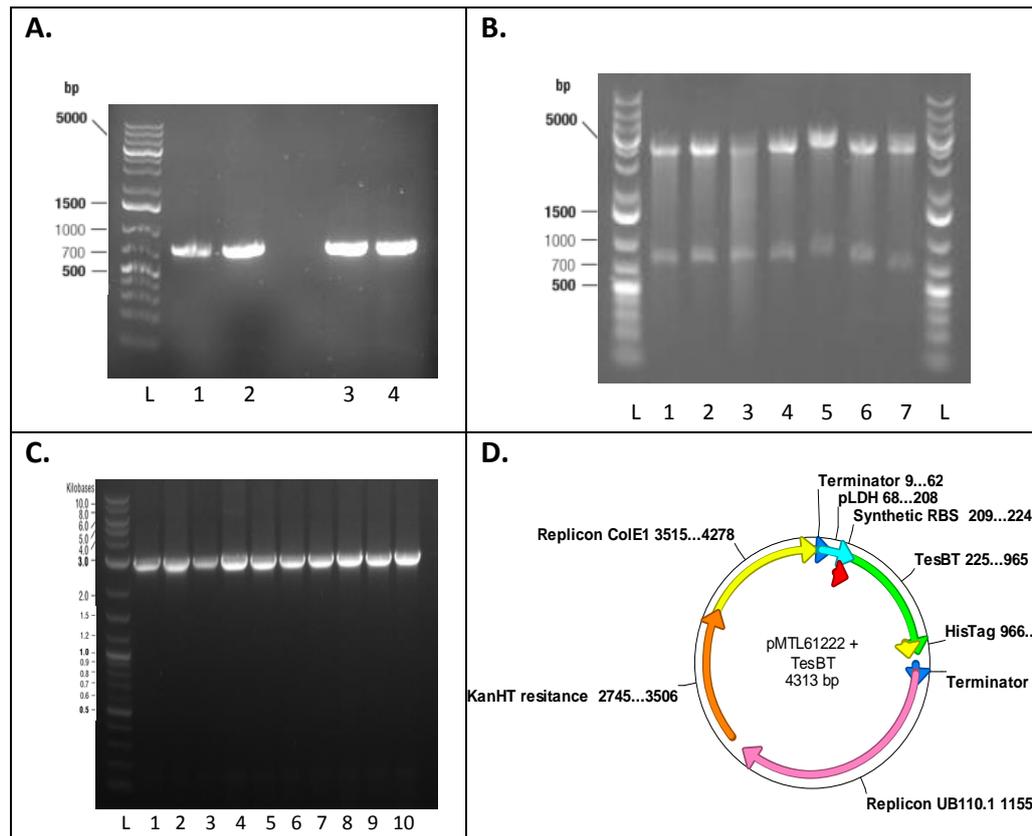


Figure 5.2. Cloning for construction of the pMTL61222+*tesBT* vector. A. Visualization of PCR products run on 1% agarose gel. Photographed under UV light, bands visible due to binding of ethidium bromide. Lanes 1 to 4 show PCR amplification of the *tesBT* gene. The gene plus addition of RBS and restriction enzyme sites is 796 bp in size. Lane L is O'GeneRuler 1 kb Plus DNA ladder (Thermo Scientific). The PCR was carried out using Phusion polymerase. Cycling conditions; 98°C for 2 minutes, 35x cycles of; 98°C for 30 seconds, 72-60°C for 30 seconds and 72°C for 30 seconds, followed by a final extension of 72°C for 10 minutes and 4°C hold. Gradient PCR annealing temperatures: Lanes 1 and 2; 55°C, lanes 3 and 4; 65°C. The PCR primers used are detailed in appendix A1. 5 µl of each reaction was run on the gel to confirm amplification at the correct size. **B.** Confirmatory digestion of screened *E. coli* colonies. DNA preparations from *E. coli* transformants were digested using *Spe*I and *Pst*I restriction enzymes. Two bands of size; 3533 and 796 bp, show the vector and insert fragments, respectively. **C.** Colony PCR of 10 *Geobacillus* transformants. The PCR primers used are detailed in appendix A1. Colony PCR using DreamTaq

CoA INDEPENDENT n-BUTANOL PRODUCTION

MasterMix (Thermo Fisher Scientific) amplified a target region of 3 kbp on the pMTL61222+*tesBT* vector. With the forward primer located on the vector backbone and the reverse primer located in the *tesBT* gene, confirming the gene has been introduced into the *Geobacillus* host. **D.** Plasmid map of the pMTL shuttle vector containing the *tesBT* gene. The vector was constructed, as described, before transformation into *Geobacillus* for expression.

The PCR reactions shown in lanes 3 and 4 of Figure 5.2 A. were processed using a PCR clean-up kit (Thermo Scientific). Eluted DNA was pooled the concentration measured using Nanodrop. 165 ng/μl of DNA was recovered. The pMTL61122 vector (564 ng/μl) and insert DNA were digested according to the BioBrick procedure. The fragments were ligated and transformed into *E. coli*. Transformed colonies were grown overnight, a miniprep kit was used to recover the plasmid DNA which was further digested to confirm the *tesBT* insert presence, shown in Figure 5.2 B. Two aliquots of plasmid DNA preparation, which were confirmed to digest with the correct band sizes, lanes 1 and 2 of Figure 5.2 B., were used to transform *G. thermoglucosidasius*. Colony PCR of *G. thermoglucosidasius* transformants was performed to confirm successful introduction of the vector into the final host strain. Figure 5.2 C. shows all ten colonies screened were positive clones, confirming presence of the *tesBT* gene.

Following confirmation of successful cloning and introduction of the pMTL61222+*tesBT* vector into *Geobacillus*, shown in Figure 5.2, protein expression and enzyme activity were assayed. The assays were carried out to determine if the *tesBT* gene was functional.

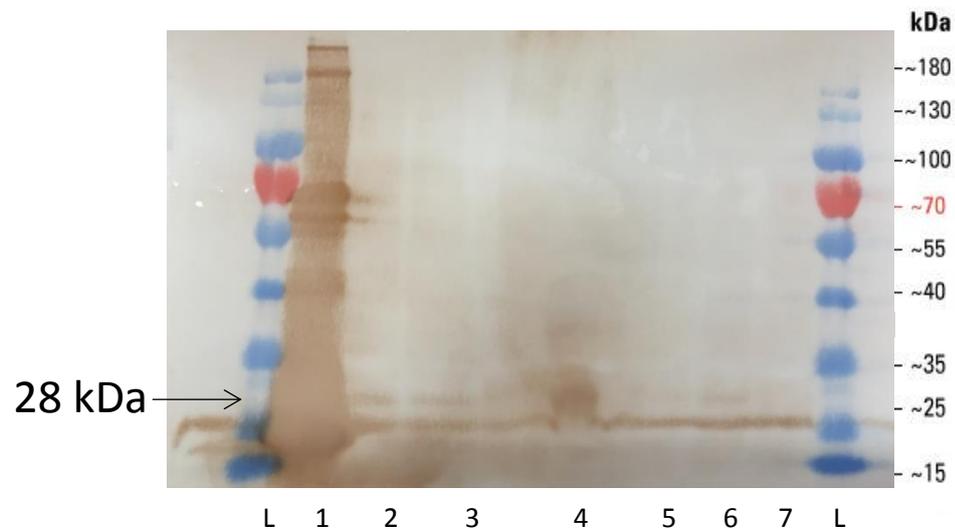
Western blot analysis

Figure 5.3. Western blot of *Geobacillus* expressing the TesBT protein. *G. thermoglucosidasius* harbouring pMTL61222+*tesBT* were prepared for Western blot as described in section 2.7.2, with cultures adjusted to the concentrations indicated. Here Mouse-anti-Penta-His (Qiagen) and anti-mouse-HRP (Promega) antibody was used. Lanes L; PageRuler prestained protein ladder (Thermo Fisher), 1; positive control, a known functional His-tagged protein, 2 and 3; supernatant at 0.8 OD concentration, 4; pellet, 5 and 6; supernatant at 0.4 OD concentration, 7; wild type supernatant at 0.8 OD concentration. Wild type cells were included as a negative control.

The protein visualisation by Western blotting shown in Figure 5.3 confirms plasmid based expression of the thioesterase gene *tesBT* in *Geobacillus*. Analysis of the soluble fraction showed some expression, with bands of 28 kDa visible in lanes 2, 3, 5 and 6 of Figure 5.3. However a stronger band present in lane 4 of the Figure indicates the protein is accumulating in the insoluble fraction as inclusion bodies. This suggests the protein is aggregating in the cytoplasm. Aggregates are formed by the association of partially folded protein or misfolded proteins. This indication of misfolded protein accumulation suggests the gene is not compatible with a *Geobacillus* host, as only correctly folded proteins are functional.

Enzyme assay

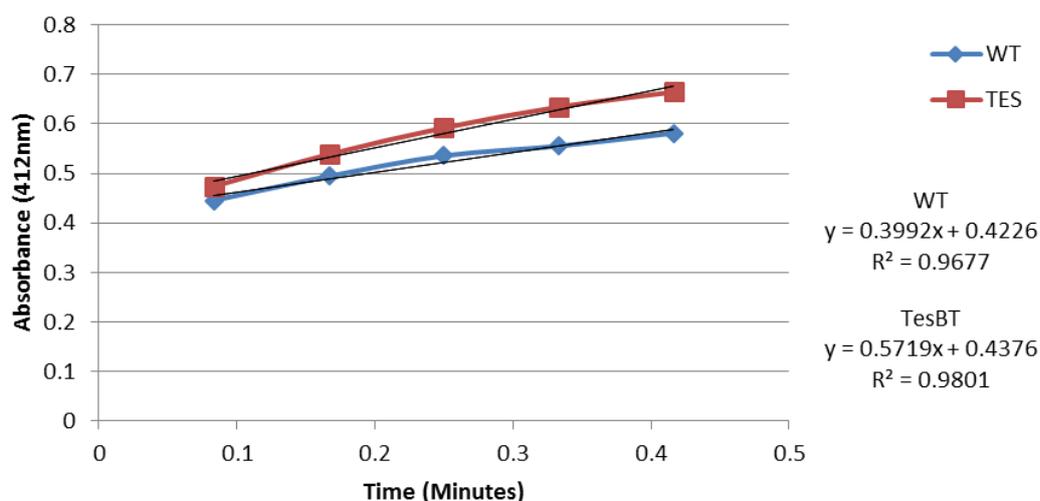


Figure 5.4. Thioesterase enzyme activity assay. Enzyme activity of *G. thermoglucosidasius* harbouring pMTL61222+tesBT was determined. The culture was grown in 2SPYNG at 52°C before protein extraction by sonication. 200 µg protein was measured. The assay was performed at 52°C. Briefly; the reaction mixture contained 60 mM potassium phosphate buffer (pH 7.4), 100 µM DTNB, 0.08 mg bovine serum albumin, 20 µM butyryl-CoA and crude enzyme from cell extract. The reaction was started with the addition of enzyme. Reduction of DTNB by the CoA liberated in the thioesterase reaction was measured spectrophotometrically by an increase in absorbance at 412 nm.

The enzyme assay result shown in Figure 5.4, quantifies TesBT activity as 63.5 nmole/min/mg. Where one unit of enzyme activity is defined as the amount of enzyme catalyzing the cleavage of 1 nmol of acyl-CoA per minute. This represents a significant level of activity. Jawed et al. (2016) reported activity of 2.91 nmole/min/mg for the same TesBT enzyme in an *E. coli* host. Activity of 2.91 nmole/min/mg produced a butyric acid titre of 1.46 g/l (Jawed et al., 2016).

5.3.3 Culture for butyric acid production

In order to determine if the addition of plasmid based *tesBT* resulted in butyric acid production, the *G. thermoglucosidasius* strain harbouring pMTL61222+*tesBT* was cultured for product analysis.

Glycerol was initially selected as the carbon source. In addition to providing a carbon source, glycerol is an essential precursor for the synthesis of lipids. Presence of glycerol in the culture medium could upregulate the fatty acid biosynthesis pathway. Butanol production from butyric acid requires one ATP and two NADH. Therefore an optimal electron donor satisfying both requirements is needed. Mattam and Yazdani (2013) compared glucose and glycerol as energy and electron source for butanol production. They found, in *E. coli*, butanol yield with respect to butyric acid was similar (>85% of theoretical maxima) for either substrate. However, butanol yield with respect to glycerol was approximately doubled as compared to that of glucose (Mattam & Yazdani, 2013).

G. thermoglucosidasius pMTL61222+*tesBT* was grown in AYSE medium with 1% glycerol. Cultures were incubated aerobically at 52°C shaking at 250 rpm for 72 hours. Samples were collected for HPLC analysis at 12 h intervals throughout. The HPLC data was analysed for presence of a peak at the butyric acid retention time of 44.834 minutes.

The initial experimentation using AYSE medium with 1% glycerol showed no butyric acid production. Following this result a range of further media were screened; AYSE with 1% glucose, 2SPY with 0.4% and 0.8% glycerol, Terrific broth with 0.4% and 0.8% glycerol, LB with 0.4% and 0.8% glycerol, minimal media; AYSE with no yeast extract, UYSE with no yeast extract and MOPS. The culture conditions as described previously, were used. This media screening was carried out as growth medium plays a key role in energy metabolism of an

organism. Culture of the strain in optimal conditions with the required nutritional components could be essential in achieving product formation.

The subsequent culture experiments again showed no butyric acid production, across the range of growth media tested. The HPLC data showed production of lactate, acetate, succinate and ethanol. No butyric acid or *n*-butanol was detected. Analysis of the cultures grown in minimal media showed low OD and no glycerol consumption. This indicates minimal media is insufficient for growth and metabolism of *G. thermoglucosidasius*.

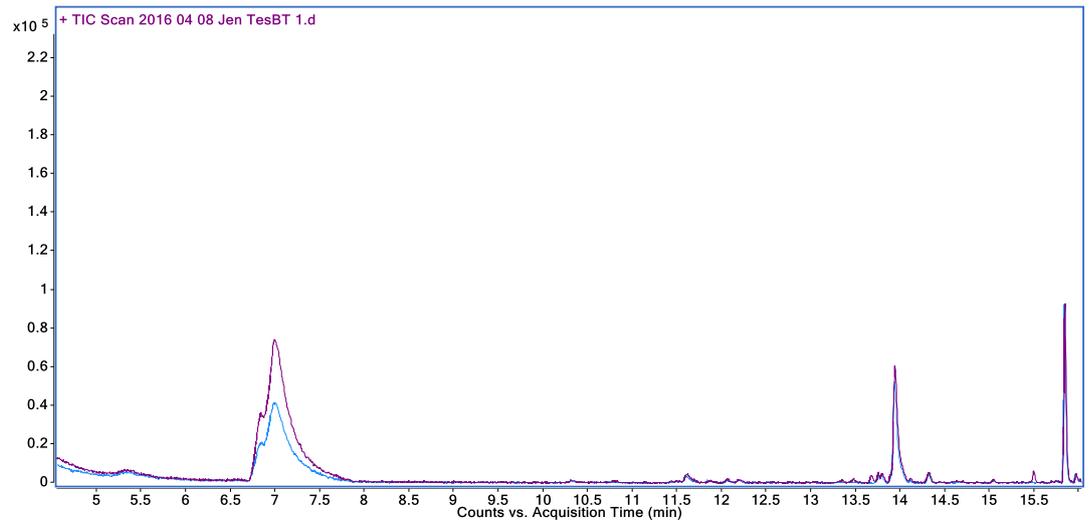
5.3.4 Fatty acid profiling

Following no detection of butyric acid from cultures of *G. thermoglucosidasius* pMTL61222+*tesBT*, analysis of free fatty acids was carried out to determine if the addition of the thioesterase gene altered the fatty acid composition of the strain.

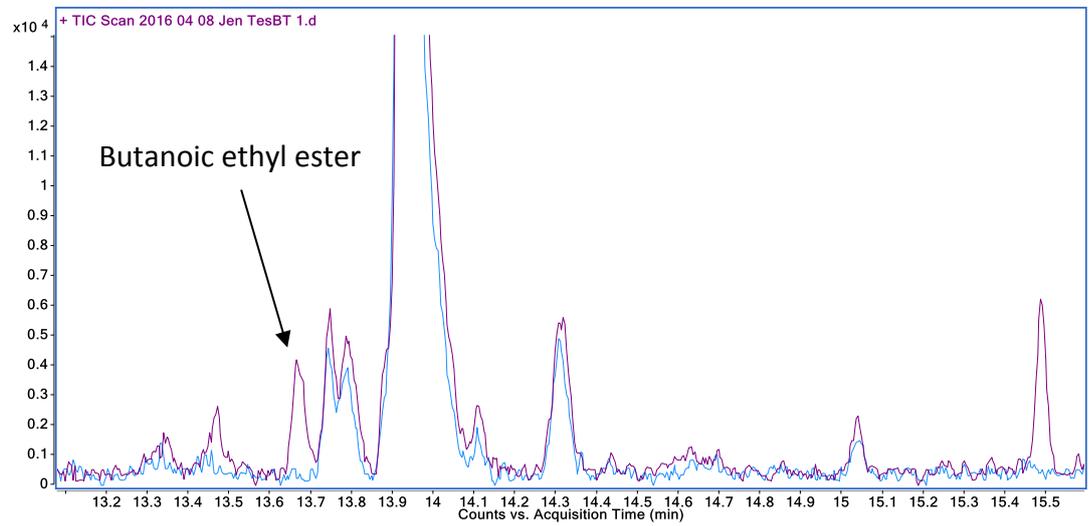
Fatty acid ethyl ester extraction of wild type *G. thermoglucosidasius* and *G. thermoglucosidasius* pMTL61222+*tesBT* strains was carried out, as previously described in section 2.11.3. The extraction and analysis was carried out in duplicate. The duplicate samples showed the same profile, with one selected for presentation here.

CoA INDEPENDENT n-BUTANOL PRODUCTION

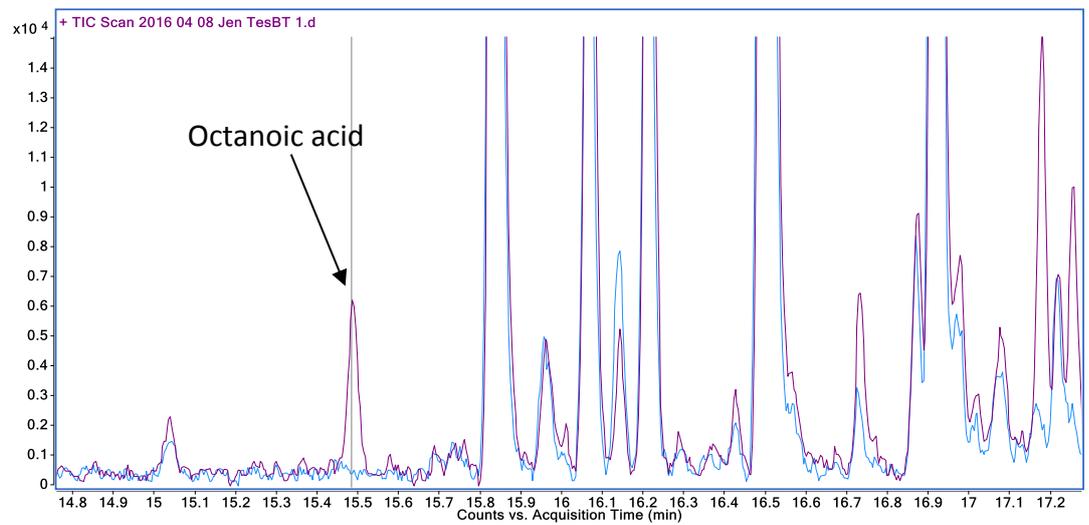
A Full profile



B

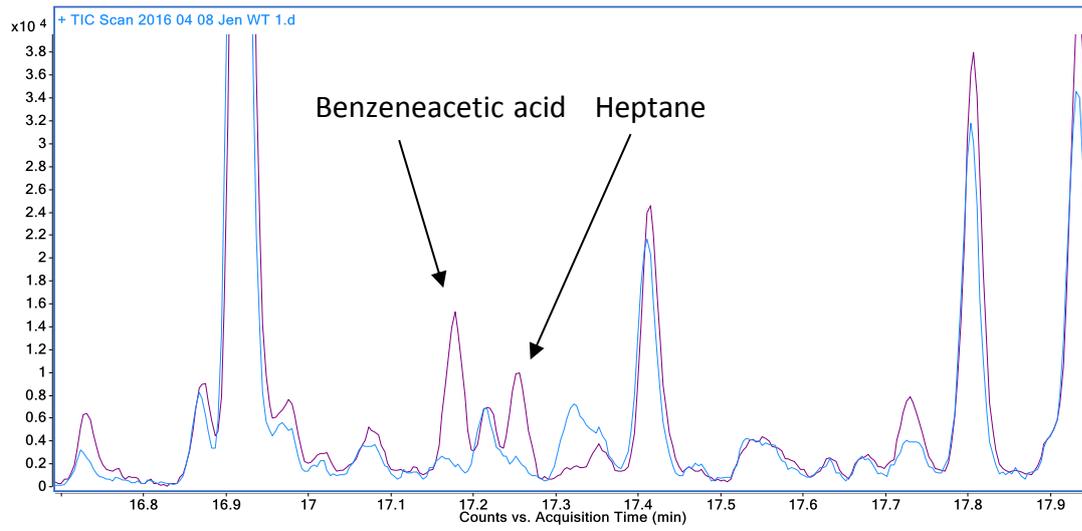


C

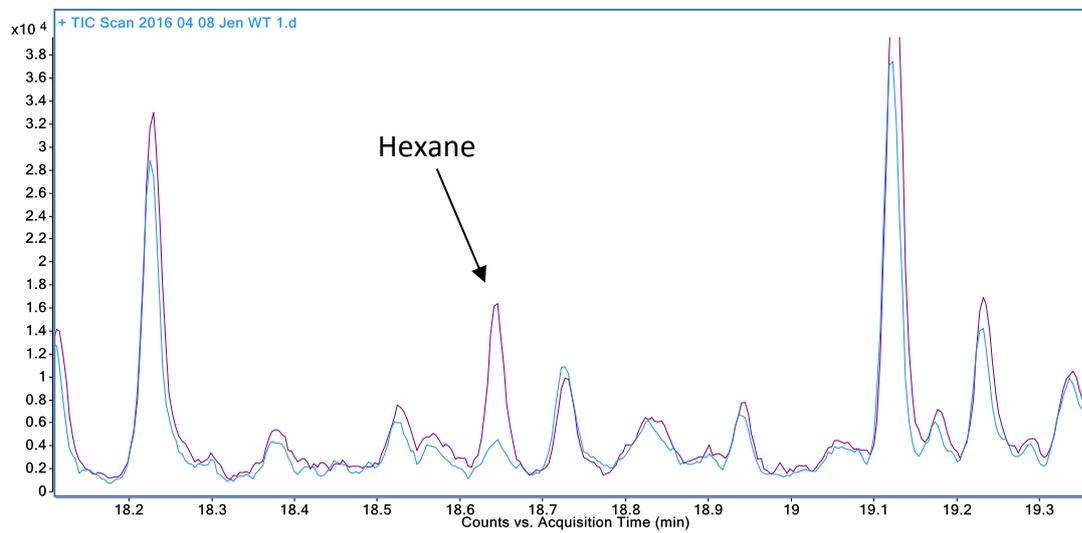


CoA INDEPENDENT n-BUTANOL PRODUCTION

D



E



F

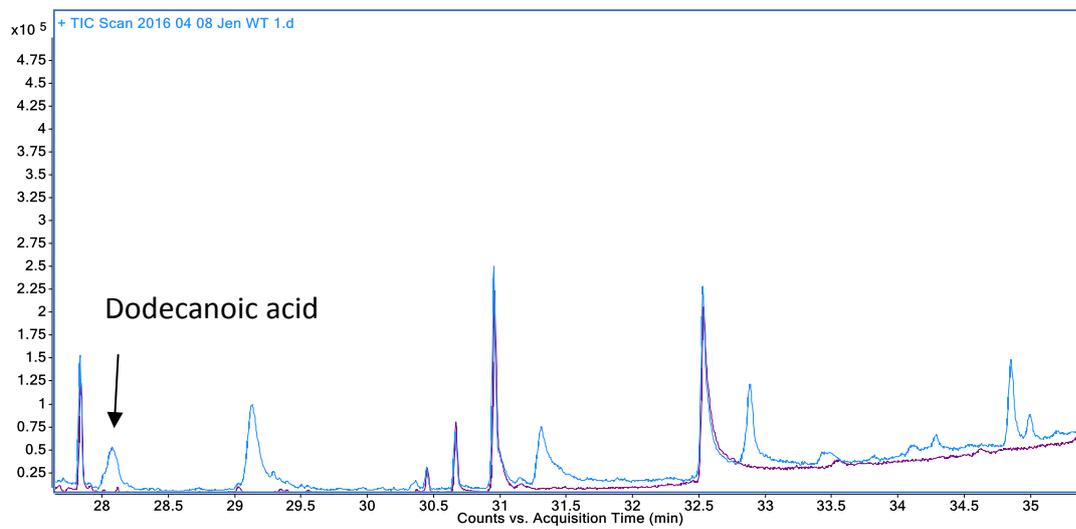


Figure 5.5. Free fatty acid profiles of *G. thermoglucosidasius* wild type and *G. thermoglucosidasius* pMTL61222+tesBT cells. Fatty acids were analyzed by GC-MS in full scan mode using Agilent 7890 GC with an Agilent 7000 Triple Quadrupole GC-MS system. The column used was Zebron ZB-5MS (length: 30 m; diameter: 0.25 mm; film thickness: 0.25 μ m). The injection volume was 1 μ L without split mode and oven temperature ramped from 40 to 325°C (initial ramp at the rate of 5°C per min from 40 to 75°C and then 10°C per min until 325°C). The total run time was 41 min. Compound identities were determined via GC retention times of known standards (FAEES mix from Sigma Aldrich) and verified by mass spectra using the National Institute of Standards and Technology (NIST) reference library. Wild type profile is shown as the blue trace. The TesBT strain is shown as the purple trace. **A.** Full profile, **B.** Butanoic ethyl ester peak highlighted, **C.** Octanoic acid peak highlighted, **D.1,** Benzeneacetic acid ethyl ester peak highlighted, **D.2,** Heptane peak highlighted, **E.** Hexane peak highlighted, **F.** Dodecanoic acid peak highlighted.

The comparison of GC-MS traces shown in Figure 5.5 highlight the differences in the fatty acid composition of wild type and TesBT *G. thermoglucosidasius* cells. These differences indicate expression of the *tesBT* gene is having an impact on the host strain. Image A shows the full profile, with the prominent peaks being C7, C14 and C17 internal standards. The subsequent images show regions of the profile in more detail. Image B shows a butanoic/butyric ethyl ester peak. This confirms butyric acid is being produced by *G. thermoglucosidasius* pMTL61222+tesBT, with none detected from the wild type trace. Image F shows the wild type strain's blue trace has many larger compounds not seen in the TesBT strain, including dodecanoic acid, the peak highlighted. This gives further indication of the action of the TesBT enzyme in interrupting fatty acid synthesis. By termination of acyl chain elongation, this prevents synthesis of the longer chain fatty acids.

5.3.5 Supplemented culture

The GC-MS analysis of free fatty acids in wild type and TesBT *Geobacillus* strains, showed butyric acid was formed in the strain expressing thioesterase. However, no butyric acid was detected by HPLC analysis of the same culture. This could indicate the bacteria are able to utilise the butyric acid. If the acid is being formed, then subsequently converted to another compound, it may not be detected by HPLC analysis.

Geobacillus wild type was cultured in 2SPYNG supplemented with butyric acid 1 g/l. HPLC analysis of the culture showed no reduction in the butyric acid concentration over 48hrs. This indicates the bacteria are not generating then subsequently metabolising the acid.

5.4 Conversion of butyric acid to *n*-butanol

The route to *n*-butanol via a CoA independent pathway requires multiple genes in a set of reactions. Following generation of butyric acid by a thioesterase, additional genes are required for subsequent conversion of butyric acid to the corresponding alcohol *n*-butanol.

Here the full pathway was introduced into *G. thermoglucosidasius* in a plasmid-based operon. In attempt to direct the metabolic flux through the pathway. As the intermediates in this *n*-butanol synthesis pathway are common to many cellular reactions they may be utilized by the cell for other, native processes. By overexpressing the complete pathway in an operon, high levels of enzymes in close proximity may be sufficient to prevent diversion pulling the intermediates through the pathway to form the desired end product *n*-butanol.

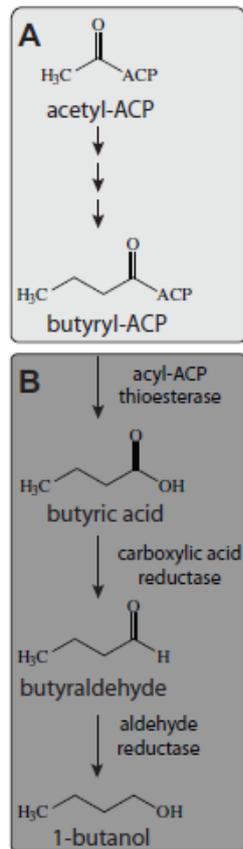


Figure 5.6. Oxygen tolerant *n*-butanol synthesis (Pasztor et al., 2014). A. Conversion of acetyl-ACP to butyryl-ACP carried out by the host's native bacterial fatty acid synthesis. B. Introduction of a synthetic heterologous ACP pathway for conversion of butyryl-ACP to *n*-butanol. This pathway requires; acyl-ACP-thioesterase, carboxylic acid reductase and aldehyde reductase genes.

Thioesterase

The *tesBT* gene from *Bacteroides thetaiotaomicron*, as discussed in section 5.3.1, was selected for further investigation. The work previously discussed demonstrated some activity of the enzyme in *Geobacillus*, here the gene is codon optimized for *Geobacillus*, to improve compatibility with the host.

Carboxylic acid reductase (Car)

An oxygen tolerant carboxylic acid reductase (Car) from *Mycobacterium marinum* was selected to carry out the reduction of butyric acid to butyraldehyde. Akhtar et al. (2013) showed Car was able to convert a wide range of aliphatic fatty acids into corresponding aldehydes in the presence of ATP and NADPH.

The reaction mechanism involves; adenylation of the bound fatty acid substrate to form an AMP-fatty acyl complex (carboxyphosphate mixed anhydride), formation of a thioester linkage between the fatty acyl moiety and the phosphopantetheine prosthetic group and reduction of the thioester intermediate to the aldehyde (Akhtar et al., 2013). This reaction is illustrated in Figure 5.7.

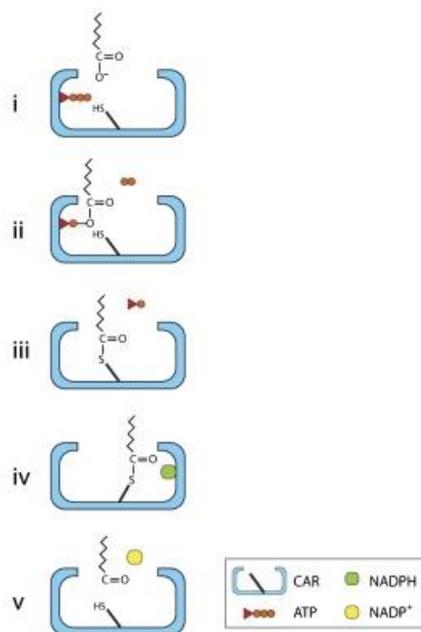


Figure 5.7. The catalytic cycle of Car (Akhtar et al., 2013). i. The fatty acid substrate enters the active site and binds within the vicinity of the ATP domain. ii. An adenosyl moiety is added to the fatty acid, releasing diphosphate. iii. The thiol group residing on the phosphopantetheine arm (angled solid line) reacts with the electrophilic carbonyl group, resulting in formation of a thioester bond. AMP is released in the process. iv. The phosphopantetheine arm

repositions the thioester intermediate within the NADPH domain. v. The thioester intermediate is reduced, via hydride transfer from NADPH, to form the aldehyde product. The product along with NADP⁺ is released and the catalytic cycle is repeated.

To facilitate the conversion of fatty acid to aldehyde, the fatty acid substrate remains tethered to Car by a thioester linkage. This mechanism provides a spatial and steric advantage over alternative reactions dependent on sequential release of intermediates from the enzyme active site (Akhtar et al., 2013).

Phosphopantetheinyl transferase maturation factor

As shown in Figure 5.7 a catalytic feature of Car enzymes is a prosthetic group 4'-phosphopantetheine. The prosthetic group is covalently linked to the Car by a phosphodiester bond to a serine residue. The formation and insertion of this prosthetic group is mediated by a phosphopantetheinyl transferase. Venkitasubramanian et al. (2007) describe Car as an apoenzyme converted to a holoenzyme by post-translational modification via phosphopantetheinylation. They found post-translational phosphopantetheinylation of Car is required for maximum enzyme activity. Co-expression of Car and a phosphopantetheine transferase in *E. coli* gave a reductase with nearly 20-fold higher specific activity than Car alone (Venkitasubramanian et al., 2007).

The phosphopantetheinyl transferase maturation factor (Sfp) from *Bacillus subtilis* was selected for co-expression with the Car, in this work. This combination of enzymes has previously been demonstrated to enable *n*-butanol production in *E. coli* (Pasztor et al., 2014).

Aldehyde reductase (Ahr)

For reduction of fatty aldehyde to fatty alcohol, the Ahr enzyme from *E. coli* was selected for use in this work. Akhtar et al. (2013) demonstrate that this

enzyme catalyzes the reduction of aliphatic aldehydes to alcohols with a substrate profile ranging from C4 to C16, with a selective preference for NADPH. Further Pasztor et al. (2014) show Ahr has specificity for butyraldehyde. Over-expression of *ahr* resulted in an increase in butanol production in aerobic conditions in *E. coli* (Pasztor et al., 2014).

An N-terminal His-tag is added to the final gene in the operon; *ahr*. Akhtar et al. (2013) showed a His-tagged variant of the *ahr* gene resulted in higher protein expression levels compared to the untagged variant. This result suggests inclusion of the His sequence may impart stability at the post transcriptional and/or post translational level. As this operon will be expressed in a thermophilic host, stability of the genes will be essential for a functional pathway. The *ahr* gene is the only gene in the operon to be modified with addition of the His-tag. This is in order to prevent regions of homology in the plasmid gene sequence.

The operon was synthesized by GeneArt (Thermo Fisher Scientific). The sequences were codon optimized for *Geobacillus*. According to the codon optimization table, appendix A8. The gene sequences can be found in appendix A9. Expression of the operon is driven by P_{ldh}. Four unique RBS sequences originating from *Clostridium sporogenes*, were selected for protein translation. The RBS screening work was previously discussed in chapter 3 section 3.4.

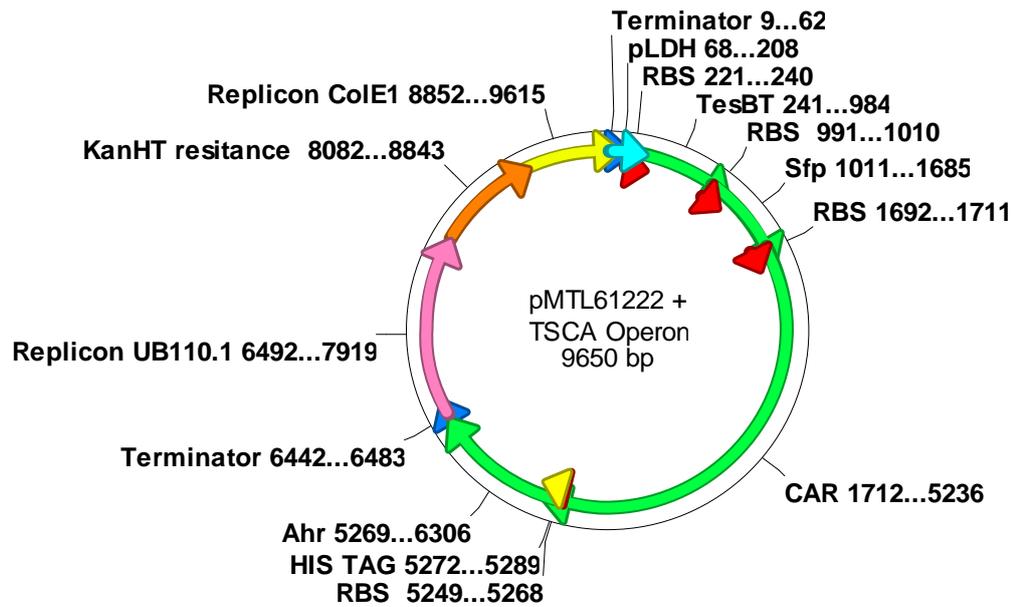


Figure 5.8. A synthetic TSCA operon in the pMTL *Geobacillus* shuttle vector. A four gene operon was synthesised and inserted into the pMTL61222 *Geobacillus* shuttle vector. The genes; *tesBT*, *sfp*, *car* and *ahr* (TSCA) encode the full CoA independent *n*-butanol pathway.

5.4.1 Expression in *Geobacillus*

Following transformation of the pMTL61222+TSCA vector into a *G. thermoglucosidasius* host, Western blot was performed to assay protein expression.

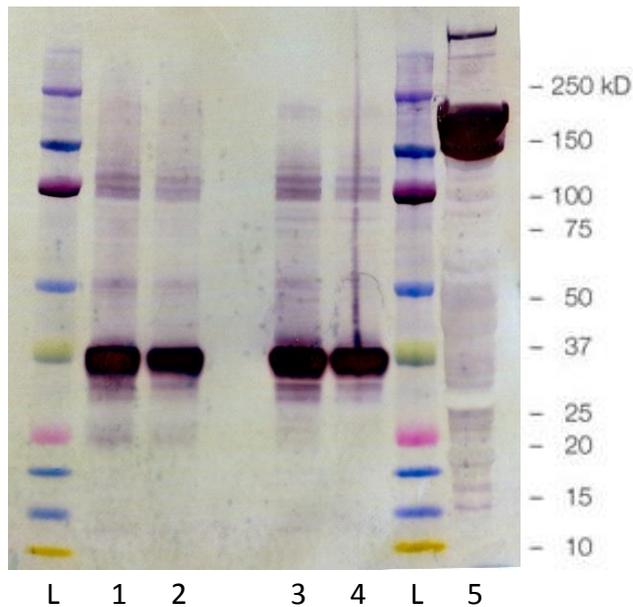


Figure 5.9. Western blot of *Geobacillus* expressing the TSCA operon. *G. thermoglucosidasius* harbouring pMTL61222+TSCA were prepared for Western blot as described previously. Briefly; cultures were adjusted to concentrations of OD 0.4 and 0.8, as indicated and centrifuged at 16,000 x g for 5 minutes. The cell pellet was re-suspended in SDS sample buffer and boiled. The samples were centrifuged and the supernatant recovered providing the soluble cell lysate fraction. The samples were separated on 4-12% Bis-Tris gel (Invitrogen). Following electrophoresis, the proteins were transferred to a PVDF membrane. The membrane was blocked before being probed with antibody. 3 μ l Mouse-anti-Penta-His (Qiagen) and anti-mouse-HRP (Promega). The membrane was washed with TBS with TWEEN[®]20 before 3,3',5,5'-Tetramethylbenzidine (TMB) (Sigma Aldrich) was applied to the membrane for antibody detection. Lanes L; Precision Plus Kaleidoscope protein ladder (Bio-Rad), 1 and 2; supernatant at 0.4 OD concentration, 3 and 4; supernatant at 0.8 OD concentration, 5; positive control, a known functional His-tagged protein.

The protein visualisation by Western blot, shown in Figure 5.9, confirms plasmid based expression of the TSCA operon in *Geobacillus*. Strong bands of 38 kDa size, present in lanes 1 to 4, show expression of *ahr*, the final gene in

the TSCA operon. This demonstrates the *ldh* promoter is sufficient to drive expression of the whole operon.

5.4.2 Culture for *n*-butanol production

In order to determine if the addition of the plasmid based TSCA operon resulted in *n*-butanol production, the *G. thermoglucosidasius* strain harbouring pMTL61222+TSCA was cultured for product analysis.

G. thermoglucosidasius pMTL61222+TSCA was grown in UYSE medium with 20 g/l glucose. Cultures were incubated aerobically at 55°C shaking at 250 rpm for the first 6 hours. At 6 h the cultures either remained in an aerobic environment or were switched to micro-aerobic culture for the remaining 18 hours incubation. Samples were collected for growth, pH and GC analysis at 6, 12 and 24 hour intervals.

CoA INDEPENDENT n-BUTANOL PRODUCTION

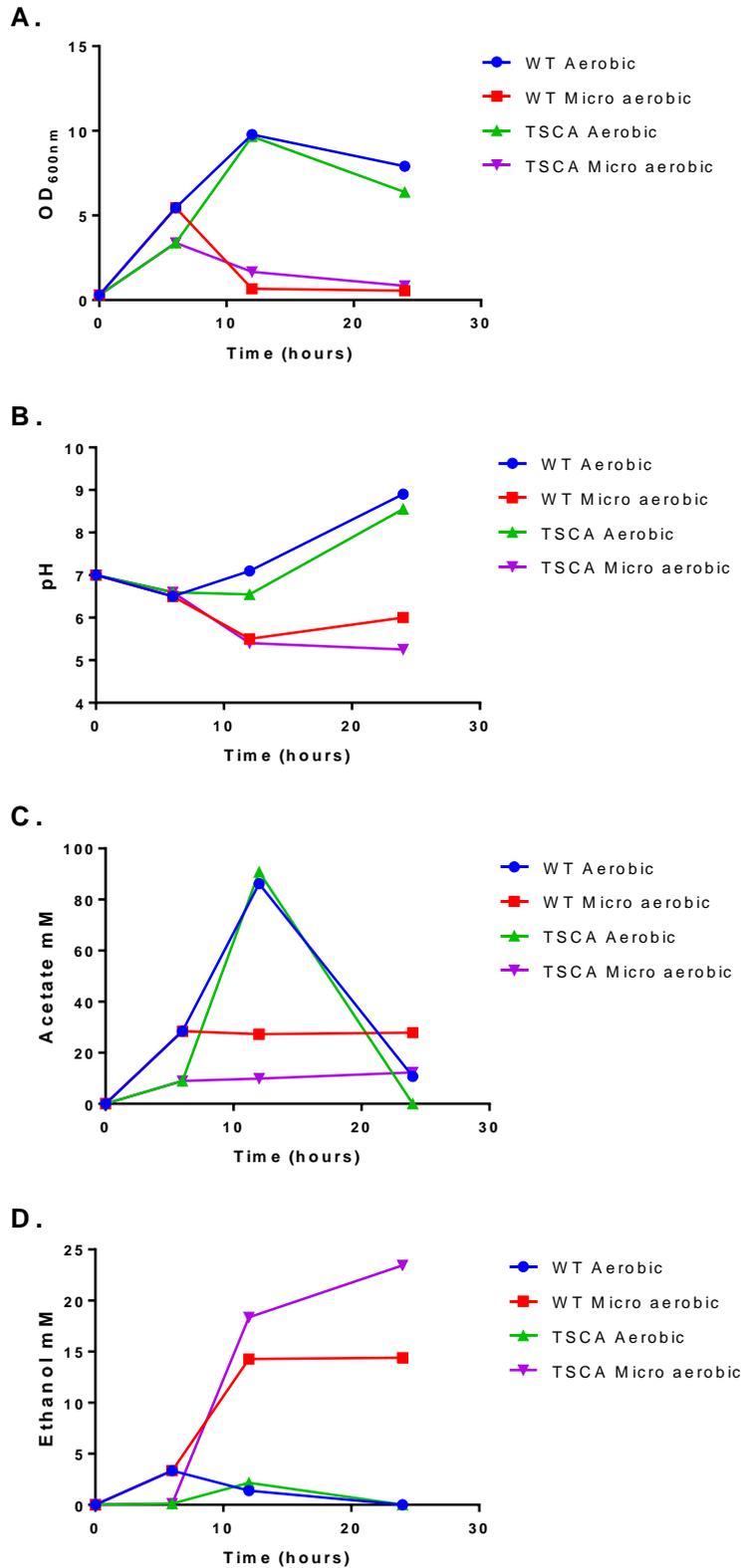


Figure 5.10. Growth, pH and product formation of *G. thermoglucosidasius* expressing the TSCA operon. Graphs show the result of OD, pH and GC solvent

analysis in *G. thermoglucosidasius* wild type and *G. thermoglucosidasius* pMTL61222+TSCA cultures. 50 ml cultures, from a starting OD of 0.3 were grown in UYSE medium with 20 g/l glucose. Cultures were incubated at 55°C for 24 hours. Cultures were grown aerobically in baffled flasks or grown aerobically in baffled flask for 6 h before being switched to micro-aerobic conditions in a 50 ml falcon tube for the remaining incubation. Solvent analysis was carried out by GC, as described previously. All data points represent the average of triplicate data. A. OD₆₀₀, B. pH, C. acetate concentration, D. ethanol concentration.

No *n*-butanol was detected in the cultures. Figure 5.10 shows addition of the TSCA operon did not have any observable effect, on the parameters shown here, compared to WT. Figure 5.10 shows differences in the OD and product profiles when cultured aerobically vs micro-aerobically. In the cultures switched to a micro-aerobic environment, shown as red and purple lines in Figure 5.10 A., at 6 hours the OD drops. When cultured in an aerobic environment, shown as blue and green lines in Figure 5.10 A., the OD increased beyond 6 hours. This is due to *G. thermoglucosidasius*' oxygen requirement for growth. The aerobic cultures produced acetate to a high concentration, reaching a maximum peak at 12 hours. The acetate was then assimilated by the growing culture by 24 h (Figure 5.10 C. blue and green lines). The micro-aerobic cultures produced a lower concentration of acetate, however it continued to increase throughout the 25h period (Figure 5.10 C. red and purple lines). This indicates the glucose was not fully consumed in the micro-aerobic cultures due to the drop in cell growth. Ethanol was produced in cultures grown micro-aerobically (Figure 5.10 D. red and purple lines). Low levels of ethanol are produced from aerobic culture (Figure 5.10 D. blue and green lines) as ethanol is generated via a fermentative pathway.

5.4.3 Fatty acid profiling

Following no detection of *n*-butanol from cultures of *G. thermoglucosidasius* pMTL61222+TSCA, analysis of free fatty acids was carried out to determine if the addition of the pathway altered the fatty acid composition of the strain.

Fatty acid ethyl ester extraction of wild type *G. thermoglucosidasius* and *G. thermoglucosidasius* pMTL61222+TSCA strains was carried out, as previously described. Briefly; 0.4 mL of culture supernatant was supplemented with 50 μ L of 10% (w/v) NaCl and 4 μ L of internal standard. 50 μ L of glacial acetic acid and 200 μ L of ethyl acetate were added. The mixture was centrifuged at 16,000 x g for 10 min and esterified with ethanol by addition of 900 μ L of a 30:1 mixture of ethanol and 37% (v/v) HCl to 100 μ L of the organic phase, and incubated at 55°C for 1 hr. The fatty acyl ethyl esters were extracted by adding 500 μ L dH₂O and 500 μ L hexane and vortexing for 20 seconds. The top hexane layer extract (300 μ L) was analyzed by GC-MS. The extraction and analysis was carried out in triplicate. The triplicate samples showed the same profile, with one selected for presentation here.

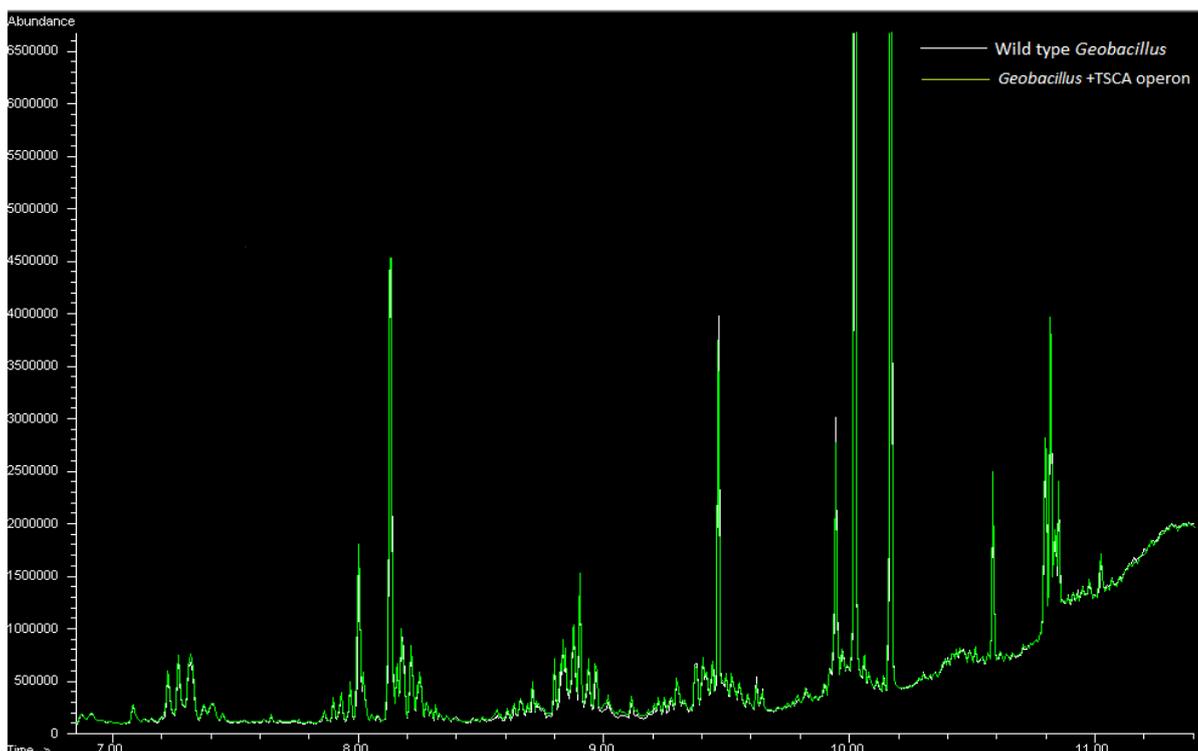


Figure 5.11. Free fatty acid profiles of *G. thermoglucosidasius* wild type and *G. thermoglucosidasius* pMTL61222+TSCA cells. Fatty acids were analyzed by GC-MS in full scan mode using Agilent 7890 GC with an Agilent 7000 Triple Quadrupole GC-MS system. The column used was Zebron ZB-5MS (length: 30 m; diameter: 0.25 mm; film thickness: 0.25 μ m). The injection volume was 1 μ L without split mode and oven temperature ramped from 40 to 325°C (initial ramp at the rate of 5°C per min from 40 to 75°C and then 10°C per min until 325°C). The total run time was 41 min. Wild type profile is shown as the white trace. The TSCA strain is shown as the green trace.

Figure 5.11 shows the full fatty acid profile, with no major differences observable in the comparison of the two strains tested.

5.5 Discussion

The CoA dependent *n*-butanol pathway is not suited to application in an aerobic host strain. The work carried out in this chapter explored a possible alternative, oxygen tolerant route to *n*-butanol. With the desire to identify a pathway which

would not be limited by an aerobic environment and which may be better suited to the *Geobacillus* host. A synthetic oxygen tolerant *n*-butanol pathway utilizing the native fatty acid biosynthesis pathway was constructed and expressed in *G. thermoglucosidasius*. This pathway has previously been tested in a mesophilic *E. coli* host. Pasztor et al. (2014) demonstrated bacterial *n*-butanol production under aerobic conditions. This indicates the feasibility of potential application of this pathway in other host organisms. Although there are many hurdles to overcome when working with a relatively novel organism, and especially a thermophile. In comparison to *E. coli* very little is known of the *Geobacillus* metabolism or regulation. Making application of systems, such as introduction of this pathway, complex.

5.5.1 Thioesterase

The initial step in the CoA independent pathway is the interruption of the host's native fatty acid synthesis, specifically at C4 chain length, producing butyric acid. This step requires a C4 specific thioesterase. The enzyme that determines fatty acid chain length is acylacyl carrier protein thioesterase. This enzyme catalyzes the terminal reaction of fatty acid biosynthesis, acyl-ACP thioester bond hydrolysis to release a free fatty acid and ACP. As *Geobacillus* does not natively produce butyric acid, this required introduction of a heterologous enzyme. A review of screening carried out by Jing et al. (2011) identified a potential candidate gene from *Bacteroides thetaiotaomicron*. Although the total fatty acid concentration produced was not the highest in the study, 27% of the total fatty acid was C4. This suggested selection of this enzyme to be the best candidate for butyric acid production. However Western blot showed much of the protein accumulated as inclusion bodies. This result suggested that the enzyme was not compatible with *Geobacillus*. Factors which may contribute to aggregate formation include; lack of post-translational modification, lack of access to chaperones and enzymes catalyzing folding such as cis-trans isomerase and high concentrations of protein coupled with limited solubility of folding intermediates. Any number of these factors could be

causing compatibility problems. It is generally considered that protein aggregation in bacteria implies a loss of biological activity. However Sánchez de Groota & Ventura (2006) suggested this may not necessarily be the case. They found proteins were active in inclusion bodies. In this work an enzyme assay showed activity of 63.5 nmole/min/mg. This activity could be attributed to the smaller fraction of the protein produced which was soluble. Or it could be from activity of the protein in inclusion bodies. It is likely the high growth temperature is impacting the protein production and folding. Here an incubation temperature of 52°C was used. An initial assay of lysate from *E. coli* expressing the *tesBT* gene, tested at 50°C, showed activity (data not shown). This indicated the enzyme would be thermostable. However, both *Bacteroides* and *E. coli*, the two species in which functionality of the gene has been shown, are mesophilic. Production of the enzyme in a soluble, properly folded and active conformation may not be occurring at the higher temperature. In addition the gene used in this work was codon optimized for *E. coli*, this may hinder the optimal application in an alternative host.

The lack of quantifiable butyric acid or *n*-butanol in this work could be attributed to insufficient or no specificity of the TesBT thioesterase for the required C4 carbon chain length. Although a butyric acid peak was detected on the GC-MS trace, this was not in sufficient quantity to be detected as a product from culture of the strain. In a study by Jawed et al. (2016) the chain length specificities of the TesBT thioesterase used in this work was measured, shown in Table 5.2. Their results showed 2–3 fold higher specific activities of TesBT against longer chain fatty acid precursors of C6 and C8 chain length as compared to butyric acid precursor C4. This result, along with the GC-MS profile obtained in this work, indicate the addition of TesBT to *Geobacillus* is generating longer chain fatty acids. In most organisms the native fatty acid synthesis process optimally produces 16- and 18-carbon (C16 and C18) fatty acids. As a critical acyl chain termination enzyme, a thioesterase with desired substrate specificity is important for engineering this pathway.

Table 5.2 Chain length specificity of thioesterases (Jawed et al., 2016).

Substrate	Thioesterase	
	TesBF (nmol min ⁻¹ mg ⁻¹)	TesBT (nmol min ⁻¹ mg ⁻¹)
Acetyl CoA (C2)	0.11	4.62
Malonyl CoA (C3)	0.00	0.18
Acetoacetyl CoA (C4)	0.76	1.38
Butenoyl CoA (C4:1)	0.43	0.79
Butyryl CoA (C4)	1.28	2.91
Hexanoyl CoA (C6)	2.32	6.63
Octanoyl CoA (C8)	2.06	5.57

The thioesterase gene tested here originates from a strictly anaerobic bacterium; *Bacteroides thetaiotaomicron*. The gene, therefore, may be incompatible with the aerobic growth conditions used to culture the host strain here. In the screening carried out by (Jing et al., 2011) the only aerobic bacterial thioesterase sources were *Geobacillus* and *Bdellovibrio bacteriovorus*. Both generated very low percentages of C4:0 fatty acid; 2.35% and 0.1%, respectively. Further screening of aerobic bacterial strains could determine if this is a trend across aerobic organisms, or if a short chain specific aerobic native alternative gene can be identified. However as the basis of this pathway is to provide an oxygen tolerant system, as demonstrated by Pasztor et al. (2014), this suggests oxygen tolerance will not be the decisive issue at play here.

The *G. thermoglucosidasius* genome has 3 native thioesterases present; phenylacetic acid degradation-related protein; K02614 acyl-CoA thioesterase [EC:3.1.2.-], 4-hydroxybenzoyl-CoA thioesterase; K07107 acyl-CoA thioester hydrolase [EC:3.1.2.-] and thioesterase superfamily protein; K07107 acyl-CoA thioester hydrolase [EC:3.1.2.-]. Identified by analysis of the *G. thermoglucosidasius* genome using KEGG. The substrate specificities of these thioesterases is unknown. However as the fatty acid profile of the wild type strain, obtained in this work, showed no butyric acid peak it is unlikely the genes will show specificity for C4 chain length substrate.

As an alternative thioesterase, a novel short chain specific thermophilic esterase, Fnod, should be considered. The gene source, *Fervidobacterium nodosum*, is a thermophilic bacterium with optimal growth temperature of 80°C. The enzyme will, therefore, be highly thermostable. Yu et al. (2010) report short chain specificity with catalytic activity for C4 substrate to be 5.8 ($k_{\text{cat}}/K_{\text{m}}$). This is a potential candidate for future investigation.

5.5.2 Reduction of butyric acid to butyraldehyde

The enzymatic reduction of butyric acid to butyraldehyde could be a limiting step. The Car enzyme responsible for this step has a low affinity towards shorter chain substrates (Akhtar et al., 2013). Although the Car enzyme was found to be a versatile enzyme capable of accepting a broad range of aliphatic fatty acids. Kinetic characterization, shown in Figure 5.12, reported greater catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$) toward longer chain substrates (Akhtar et al., 2013). C4 fatty acid showed a low catalytic efficiency due to a higher K_{m} , compared with longer chain length substrates. Therefore a large concentration of substrate is needed for efficient conversion.

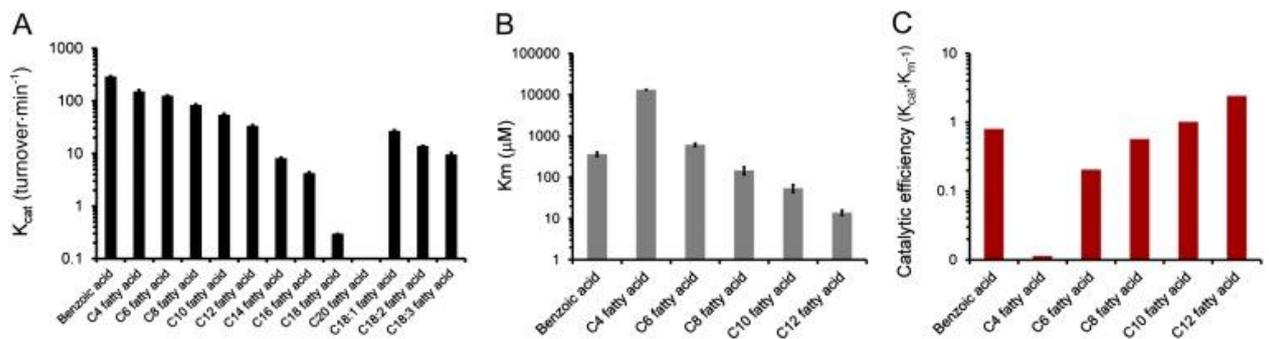


Figure 5.12. The kinetic characterization of Car (Akhtar et al., 2013). A. The k_{cat} values were obtained for fatty acid substrates ranging from C4 to C18. B. K_{m} C. catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$) were determined for C4–C12 fatty acid substrates.

As with Car, Ahr was found to accept a broad range of substrates with specificity for NADPH (Akhtar et al., 2013). The promiscuity of both enzymes means end products of different chain lengths will be formed.

5.5.3 Host strain metabolic engineering

Further modifications to the host strain could be made to improve the production of butyric acid and therefore *n*-butanol. Optimization of host metabolism and stimulation of fatty acid biosynthesis could lead to higher fatty alcohol yields. The production of free fatty acids could be increased by; enhancing biosynthesis, blocking degradation or controlling regulation of both. Deletion of the beta oxidation genes encoding *fadD* and *fadL* would eliminate fatty acid degradation (Lu et al., 2008, Liu et al., 2012). Zhang et al. (2012) enhanced fatty acid production by the expression of the regulatory transcription factor FadR. In addition competing pathways, such as acetic acid formation, could be knocked out or down regulated.

The accumulation of acyl-ACP in the cell has an inhibitory effect of on the fatty acid synthesis pathway. Production of short chain fatty acids could be improved by removing negative feedback. One approach is the addition of a long chain acyl-ACP thioesterase. This means any fats escaping the short chain thioesterase are removed by the second long chain thioesterase, preventing feedback inhibition. The long chain fats can be used to provide the cell with energy via beta oxidation.

5.5.4 Conclusions

Although *n*-butanol production was not demonstrated in this work, the CoA independent pathway remains a promising route for further investigation. Some areas for further consideration have been highlighted. Many possible alterations to the system require investigation if high yield, titre and productivity are to be achieved in *Geobacillus*. The major challenges faced include; identification of compatible enzymes with regard to thermostability

and oxygen tolerance; identification of enzymes with the desired specificity; improvement in enzyme activity across the entire metabolic pathway; recycling or replacing cofactors for enzymatic reactions; enriching precursors and eliminating byproducts; optimizing and balancing the fluxes of metabolic networks to reduce burden on the host and removing negative feedback regulation.

CoA INDEPENDENT n-BUTANOL PRODUCTION

CHAPTER 6

HOST STRAIN CHARACTERISATION

6.1 Introduction

As a relatively novel host organism, little is known about the properties of *G. thermoglucosidasius*. Desirable traits for an industrial process organism include; reliable and consistent growth and metabolite formation, resistance to phage contamination, asporogenous, no biofilm formation and high product tolerance. Currently butanol fermentation is hampered by phage contamination, solvent stress and toxicity (Nanda et al., 2017). Most industrial producers use *S. cerevisiae* as the organism of choice because of its robustness under industrial conditions. Here *G. thermoglucosidasius* will be evaluated to determine its suitability as an industrial process organism.

6.2 Prophage

Bacteriophage (phage) are virus' which infect and replicate within bacteria. Clostridia are susceptible to phage infection. Butanol fermentations are commonly hindered by phage infection. Phage infection results in host cell death and therefore reduced productivity. The first reported phage infection of a butanol fermentation occurred in the USA in 1923. In this case, solvent yields from a *C. acetobutylicum* fermentation halved (Jones et al., 2000).

In order to determine if *G. thermoglucosidasius* culture is vulnerable to phage contamination, here the presence of prophages was examined. Prophages are phage genomes integrated into the host genome. If susceptible to phage infection, dormant phage may be present as prophage in the *G. thermoglucosidasius* genome. Prophages are a source of new genes and functions in bacterial genomes. Introduction of prophages into the genome cause gene disruption which could positively or negatively influence the characteristics of a production strain. Any prophage present in the genome could be activated into a lytic form under stressful conditions, such as an industrial fermentation. This would result in cell death and prevention of product formation. Prophage presence and susceptibility of the host strain to phage infection is therefore undesirable.

Various attempts to induce prophage yielded no plaques. UV exposure, high temperatures, aged culture and exposure to toxic levels of *n*-butanol were used to stress *G. thermoglucosidasius* cells. These strategies were designed to cause cell damage and thus induce the prophage to excise from the host chromosome into a lytic cycle. These experimental results suggest no prophage are present in the *G. thermoglucosidasius* genome.

Subsequent analysis of the *G. thermoglucosidasius* genome using PHASTER (PHAge Search Tool Enhanced Release, <http://phaster.ca/>) confirmed no complete phage sequences are present in either the chromosome or the two megaplasמידs. Six incomplete phage regions were identified on the chromosome. No partial phage sequences were identified in either of the megaplasמידs.

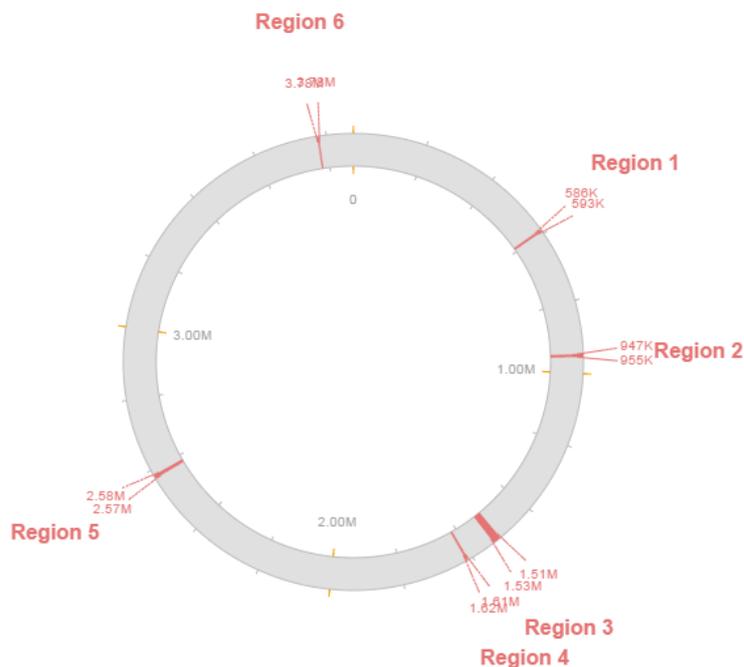


Figure 6.1. Locations of six incomplete phage regions in the *G. thermoglucosidasius* genome.

Phage infection is a major problem in ABE fermentation. However, to date, phage infection in any *Geobacillus* fermentation or culture has not been reported. This work indicates *G. thermoglucosidasius* does not contain any intact prophages. This is a positive feature of the strain for industrial application. Any bacterial strain containing complete prophage sequences has the potential to induce lytic phage infection when exposed to stressful fermentation conditions. Phage infection results in reduced product formation and would halt fermentation. The lack of complete prophage sequences and presence of partial sequences within the *G. thermoglucosidasius* genome could be attributed to a high number of CRISPR elements within its genome (Hussein et al., 2015). CRISPR elements play a role in bacteria's defence system. Possession of a 'strong' immune system is beneficial and a desirable trait for an industrial production strain.

6.3 Product toxicity

Biological butanol production is currently limited by the sensitivity of solvent producing microbes towards butanol. In order to develop a process organism suitable for high yield fermentation, a high tolerance to the product of interest is required. In this case it would be impossible to accumulate a high concentration of *n*-butanol if the butanol producing bacteria were sensitive to butanol. Exposure to toxic levels of *n*-butanol will induce a stress response, cause cell damage or death. Such effects would inhibit or prevent product formation.

In addition to *n*-butanol tolerance the strain should ideally also be tolerant of a range of other inhibitors. Compounds known to be inhibitory to bacteria make up rice straw derived lignocellulosic hydrolysate, the target fermentation substrate for this work. These compounds include; acetic acid, sulphate, furfural, hydroxymethylfurfural and water soluble lignin. If the bacterial production strain were tolerant to these compounds it could result in cheaper processing as less pre-treatment would be required.

6.3.1 Toxicity determination

In order to characterise the strain and to understand the potential maximum yield obtainable, the tolerance of *G. thermoglucosidasius* to *n*-butanol was determined.

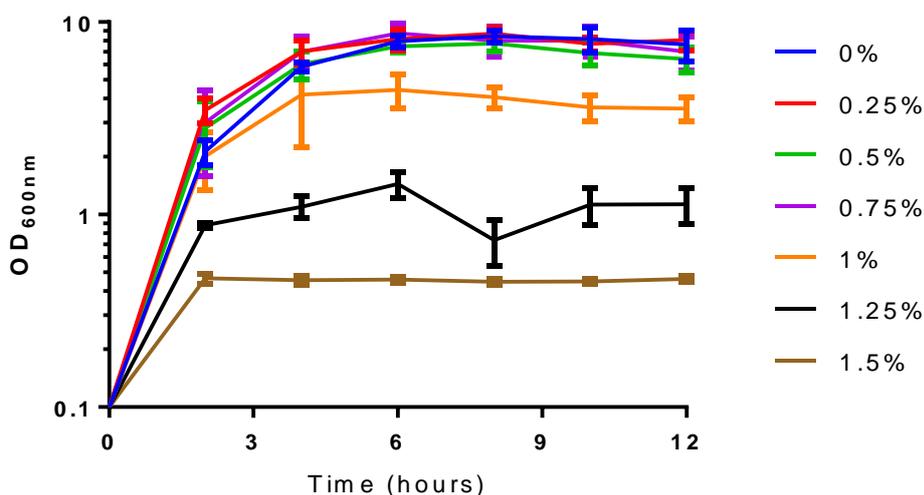


Figure 6.2. Growth of *G. thermoglucosidasius* when exposed to a range of *n*-butanol concentrations. Wild type *G. thermoglucosidasius* cells were cultured from a starting OD₆₀₀ of 0.1 at 52°C with shaking at 250 rpm in aerobic conditions. The cells were cultured in 2SPYNG medium supplemented with *n*-butanol to the percentages (v/v) as indicated. Triplicate data with average and standard deviation are shown here.

The minimum inhibitory concentration (MIC) of *n*-butanol was determined to be 1% (v/v), in accordance with observation of impaired growth. This level of product tolerance is too low for the strain to be considered an advantageous industrial production organism. In general, *n*-butanol inhibits the growth of *Clostridium* sp. at 2.0% (v/v) (Liu and Qureshi, 2009). Many attempts to identify *n*-butanol tolerant organisms have previously been made. Microbes capable of growth in the presence of >2.0% (v/v) *n*-butanol have been identified. Such microbes include; *Pseudomonas*, *Zymomonas*, *Enterococcus*, *Lactobacillus*, *Bacillus*, *Eubacterium* and *Saccharomyces* (Kanno et al., 2013; Kataoka et al.,

2011; Knoshaug and Zhang, 2009; Li et al., 2010; Rühl et al., 2009; Ting et al., 2012). Ting et al. (2012) isolated an *Enterococcus faecium* strain tolerant of *n*-butanol up to 3% (w/v). This strain was capable of growth both aerobic and anaerobically however it did not produce *n*-butanol aerobically. Adapted *Pseudomonas putida* strains have been reported to grow in the presence of up to 6% (v/v) *n*-butanol (Rühl et al., 2009). This is the highest reported *n*-butanol concentration tolerated by a microbe, to date. Although this strain did consume *n*-butanol. No isolated or adapted strains with high *n*-butanol tolerance have demonstrated aerobic *n*-butanol production.

6.3.2 Directed evolution for increased *n*-butanol tolerance

For *G. thermoglucosidasius* to be considered an industrial organism of choice, its tolerance toward *n*-butanol requires improvement. High *n*-butanol resistance is a crucial characteristic for butanol-production strains. In addition a reduction in sensitivity to *n*-butanol should allow for higher production yields.

Due to limited knowledge of the complex biological systems of *G. thermoglucosidasius*, rational design does not offer a viable strategy for alteration of the phenotype for increased resistance to *n*-butanol. Directed evolution offers an alternative approach, which does not require detailed functional, structural or mechanistic information of the biological system. Directed evolution can be utilised as a powerful tool to bridge the gap between current knowledge and the desired destination (Cobb et al., 2013).

In this work directed evolution was used to increase the *n*-butanol tolerance of *G. thermoglucosidasius*. Exposure to sub-inhibitory concentrations was used to select for bacterial resistance as a tool for directed evolution. Strain adaptation as a result of changing growth environment, places evolutionary pressure on metabolically active cells, resulting in heritable changes in the host chromosomal DNA (Jilani et al., 2017).

Here, a directed evolution approach was employed with the aim to develop a *G. thermoglucosidasius* strain with increased *n*-butanol tolerance. Directed evolution is used to direct cell adaptation for growth in the presence of increasing *n*-butanol concentrations. Wild type *G. thermoglucosidasius* cells were used to inoculate 5 ml 2SPYNG medium. The cultures were incubated at 52°C with 250 rpm shaking. Cultures were serially passaged after overnight growth by inoculation of 5 ml fresh 2SPYNG medium with 50 µl of previous overnight growth. In some cases slower growth resulted in passaging after 2 days of incubation. To the second inoculation 0.8% (v/v) *n*-butanol was added to the medium. The concentration of *n*-butanol was increased by 0.1% with each transfer.

After 17 passages the *G. thermoglucosidasius* cells were able to grow in the presence of 2.5% (v/v) *n*-butanol. This result demonstrated a significant increase in *n*-butanol tolerance. With concentrations greater than 2.5% (v/v) *n*-butanol, no growth was observed.

6.3.3 Sequencing of the evolved strain

Following observation of increased tolerance towards *n*-butanol, the adapted *G. thermoglucosidasius* strain resulting from directed evolution was subject to whole genome sequencing. The phenotypic change observed could be due to gene mutations or changes to gene expression levels (Cobb et al., 2013). In attempt to identify changes in the genome sequence which may be responsible for the increased tolerance, the strain was sequenced. Whole genome sequencing was carried out by MicrobesNG (<http://www.microbesng.uk>), at the University of Birmingham. High throughput Illumina MiSeq using 2x250 bp paired-end reads. Paired reads of the adapted strain were mapped against the *G. thermoglucosidasius* NCIMB 11955 reference genome (Sheng et al., 2016).

99.75% of the genome was mapped. 94 variants were identified in comparison to the reference genome. 93 of which were located on the chromosome, with

HOST STRAIN CHARACTERISATION

only one variant identified on the smaller of the megaplasms, pNCI002. Of these 94 variants 31 resulted in amino acid changes. 37 variants were identified in inter gene areas. These could include promoter regions or control elements.

81 single nucleotide variants (SNV) resulted in 25 amino acid changes

7 multi-nucleotide variants (MNV) resulted in 3 amino acid changes

3 insertions resulted in 2 amino acid changes

2 deletions resulted in 1 amino acid change

1 replacement resulted in 0 amino acid changes.

The SNV, MNV, insertions and deletions resulting in amino acid changes are detailed in Tables 6.1 and 6.2. The silent variants are not detailed here.

Table 6.1. Single nucleotide variations in the *n*-butanol resistant strain in comparison to the *G. thermoglucosidasius* NCIMB 11955 reference strain.

Type	Length	Reference	Allele	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change	Product
SNV	1	C	T	28	100	Gene: BCV53_00905	ANZ32034.1:c.[1-169303G>A]; ANZ28799.1:c.[353G>A]	ANZ28799.1:p.Arg118His	HPr kinase/phosphotransferase
SNV	1	G	A	28	100	Gene: BCV53_01210	ANZ28854.1:c.[628C>T]; ANZ32034.1:c.[1-229318C>T]	ANZ28854.1:p.Arg210Trp	DNA-binding response regulator
SNV	1	A	G	70	100	Gene: BCV53_01465	ANZ32034.1:c.[1-296647T>C]; ANZ28896.1:c.[434T>C]	ANZ28896.1:p.Met145Thr	Transposase
SNV	1	G	A	19	100	Gene: BCV53_02900	ANZ29139.1:c.[334C>T]; ANZ32034.1:c.[1-575804C>T]	ANZ29139.1:p.Arg112Trp	DNA polymerase III subunit delta
SNV	1	G	A	26	100	Gene: BCV53_03900	BCV53_03900:c.[192G>A]; ANZ32034.1:c.[1-778695C>T]	BCV53_03900:p.Trp64*	Transposase
SNV	1	T	A	23	100	Gene: BCV53_07415	ANZ29921.1:c.[194T>A]; ANZ32034.1:c.[1-1497643A>T]	ANZ29921.1:p.Leu65His	Cardiolipin synthase
SNV	1	C	T	30	100	Gene: BCV53_08230	ANZ32034.1:c.[1-1655933G>A]; ANZ30072.1:c.[32C>T]	ANZ30072.1:p.Ala11Val	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
SNV	1	C	T	24	100	Gene: BCV53_09190	ANZ30251.1:c.[379C>T]; ANZ32034.1:c.[1-1839032G>A]	ANZ30251.1:p.Gln127*	Flagellar biosynthesis protein FlhF
SNV	1	T	G	22	100	Gene: BCV53_11320	ANZ32034.1:c.[693+1575626A>C]; ANZ30630.1:c.[881A>C]	ANZ30630.1:p.Tyr294Ser	Betaine-aldehyde dehydrogenase
SNV	1	C	T	21	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157164G>A]; ANZ30980.1:c.[325G>A]	ANZ30980.1:p.Glu109Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	T	19	100	Gene: BCV53_13285	ANZ30980.1:c.[287G>A]; ANZ32034.1:c.[693+1157126G>A]	ANZ30980.1:p.Trp96*	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	T	20	100	Gene: BCV53_13285	ANZ30980.1:c.[262G>A]; ANZ32034.1:c.[693+1157101G>A]	ANZ30980.1:p.Glu88Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	T	19	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157093G>A]; ANZ30980.1:c.[254G>A]	ANZ30980.1:p.Arg85Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	A	C	22	100	Gene: BCV53_13285	ANZ30980.1:c.[216T>G]; ANZ32034.1:c.[693+1157055T>G]	ANZ30980.1:p.Asn72Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	T	G	18	100	Gene: BCV53_13285	ANZ30980.1:c.[207A>C]; ANZ32034.1:c.[693+1157046A>C]	ANZ30980.1:p.Arg69Ser	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	G	15	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157034G>C]; ANZ30980.1:c.[195G>C]	ANZ30980.1:p.Gln65His	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	T	A	18	100	Gene: BCV53_13285	ANZ30980.1:c.[186A>T]; ANZ32034.1:c.[693+1157025A>T]	ANZ30980.1:p.Leu62Phe	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	T	C	17	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157017A>G]; ANZ30980.1:c.[178A>G]	ANZ30980.1:p.Asn60Asp	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	G	19	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157008G>C]; ANZ30980.1:c.[169G>C]	ANZ30980.1:p.Glu57Gln	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	T	22	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1156933G>A]; ANZ30980.1:c.[94G>A]	ANZ30980.1:p.Glu32Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	T	G	24	100	Gene: BCV53_13285	ANZ30980.1:c.[92A>C]; ANZ32034.1:c.[693+1156931A>C]	ANZ30980.1:p.Gln31Pro	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	A	C	23	100	Gene: BCV53_13285	ANZ30980.1:c.[90T>G]; ANZ32034.1:c.[693+1156929T>G]	ANZ30980.1:p.Asn30Lys	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	T	G	24	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1156927A>C]; ANZ30980.1:c.[88A>C]	ANZ30980.1:p.Asn30His	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
SNV	1	C	T	21	100	Gene: BCV53_14485	ANZ32034.1:c.[693+912306G>A]; ANZ31193.1:c.[712G>A]	ANZ31193.1:p.Val238Ile	Catechol 2,3-dioxygenase
SNV	1	T	A	16	100	Gene: BCV53_17405	ANZ32034.1:c.[693+353082A>T]; ANZ32210.1:c.[870A>T]	ANZ32210.1:p.Leu290Phe	Protein translocase subunit SecDF

Table 6.2. Multi nucleotide variants, insertions and deletions in the *n*-butanol resistant strain in comparison to the *G. thermoglucosidarius* NCIMB 11955 reference strain.

Type	Length	Reference	Allele	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change	Product
MNV	2	CA	TG	18	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157082_693+1157083delTGinsCA]; ANZ30980.1:c.[243_244delTGinsCA]	ANZ30980.1:p.Gly82Ser	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
MNV	2	TA	CC	17	100	Gene: BCV53_13285	ANZ32034.1:c.[693+1157069_693+1157070delTAinsGG]; ANZ30980.1:c.[230_231delTAinsGG]	ANZ30980.1:p.Leu77Trp	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
MNV	2	CT	TG	25	100	Gene: BCV53_13285	ANZ30980.1:c.[70_71delAGinsCA]; ANZ32034.1:c.[693+1156909_693+1156910delAGinsCA]	ANZ30980.1:p.Arg24Gln	Hypothetical protein (pyridoxal phosphate dependent aminotransferase)
Type	Length	Reference	Allele	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change	Product
Deletion	4	CAGC	-	33	93.939394	Gene: BCV53_03260	ANZ29196.1:c.[262_265delGCTG]; ANZ32034.1:c.[1-650759_1-650756delGCTG]	ANZ29196.1:p.Ala88fs	Peptide transporter
Insertion	1	-	A	26	92.307692	CDS: BCV53_07460,	ANZ32034.1:c.[1-1506121_1-1506120insT]; ANZ29930.1:c.[328_329insA]	ANZ29930.1:p.Glu113fs	GTP pyrophosphokinase
Insertion	2	-	AA	71	97.183099	Gene: BCV53_19795	BCV53_19795:c.[485_486insAA]; BCV53_19560:c.[1-5946_1-5945insAA]	BCV53_19795:p.Gly164fs	Transposase

Tables 6.1 and 6.2 detail the SNV, MNV, insertions and deletions resulting in amino acid changes in the genome sequences of wild type and *n*-butanol adapted *G. thermoglucosidasius* strains. The coding region and annotated protein affected are shown. Coverage details the number of reads obtained. Frequency details the percentage of reads which gave the result shown.

17 variants, highlighted in green, show amino acid changes in one gene. The gene affected was annotated as a hypothetical protein with BLAST analysis identifying the protein as a pyridoxal phosphate dependent aminotransferase. Highlighted in orange is the SNV resulting in an amino acid change in a catechol 2,3-dioxygenase gene. Highlighted in yellow is the one megaplasmid variant. An insertion of AA resulted in frame shift of a transposase gene on megaplasmid pNCI002.

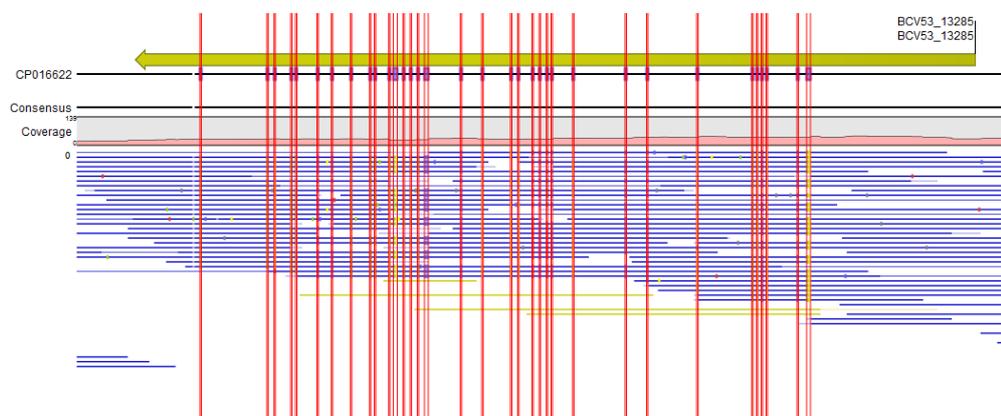


Figure 6.3. Variants identified in gene BCV53_13285. Genome sequencing data identified a total of 34 changes in a single gene, BCV53_13285. 17 of which resulted in amino acid changes, these changes are highlighted in green in Tables 6.1 and 6.2. The gene was annotated to encode a hypothetical protein. A BLAST search identified the protein to be a pyridoxal phosphate dependent aminotransferase.

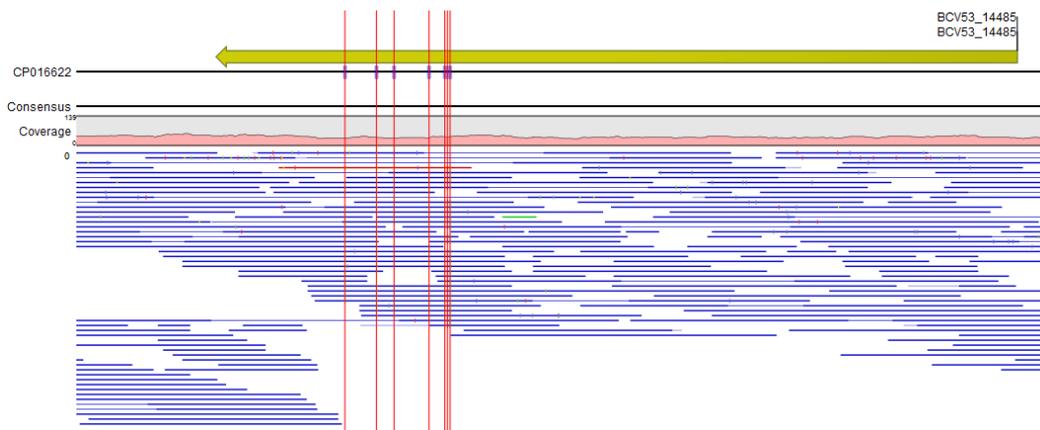


Figure 6.4. Variants identified in gene BCV53_14485. Genome sequencing data identified 7 changes in gene BCV53_14485, annotated to encode a catechol 2,3-dioxygenase. Only one of these variants resulted in an amino acid change. The SNV, highlighted in orange in Table 6.1, caused a change of base C to T which resulted in valine to isoleucine amino acid change.

It is known that the catalytic properties of catabolic enzymes can be dramatically influenced by single amino acid changes (Junca et al., 2004). Any of the proteins listed in Tables 6.1 and 6.2 could therefore, be responsible for the change in phenotype observed. Alternatively a combination of any number of the genes could play a role. It is also possible the changes observed are the result of up or down regulation of other genes, in response to *n*-butanol toxicity. Alper et al. (2005) showed that promoters can be engineered via directed evolution to achieve precise strengths and regulation to tune genetic control. Such changes may be the result of any of the 37 inter gene variants detected.

A particular gene of interest encodes a pyridoxal phosphate dependent aminotransferase (PLP). This particular gene incurred a significant number of variants. Pyridoxal phosphate is the active form of vitamin B₆ and is a versatile catalyst, acting as a coenzyme in a multitude of reactions, including decarboxylation, deamination and transamination. Pyridoxal phosphate dependent enzymes are primarily involved in the biosynthesis of amino acids

and amino acid-derived metabolites. PLP also plays a role in the biosynthetic pathways of amino sugars. By their nature such enzymes are highly catalytically versatile and have widespread involvement in cellular metabolism. PLP acts as cofactor in more than 160 different enzymes classified by the Enzyme Commission, representing 4% of all known cellular catalytic activities (Di Salvo et al., 2013). There are 12 aminotransferases annotated on the *G. thermoglucosidasius* NCIMB 11955 chromosome and 16 annotated on the *G. thermoglucosidasius* C56-YS93 genome. In addition to possible further unannotated genes, such as the one identified here. PLP-dependent enzymes serve vital roles in all living organisms and catalyze a number of diverse chemical reactions, such as transamination, decarboxylation, racemization, carbon-carbon bond cleavage and formation. PLP-dependent activities are involved in essential biosynthetic pathways including glucose and lipid metabolism, amino acid metabolism, heme and nucleotide synthesis, and neurotransmitter production. As a consequence of their crucial metabolic relevance and the presence of multiple copies on the genome, it is hard to predict if or how the gene mutation identified here would affect the cell phenotype.

A catechol 2,3-dioxygenase gene was a further site of multiple variants. Catechol degrades aromatic compounds catalyzing ring cleavage reactions. As *n*-butanol is a straight chain alcohol, this mutation identified may not be relevant in conferring the observed phenotype change.

Genome sequencing has identified amino acid variants in 3 transposase genes. A transposase is an enzyme which binds to the end of a transposon, catalyzing its movement to another part of the genome. Transposition is an important mechanism used to create genetic diversity within species. It is used to enable adaptation in response to a changing environment. Such as the changing culture conditions cells were exposed to in this work. The transposase gene mutations identified could indicate increased transposon activity in response to the

stressful environment. One of the transposase gene variants identified was on the megaplasmid pNCI002. This could indicate transposon activity on the megaplasmids in addition to the chromosome. The genes encoding transposases are widespread in the genome. Therefore changes to these genes is to be expected in any genome analysis.

To elucidate the variants responsible for the phenotypic change observed here, further investigation is required. Repair of the variants in the adapted strain and gene knock-out in the wild type strain could be used to observe the resultant phenotypes. In addition RNA analysis should be employed to determine up or down regulation of genes which could cause phenotypic changes. Such transcriptional changes would not be detected by genome sequencing studies.

6.4 Sporulation

Endospore formation is a process in which some bacterial species are able to respond to unfavourable growth conditions by differentiating into dormant spores. Endospores are resistant to extreme conditions such as pressure, extreme heat or cold, drought, starvation, biocides and UV irradiation (Moeller et al., 2008). Spores are able to reactivate when favourable conditions return, ensuring survival. *Geobacillus* readily sporulate. This characteristic has allowed wide spread distribution of the bacteria across the globe and explains isolation of the bacteria from environments not conducive to *Geobacillus* growth or survival.

6.4.1 Generation of a sporulation mutant

Sporulation deficiency is a desired strain characteristic in industrial settings. Industrial fermentation processes expose cells to stressful conditions. Such processing would trigger stress response and sporulation. When dormant, cells are no longer metabolically active, therefore production of the compound of interest would cease.

Sporulation is initiated by activation of the master transcription regulator SpoOA. Activation of SpoOA leads to synthesis of the transcription factors SigmaF and SigmaE. These regulatory components present targets for disruption of the sporulation mechanism.

SpoOA

SpoOA is a transcriptional regulator controlling the initiation of sporulation, the development of competence for DNA uptake and many other stationary phase-associated processes (Harris et al., 2002). As a master regulator, SpoOA is responsible for the cellular stress response resulting in many phenotypic changes, including sporulation.

The SpoOA transcription factor initiates sporulation and solvent production in *C. acetobutylicum* and *C. beijerinckii* by activating transcription of *adc* encoding acetoacetate decarboxylase, *adhE* and *ctfAB* encoding CoA transferase genes. The transcription factor binds to the *ptb* promoter, resulting in downregulation of *ptb* gene expression encoding phosphotransbutyrylase. This results in enhancement of butanol production (Zheng et al., 2009). Although SpoOA activity results in increased *n*-butanol production, this effect is short lived as the initiation of sporulation results in a metabolically inactive, dormant cell. By preventing or delaying sporulation it may be possible to produce more solvent, including *n*-butanol by prolonging the duration of active culture.

In-frame deletion of *spoOA* was carried out according to the method described in Sheng et al. (2017). The protocol utilised a $\Delta pyrE$ *G. thermoglucosidasius* strain, previously created by Sheng et al., (2017). The $\Delta pyrE$ strain provided a counter-selection marker by uracil auxotrophy and resistance to 5-fluoroorotic acid (5-FOA). A *spoOA* knock-out vector, pMTL_LS5k_ *spoOA*, was constructed by addition of left and right homology arms (LHA/RHA) of 500 bp upstream and

downstream of the *spoOA* gene, using NEBuilder HiFi assembly. Homology arm DNA fragments were amplified from *G. thermoglucosidasius* genomic DNA using primers detailed in appendix A1. The pMTL_LS5k_ *spoOA* vector was transformed into the Δ *pyrE* *G. thermoglucosidasius* strain. Colonies from the transformation were re-streaked on CBM1X without uracil and incubated at 60°C, followed by re-streaking colonies on CBM1X with 20 µg/ml uracil and 300 µg/ml 5-FOA. Colonies were screened by colony PCR using primers flanking the *spoOA* knock-out region. A second PCR reaction was also performed to ensure the truncated *pyrE* gene was retained. Wild type genomic DNA was used as a positive control.

This protocol resulted in no positive gene deletions. Attempts to delete the *spoOA* gene were unsuccessful with all screening results showing the full gene remained.

SigF

After attempts to generate a sporulation deficient *G. thermoglucosidasius* strain by deletion of the *spoOA* gene proved unsuccessful an alternative gene target was considered. Here *sigF* (*spoIIAC*) deletion was carried out. The *sigF* gene encodes an RNA polymerase sigma-F factor. Sigma factors are initiation factors that promote the attachment of RNA polymerase to specific initiation sites and are then released. This sigma factor is responsible for the expression of sporulation specific genes. *sigF* has been shown to be involved in the onset of *Bacillus* sporulation. Bosma et al. (2015) constructed a sporulation deficient *B. smithii* Δ *sigF* mutant strain. Thus confirming removal of the *sigF* gene resulted in a sporulation deficient phenotype, in *B. smithii* (Bosma et al., 2015).

In-frame deletion of *sigF* was carried out using the same protocol as described previously for *spoOA* deletion. However, vector pMTL_LS5k_ *sigF* was constructed to contain left and right homology arms of 500 bp of DNA up and downstream of the *sigF* gene. pMTL_LS5k_ *sigF* was transformed into Δ *pyrE* *G.*

thermoglucosidasius. Transformed colonies were re-streaked as detailed previously and screened for *sigF* deletion.

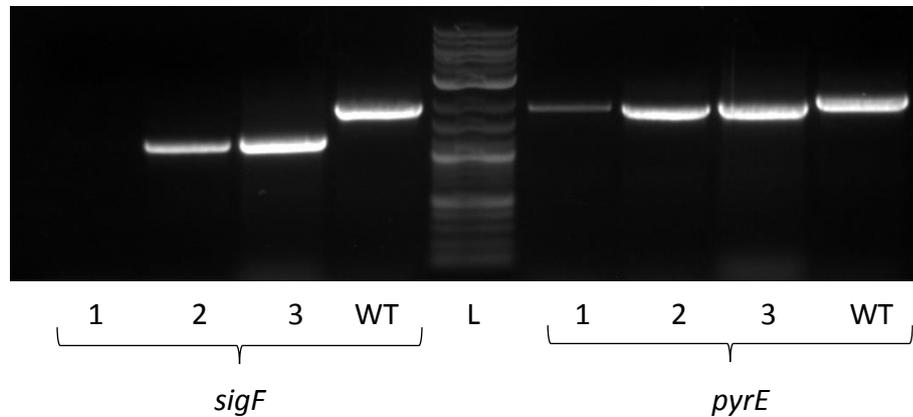


Figure 6.5. Screening of $\Delta sigF$ *G. thermoglucosidasius* mutant strains. PCR screening of three $\Delta sigF$ candidates in comparison to *G. thermoglucosidasius* wild type DNA. Isolates 2 and 3 show positive deletions of both *sigF* and *pyrE* with a reduction in band size in comparison to WT. Amplification of the *sigF* region; WT 2000 bp, $\Delta sigF$ 1282 bp. Amplification of the *pyrE* region; WT 2101 bp, $\Delta pyrE$ 1876 bp.

Sanger sequencing of the *sigF* region of isolates 2 and 3 confirmed the knock-out. A 718 bp deletion was removed from the *sigF* gene.

Following confirmation the $\Delta sigF$ strain had been successfully generated, the sporulation phenotype of *G. thermoglucosidasius* $\Delta sigF$ was characterised. Microscopic visualisation of the mutant strain showed spores still present in the culture, shown in Figure 6.6. This result indicates SigF deletion has not prevented sporulation.

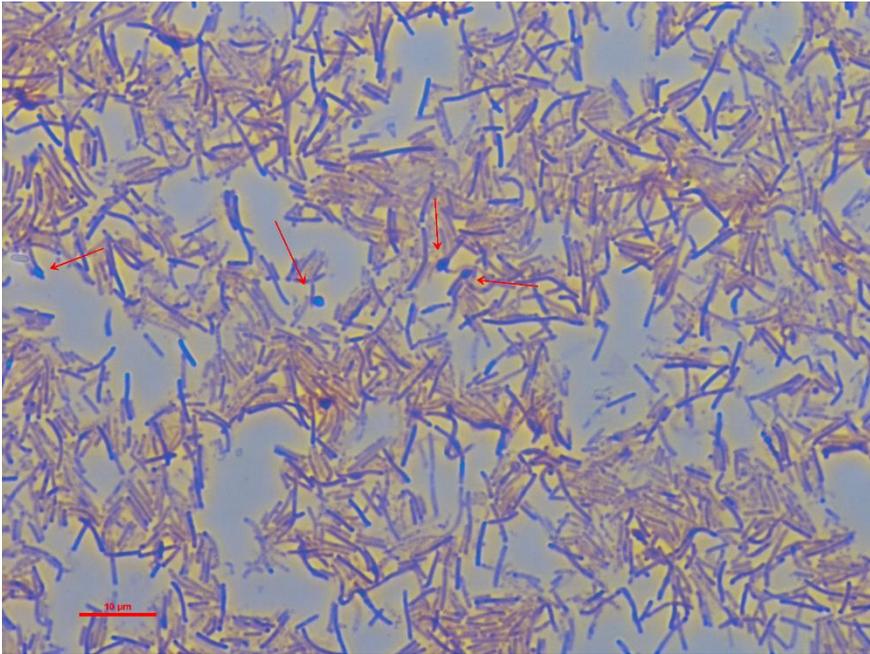


Figure 6.6. Light microscopy image of *G. thermoglucosidasius* $\Delta sigF$ cells. The strain was cultured and observed by light microscopy using 100 x objective lens with oil immersion. The scale bar represents 10 μm . Vegetative cells and terminal endospores are visible. Endospores highlighted by red arrows.

6.5 Phenotypic observations

Very little has been published with regard to the phenotypic growth properties of *G. thermoglucosidasius*. When selecting a strain for industrial production, this is an important consideration. In the forthcoming section observations of the characteristics of *G. thermoglucosidasius* culture are presented. Such observations may assist in developing an improved picture of this organism in order to better evaluate its potential for industrial application.

6.5.1 Cell aggregation

When grown in liquid culture *G. thermoglucosidasius* cells often aggregate forming clumps. This aggregation was observed in response to a variety of conditions. Cell aggregation can be induced by; lack of aeration, a drop in pH, lack of nutrients or other stress factors. When clumps were visible no further carbon consumption or metabolite production occurred.

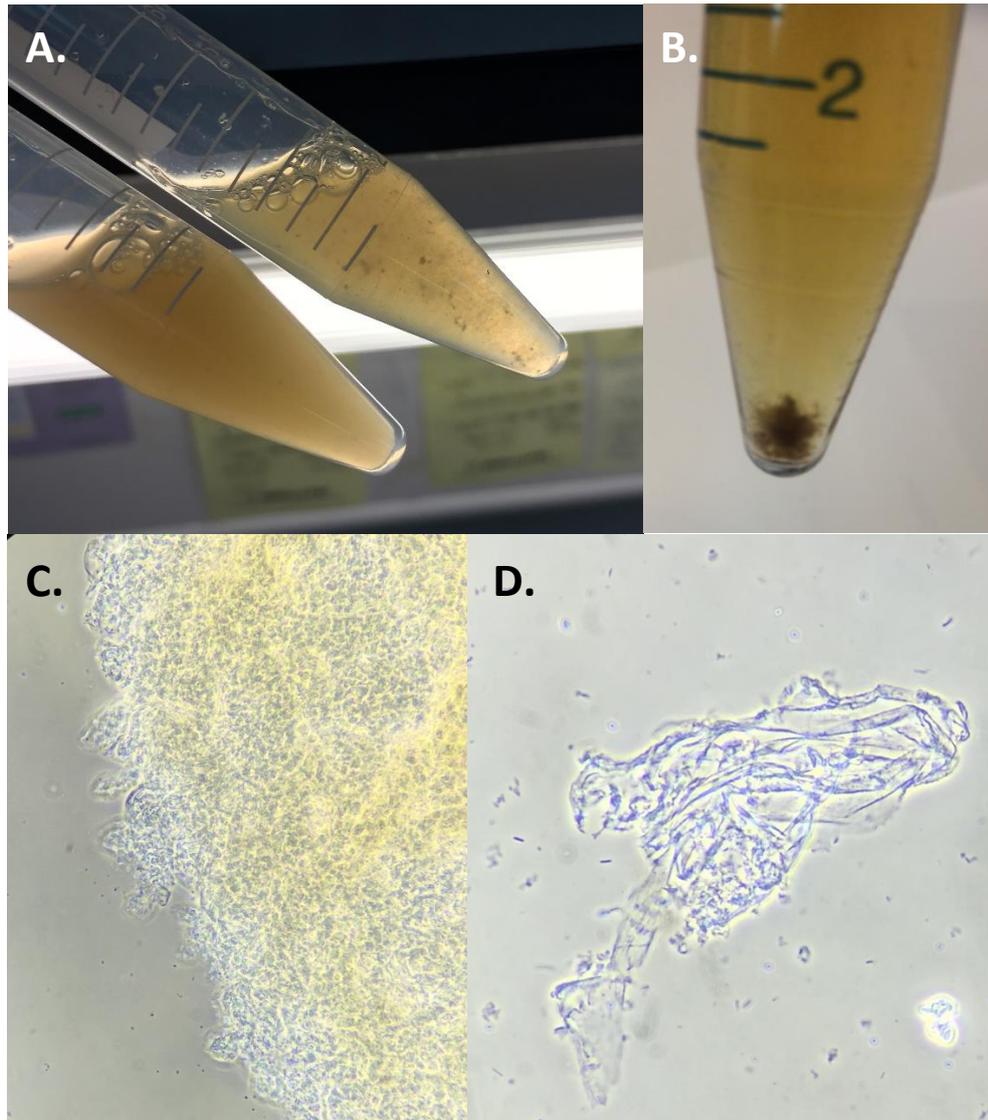


Figure 6.7. *G. thermoglucosidasius* cell aggregation. When grown in liquid culture aggregation of *G. thermoglucosidasius* cells was often observed. Here wild type cells were grown in 2SPYNG medium at 52°C with shaking of 250 rpm for 8 hours. A. Comparison of homogenous culture and aggregation with visible clumps present. B. Formation of a larger clump. C. and D. show visualisation of the material using light microscopy. Here a sample of the clump was mounted on glass slide for more detailed observation. C. 10 x magnification. D. 40 x magnification.

Microscopic visualisation of the cell aggregation shown in images C and D of Figure 6.7, indicate production of extracellular material by *G. thermoglucosidasius*.

6.5.2 Floating biofilm

A further growth characteristic of note, floating biofilms were often produced by *G. thermoglucosidasius* culture.

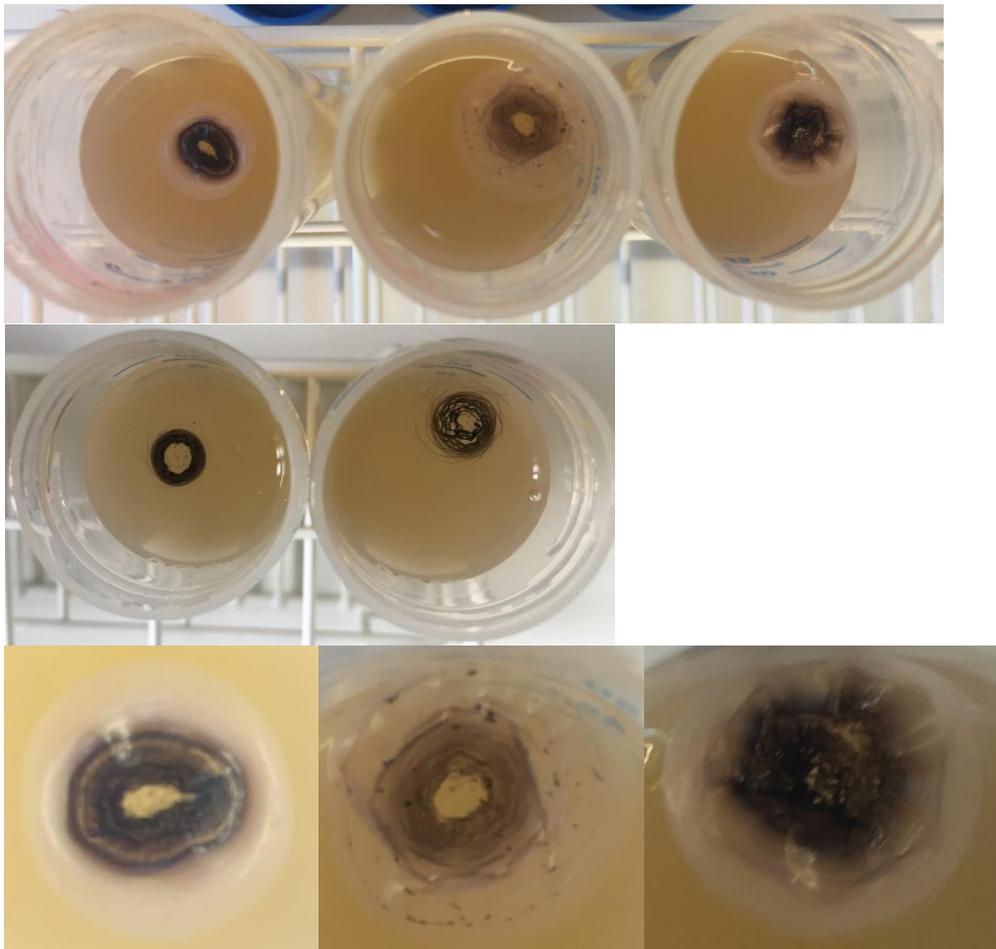


Figure 6.8. Floating biofilm formation. Wild type and modified *G. thermoglucosidasius* strains were grown micro aerobically in 40 ml UYSE medium in falcon tubes. The cultures were incubated at 52°C with 250 rpm shaking for 24 hours. All formed floating biofilms. The biofilm structure formed concentric rings with a hollow centre. The inner rings had black colouration and

the outer rings were cream in colour. The biofilms had a stringy/filamentous consistency. The structures were strong, keeping their shape when handled.

Biofilm-associated cells generate an extracellular polymeric substance (EPS). The composition of the EPS is strain dependent, with the majority comprised of polysaccharides, nucleic acids, lipids and proteins. These components are secreted by the bacteria, forming a matrix. The matrix encases the bacteria and provides a scaffold enabling bacteria to adhere to each other and to surfaces. The EPS matrix provides a protective environment, a nutrient source by recycling lysed cells and facilitates cell to cell communication by immobilising cells in close proximity.

The images presented in Figure 6.8 show the ability of *G. thermoglucosidasius* to form floating biofilm at air-liquid interfaces. Pellicle formation is seen on the surface of the culture as the biofilm requires an aerobic environment. Throughout this work, the biofilms have appeared as free floating circular structures rather than being surface adhered. This could be explained by the incubation of cultures with 250 rpm shaking imparting a mechanic force influencing the biofilm structure. However such shaking would usually be expected to inhibit biofilm formation.

Cell morphology during biofilm formation

Observations of *G. thermoglucosidasius* cells show markedly different morphologies. These different morphologies are thought to represent defined stages throughout the biofilm formation process.

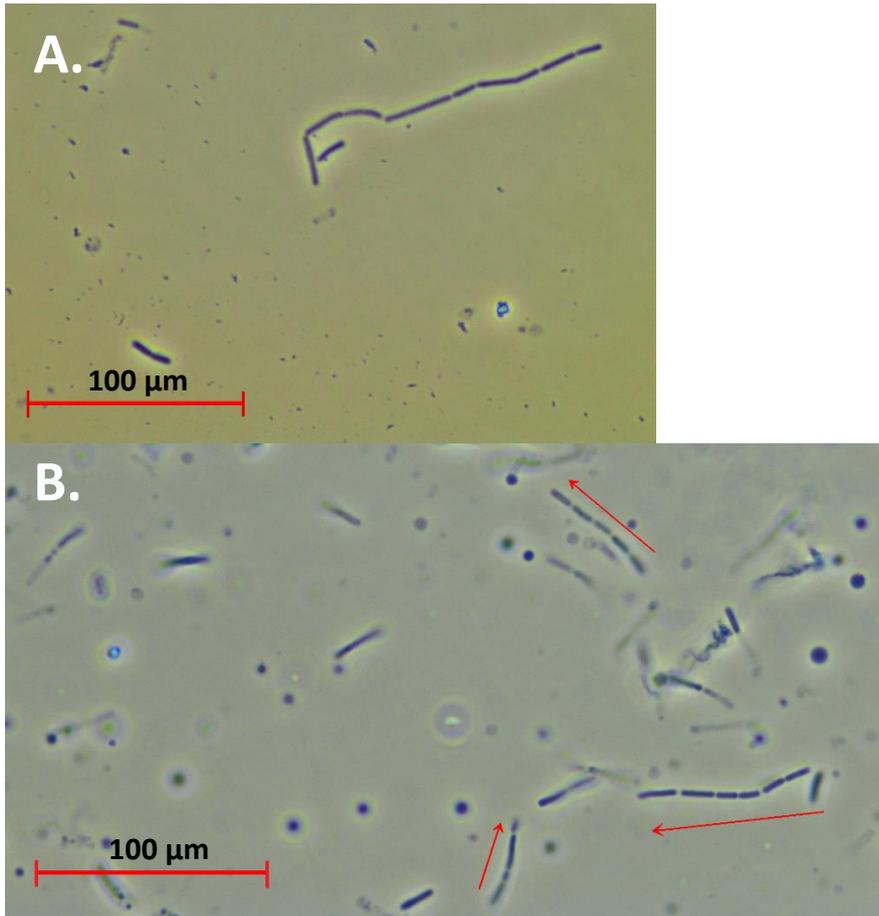


Figure 6.9. Cell morphology during biofilm formation; cells form long chains. Initial differentiation from planktonic cells sees chains of multiple cells. The chain lengths observed ranged from 2 to 13 cells. Visualisation by light microscopy x 100 magnification with oil immersion.



Figure 6.10. Cell morphology during biofilm formation; chains of cells aggregate. The chains of cells have aggregated with several chains being tightly connected forming clusters. Visualisation by light microscopy x 100 magnification with oil immersion.

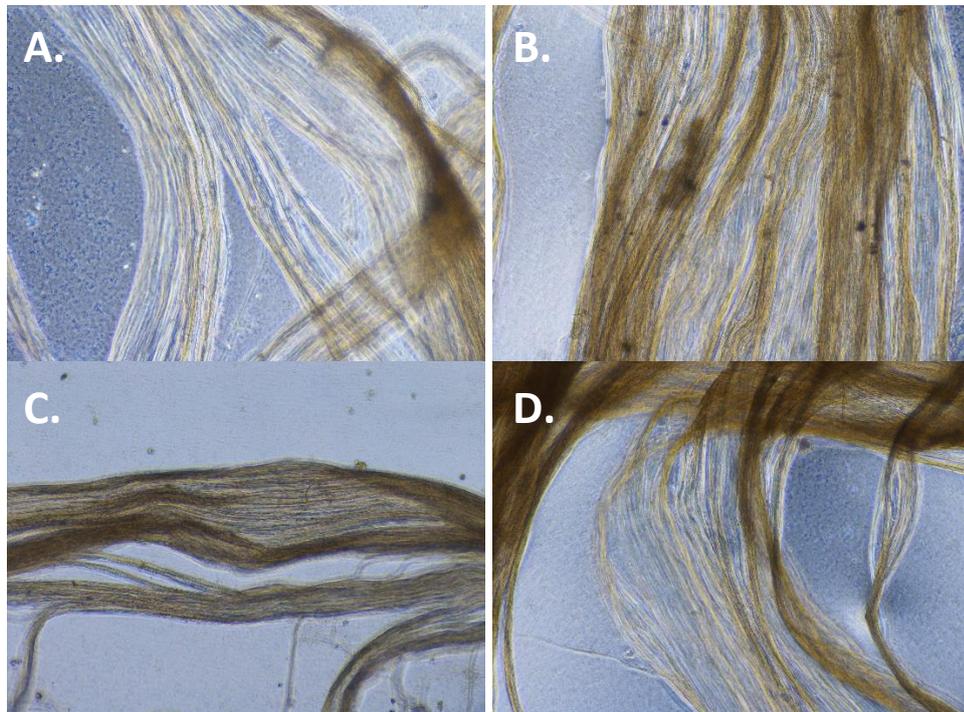


Figure 6.11. Cell morphology during biofilm formation; the clusters of cell chains form a woven string-like structure. Visualisation by light microscopy x 10 magnification.

The morphology of cells in a biofilm is markedly different from the morphology of planktonic cells. Planktonic cells exist as singular independent cells, whereas cells in a pellicle form aggregates in which cells are regularly aligned and tightly bound together. The morphologies observed, and shown in Figures 6.9 to 6.11, are concurrent with previously characterised pellicle formation in *B. subtilis* (Kobayashi, 2007). These similar observations suggests the same mechanisms are at play in both genus.

In attempt to gain a better understanding of the structure of the biofilm matrix, and the condition and location of cells within the matrix, samples of the biofilm were examined using a laser scanning fluorescent microscope. The biofilms were stained with live/dead staining using propidium iodide (PI) and SYTO9 nucleic acid binding dyes. PI is a red intercalating stain that is membrane impermeant and is therefore excluded by live cells. SYTO9 is a green intercalating membrane permeant stain and will stain all cells containing nucleic acid. This staining enabled visualisation of live and dead cells in the biofilm sample to give an indication as to the cell distribution within the matrix.

Biofilms from overnight growth of wild type *G. thermoglucosidasius* were re-suspended in 5 ml UYSE with addition of 20 g/l glucose and 100 µg/mL of 2 µM propidium iodide. The biofilms were cultured for a further 2 hours. 5 µl of 5 mM SYTO9 green fluorescent nucleic acid stain (Invitrogen) was added 30 mins prior to visualisation. The stained biofilm was visualised on a slide using a confocal laser scanning microscope (Zeiss LSM 700). PI was detected using an excitation/emission maxima of 493/636 nm. SYTO9 was detected using an excitation/emission maxima of 485/498 nm. Z-stacks were obtained and assembled to show 3D surface topography of the sample. Image assembly was carried out using the LSM 700 ZEN software.

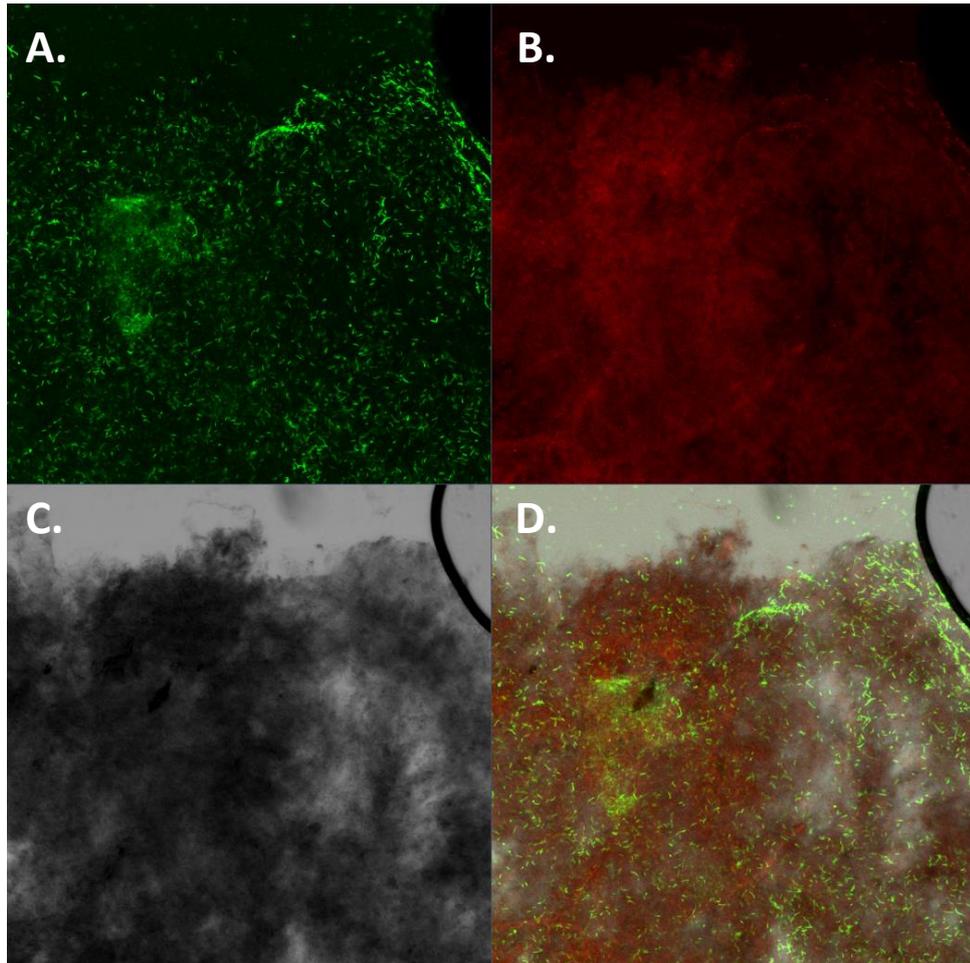


Figure 6.12. Live/dead cell staining of *G. thermoglucosidasius* biofilm with visualisation using laser confocal microscopy. The slide was visualised under oil immersion with $\times 40$ objective. A. shows live cells, B. shows dead cells, C. shows the matrix with no fluorescence and D. shows the combined live and dead channels.

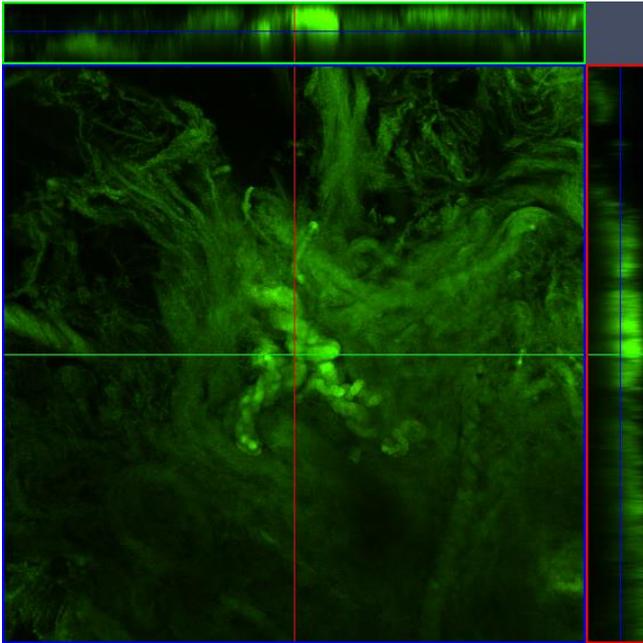


Figure 6.13. Observation of *G. thermoglucosidasius* biofilm's 3D structure visualised by Z-stacked laser confocal microscopy images. Unstained biofilm sample visualised under oil immersion with x 10 objective. This top view image shows the 3D nature of the biofilm and the woven string like nature of the composition.

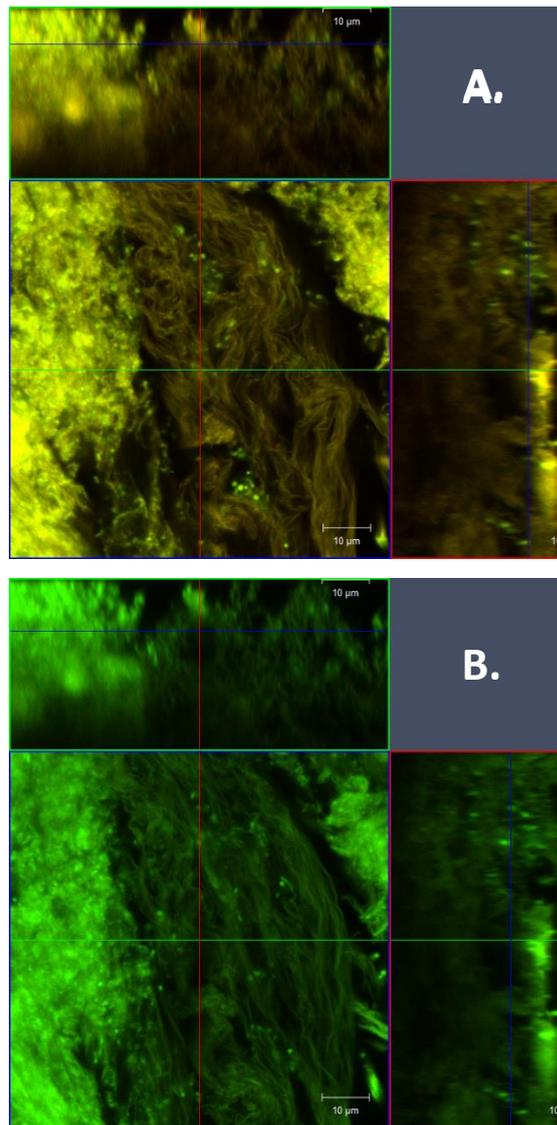


Figure 6.14. Observation of *G. thermoglucosidasius* biofilm, visualised by Z-stacked laser confocal microscopy images. Unstained biofilm sample visualised under oil immersion with x 10 objective. A. and B. show different depths in the Z axis. These images show the structure of the EPS matrix.

6.6 Discussion

6.6.1 *n*-butanol toxicity and directed evolution

Butanol toxicity is one of the most significant problems facing the ABE fermentation process. *n*-butanol is highly toxic to both natural producers and engineered hosts. *n*-butanol causes disruption to membranes, membrane related functions and metabolic networks. Its toxicity is due to its hydrophobic

nature, disrupting the phospholipid component of bacterial cell membranes, increasing membrane fluidity. Toxicity of the product presents a major challenge for bio production processes as large titres of *n*-butanol are required for economic efficiency.

The *n*-butanol inhibitory concentration of *G. thermoglucosidasius* was found to be 1% (v/v). To be considered a potential *n*-butanol production organism, it is desirable to increase this tolerance. With limited knowledge of *Geobacillus*' regulatory mechanisms, rational design of an engineering strategy for a more tolerant phenotype would likely have been unsuccessful. In this work serial passaging in the presence of increasing *n*-butanol concentrations, or directed evolution, resulted in selection of a strain with increased *n*-butanol tolerance of 2.5% (v/v). This showed the engineering of complex traits can be accomplished using a simple directed evolution strategy. Directed evolution mimics natural evolution in more controlled settings to evolve biological systems toward a desired phenotype. Whole genome sequencing of the adapted strain was carried out in attempt to determine the cause of the phenotypic change. The sequencing results identified a number of variants which could be responsible for the phenotypic change. Further investigation is required to confirm the effects caused by the gene mutations identified.

It is unclear whether an increase in tolerance will result in increased production levels. Previous studies have reported both a *C. acetobutylicum* strain with increased *n*-butanol tolerance showed improved final solvent titre (Tomas et al., 2003) and an iso-butanol tolerant strain which did not improve iso-butanol production (Atsumi et al., 2010). Such conflicting reports indicate toxicity measured by growth is not a prediction of production titre. Further research is required to investigate the underlying physiology and mechanisms of the relationship between solvent tolerance and solvent production.

6.6.2 Sporulation mutant

Despite seemingly few suitable environments for *Geobacillus* growth, the organism is widespread in the environment. It is predicted this is due to wide distribution of spores, which suggests sporulation plays an important role in *Geobacillus* survival. A lack of success in attempts to create a sporulation mutant indicates the sporulation characteristic could be essential. Although Silvaggi et al. (2004) identified genes that are activated during sporulation on a genome-wide scale and found only a small subset appeared to be essential for sporulation (Silvaggi et al., 2004). The SigF deletion did not prevent spore formation. In *B. subtilis* Sigma factor F is the first forespore specific transcription factor, controlling genes required for the early stages of prespore development. It was hypothesised this would possess the same function in *Geobacillus*, however this may not be the case. Alternatively *Geobacillus* may possess multiple copies of the gene, with other intact copies able to rescue the phenotype. The *sigF* knock out may only reduce activity rather than prevent sporulation altogether. In this case, the sporulation phenotype will not be deleted unless all copies of the gene are removed. The lack of phenotype change could be due to more complex mechanisms. Sporulation will be regulated by many genes and transcription factors. The role of SigF could be masked by other regulatory elements.

Although SpoOA was targeted for deletion this may not be beneficial for *n*-butanol production. Alsaker et al. (2004) reported that overexpression of *spoOA* increased tolerance to butanol stress and prolonged metabolism. Tomas et al. (2004) also constructed an engineered strain with overexpressed *spoOA*. Their results show higher *n*-butanol production compared to a *spoOA* inactivated strain (Tomas et al., 2004). However the strains with *spoOA* overexpression failed to produce more solvent due to an accelerated sporulation process (Harris et al., 2002). Therefore an alternative prevention mechanism should be targeted in future.

6.6.3 Phenotypic observations of clumping and biofilm formation

Biofilm formation is disadvantageous for industrial fermentation. Biofilm formation results in reduced growth rates, production of components for biofilm formation costs the bacterium energy that could be used for product formation, the EPS matrix limits the diffusion of nutrients, and as a result the cells in a biofilm are slow growing and starved. Biofilms also result in increased resistance to antibiotics. The observations of both cell aggregation and biofilm formation could have major implications for *Geobacillus*' use as a successful industrial process organism. A viable process organism should have a reliable and predictable growth profile. Greater understanding of the conditions in which biofilm cell aggregation is induced and the effects on fermentation is required in order to control and manage the culture conditions appropriately. Extending the working life of cells would increase the product titre and reduce the frequency of down-time between batches of cell growth.

One method for halting the formation of biofilms might be the development of a biofilm deficient strain by genetic manipulation. *ymcA* is the master regulator for biofilm formation in *B. subtilis*. This gene could be a possible target in *G. thermoglucosidasius*. BLAST analysis of the *B. subtilis* protein sequence against the wild type *G. thermoglucosidasius* NCIMB 11955 strain matched with 63% identity and 98% cover to a hypothetical protein. Alternatively the *yxkB* gene participates in pellicle formation at the air-liquid interface in *B. subtilis* (Nagórska et al., 2008). This presents another deletion target for future investigation.

6.6.4 *Geobacillus* as an industrial host strain

Geobacillus has been tipped as a promising industrial organism. However, to date, little research has been conducted to in order to determine if this is the case. *Geobacillus* does offer some clear advantages. Industrial production processes favour a thermophilic strain due to reduced cooling costs, less risk of contamination and ease of product separation. *Geobacillus* is able to assimilate

a range of substrates and therefore consolidated bioprocessing could be engineered in future. Strain characterisation described here found no prophage in the *G. thermoglucosidasius* genome. A lack of prophage sequences and a high number of CRISPR elements indicate a low risk of phage infection. However previously unreported and undesirable characteristics have also been noted. The wild type *G. thermoglucosidasius* had low tolerance of *n*-butanol. *Geobacillus* readily sporulates. To date an asporulant mutant strain has not been generated. In addition *G. thermoglucosidasius* clump and form biofilms during growth, making culture unreliable and inconsistent. Genetic manipulations can be hampered by the organism's ability to recombine readily, as such any strains generated may not be genetically stable. However despite potential drawbacks, the increase to *n*-butanol tolerance reported here has resulted in an improved strain with more desirable properties for industrial use. Although further work is required, *Geobacillus* could offer benefits in an industrial setting. The adapted evolution technique used here, demonstrates the organism is highly adaptable, a trait required for survival in harsh conditions.

CHAPTER 7
DISCUSSION

7.1 Background

Climate change and the burning of fossil fuels is a global issue. A continued increase in the use of renewable energies is of vital importance to protect the world's climate. Further use of biofuels as a replacement for liquid transport fuels is required to mitigate the effects of greenhouse gasses and as an alternative to finite fossil fuels. To make biofuel innovations sustainable, commercialisation on a massive scale is needed. Advanced biofuels must be economically competitive with existing products, overcoming the primary economic drivers of feedstock price, process productivity and yield. Commercialisation is supported by governments across the world with policies in place to encourage biofuel uptake. However much progress is yet to be made in order for second, third and fourth generation biofuels to become feasible and commercially viable. In this work production of the biofuel *n*-butanol was targeted. *n*-butanol is a desired product as an advanced fuel with high energy content, compatible with existing petroleum infrastructure. In comparison to bioethanol addition of two carbons significantly decreases its water solubility and volatility while increasing energy density.

Geobacillus with its natural ability to ferment various substrates, including hexose and pentose sugars, offers advantages as a process organism. As a thermophilic bacterium, *Geobacillus* is a promising potential industrial organism. Limited research had been carried out with respect to utilisation of *Geobacillus* as a process organism. Little is known of the complex regulatory networks controlling *Geobacillus* metabolism. The findings of this research provide insight into the potential challenges which may be faced when working with this bacterium.

7.2 Key findings of this work

7.2.1 Development of new tools for Gram-positive thermophiles

In recent years there has been rapid developments in the synthetic biology field. Many advanced synthetic biology tools are available. Such tools are

applicable to engineering of microbial cells for fuel production. The focus, to date, has mainly benefited the commonly used industrial workhorses *E. coli* and *S. cerevisiae*. The tools available for alternative, more novel organisms, such as *Geobacillus* remains limited. Development of basic tools required for bacterial modification was the first obstacle for this work. The tools developed facilitated subsequent metabolic engineering.

A set of reporter genes was the first necessity, to enable screening and measurement of other components. A screening of promoters was carried out to enable transcriptional engineering. A selection of Gram-positive replicons were identified and characterised. Transcript levels were also controlled by plasmid copy number. The copy number of plasmids featuring a range of Gram-positive replicons was determined. Modifications in the copy number of a gene can now be achieved by selection of the replicon for recombinant expression plasmids or by chromosomally integrating the gene. RBS screening was carried out to enable translational engineering. Synthetic, native and heterologous RBS's were characterised, providing a library for selection of desired translation strength. Codon optimization is also used to improve translation in the cell. Expansion of the modular vector series provided multiple options for each of the vector components. This provides greater flexibility for the researcher. Screening in alternative bacterial hosts also increases the potential use of the vectors, which is not limited to only *G. thermoglucosidasius*. Antibiotic sensitivities were determined and the corresponding resistance markers tested for functionality. Addition of components to the vector series enabled construction of two wholly unique plasmids. These plasmids were co-transformed into *G. thermoglucosidasius*. This was the first demonstration of co-transformation of multiple plasmids in *Geobacillus*. Such use of multiple plasmids assists in introduction of large or multi enzyme pathways, easing the construction process.

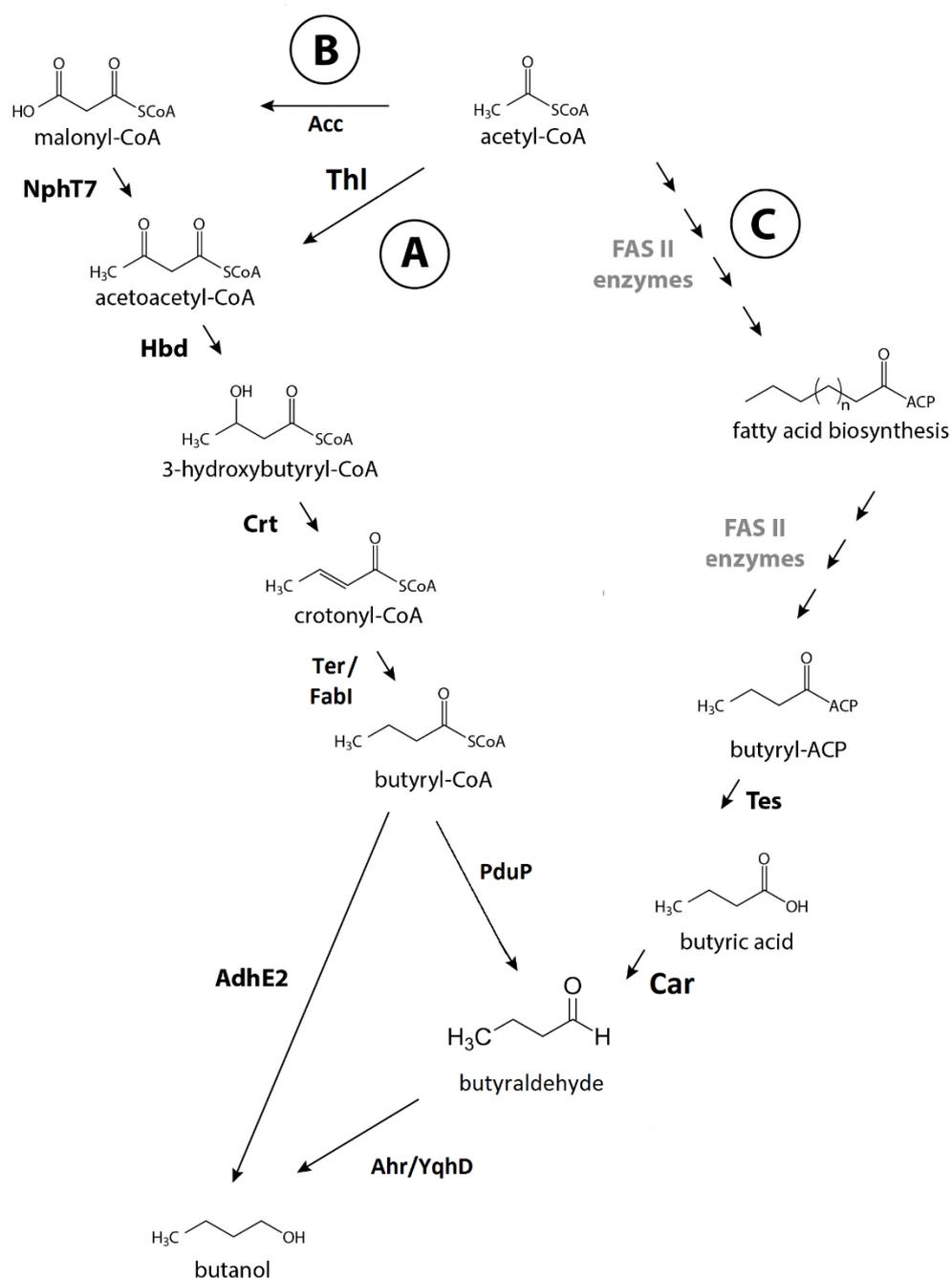
7.2.2 *n*-butanol pathways investigated

Figure 7.1. Schematic representation of the heterologous pathways for *n*-butanol biosynthesis investigated. All pathways are initiated from acetyl-CoA. Routes A and B are CoA dependent whereas route C is CoA independent. **Route A** was first considered with introduction of synthetic enzymes; Thl (thiolase), Hbd (3-hydroxybutyryl-CoA dehydrogenase), Crt (3-hydroxybutyryl-CoA

dehydratase), Ter (trans-2-enoyl-CoA reductase) and AdhE2 (aldehyde-alcohol dehydrogenase). Secondly use of enzymes native to the host strain was considered. Route A was again followed with FabI replacing Ter to give an all native pathway. Following a lack of *n*-butanol production modifications resulted in **Route B** Acc (acetyl-CoA carboxylase) and NphT7 (acetoacetyl-CoA synthase) replacing Thl to create a driving force at the initial step, PduP and YqhD replacing the incompatible AdhE2. This pathway was successful in generating *n*-butanol. **Route C** diverts endogenous fatty acid biosynthesis to convert butyryl-ACP to *n*-butanol using enzymes; Tes (acyl-ACP thioesterase), Car (carboxylic acid reductase) and Ahr (aldehyde reductase). FAS II enzymes, coloured in grey, represent native fatty acid synthase type II.

Route A is the pathway found in *Clostridium* spp., which produce butanol as part of native ABE fermentation. This pathway has been engineered in various host organisms such as *C. acetobutylicum* (Hou et al., 2013; Jang et al., 2012; Zheng et al., 2009), *C. tyrobutyricum* (Yu et al., 2011), *E. coli* (Bond-Watts et al., 2011; Shen et al., 2011), *S. cerevisiae* (Steen et al., 2008) and *S. elongatus* PCC 7942 (Lan & Liao, 2011, 2012). When introduced into *G. thermoglucosidasius* this pathway was found to be incompatible due to the anaerobic and mesophilic nature of clostridia. This CoA dependent pathway was compromised in the presence of oxygen. The AdhE was not bi-functional and Thl strongly favoured the reverse direction. A range of alternative aldehyde and alcohol dehydrogenase genes were screened in order to identify a replacement. This screening work determined those which were not compatible with the host strain and growth conditions required and also confirmed functionality of an enzyme pair. After modifications were made to the pathway, replacing some of the component genes, *n*-butanol production was achieved. Despite very low quantities of product, this result demonstrates the first *n*-butanol production in thermophilic and aerobic conditions. Metabolic engineering of the cell enabled production of a non-native compound.

n-butanol can be produced by diversion of the bacterial fatty acid synthesis pathway in a CoA independent pathway, route C of Figure 7.1. Intercepting the bacterial fatty acid synthesis pathway with a C4 specific thioesterase, terminating fatty acid elongation and generating a butyric acid intermediate. This route is reportedly O₂ tolerant. Expression of the pathway genes in *G. thermoglucosidasius* did not result in *n*-butanol production. The cause was not determined. However if the issue can be elucidated in the future, this route offers potential for increased yields in *G. thermoglucosidasius* as this ACP dependent pathway is stimulated by enhanced O₂ availability. *G. thermoglucosidasius* requires oxygen for growth, therefore this pathway has potential as a more compatible route, better suited to the *Geobacillus* host strain.

7.2.3 Host strain characterisation

Butanol toxicity is one of the most significant problems facing the ABE fermentation process. Here, the *n*-butanol inhibitory concentration of wild type *G. thermoglucosidasius* was found to be 1% (v/v). This was lower than the desired tolerance levels for an industrial process organism. A directed evolution strategy, serial passaging in the presence of increasing *n*-butanol concentrations, resulted in selection of a strain with the desired phenotype. The tolerance of *G. thermoglucosidasius* to *n*-butanol was increased to 2.5% (v/v). This showed the engineering of complex traits can be accomplished using simple directed evolution strategies. Directed evolution mimics natural evolution in more controlled settings to evolve biological systems toward a desired phenotype, allowing the cell to rewire itself in adaptation to its environment. The increased product tolerance provides a more suitable and robust production strain.

7.3 Limitations

The microbiological challenge for biofuel production processes is to improve the carbon assimilation pathways in a microbial host and channel the metabolic

flux of these pathways to produce a desirable product through natural or synthetic pathways. Although some biofuels and their intermediates are produced from various metabolic pathways that exist naturally in the microorganism, the pathways often need to be optimized or redesigned to improve efficiency. The yields and productivity of current bioprocesses are often too low to justify industrial application. For example, the current biological production of *n*-butanol produced by fermentation of sugars from biomass has disadvantages such as low butanol yield and by-product formation. Alternatively, butanol can be directly produced from ethanol through aldol condensation over metal oxides/hydroxyapatite catalysts. This chemical conversion route is more preferable than the biological process, as the reaction proceeds more quickly compared to the fermentation route and fewer steps are required to reach the final product (Ndaba & Marx, 2015). In order for biological production to compete with chemical conversion, much improvement in yield and processing conditions is required.

Biofuel production requires high volumes and low cost. Therefore high yield and high productivity are required. Achievement of such high yield, titre and productivity remains a challenge. The challenges associated include; improving the enzyme activities of an entire metabolic pathway, increasing tolerance of the host toward toxic target compounds, recycling or replacing insufficient cofactors for enzymatic reactions, enriching precursors, eliminating by products, optimizing and balancing the fluxes of whole metabolic networks to reduce burden on the host and removing negative feedback regulation.

Metabolic engineering approaches frequently fail to lead to the desired phenotypes because of unclear or complex gene structures, functions, and regulations in cellular metabolic networks (Yu et al., 2014). In order to guide more rational design strategies, knowledge of the overall network of pathways in organisms is required, rather than focusing on individual pathways in isolation. Introduction of a non-native biochemical pathway in a dynamic

metabolic network is a challenge in any host. This challenge is increased when working with a relatively novel organism. Multiple analytical and modelling tools, such as genomics, transcriptomics, proteomics, metabolomics, fluxomics, high-throughput screening and in silico studies have been utilized to elucidate metabolic engineering detail. Such studies provide useful information to predict metabolic interactions, guide strain design and maximize the efficacy of metabolic engineering. Such work has not yet been carried out for *Geobacillus*. Therefore the information is lacking to assist when working with this host strain.

Further limitations encountered during this work include; identification of compatible enzymes which are both thermostable and oxygen tolerant. The enzymes must be functional in the growth temperature range of the host organism, which in this case is thermophilic. This limits use of many enzymes which are not thermostable. Although there are native thermophilic *n*-butanol producers found in nature, none result in consistent or high yield production. Oxygen tolerance presented a further challenge as all native *n*-butanol production is exclusively anaerobic. Therefore inherent incompatibilities exist. Although an oxygen tolerant route to *n*-butanol has been reported (Pasztor et al., 2014), expression of this pathway did not result in product formation here. A further major design challenge was the reversibility of this pathway. The reactions are present in the cell for alternative functions, with some strongly favouring the reverse direction. Therefore expression of a set of functional enzymes is insufficient to result in product formation. If the desired reactions do occur, the intermediates produced are used in a range of other pathways thus diverting flux away from the intended outcome.

Genome sequencing of *Geobacillus* strains show a significant amount of genome rearrangement within the genus. With large genomic islands encompassing all the hemicellulose utilization genes and a genomic island incorporating a set of long chain alkane monooxygenase genes (Hussein et

al., 2015). This ability of *Geobacillus* to rearrange its genetic material presented difficulties when attempting to introduce heterologous genes. *Geobacillus* is a highly adaptable organism, it is therefore able to mitigate against factors posing a detrimental effect on its survival performance. As introduction of new genes and pathways doesn't provide a benefit to the host, it is able to remove these components reducing cellular burden. Thus *Geobacillus* was, at times, an intractable host.

7.4 Future perspectives

Despite limitations, the observation of any *n*-butanol production could be considered sufficient to justify further work. Although much further development is required in order to reach the point of commercial viability. This research presents only a brief introduction into the dearth of research opportunities and unanswered questions regarding the metabolic engineering of *Geobacillus*.

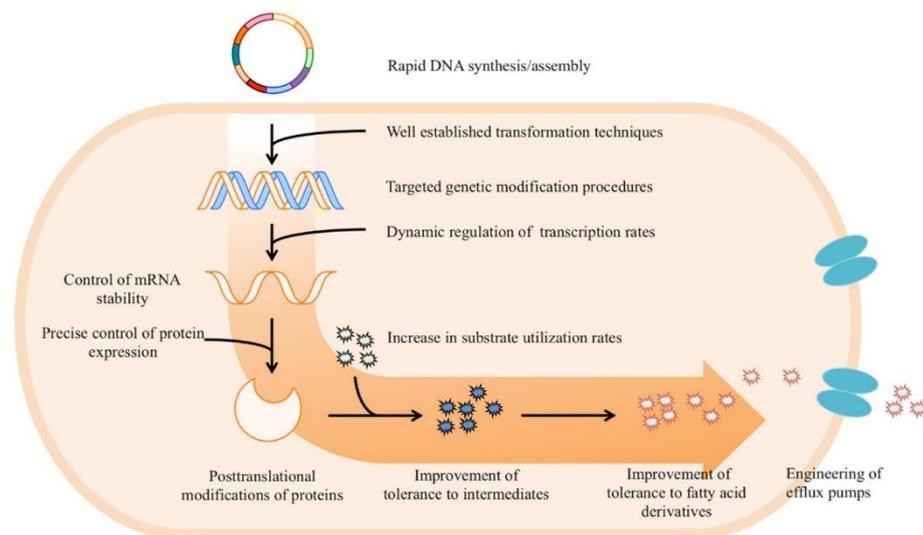


Figure 7.2. Overview of potential applications of synthetic biology tools for construction and optimization of metabolic pathways (Yu et al., 2014).

Every stage of the metabolic pathway can be optimised, as shown in Figure 7.2. Aspects of this pathway have not been considered in this work. For example engineering of efflux pumps to increase product out of the cell. This could lead

to increased product formation if the flux is pulled through the pathway, it would also reduce any effects of product toxicity.

Future work should attempt to combine many of the factors presented in this work into a single host strain. The functional *n*-butanol pathway was composed of plasmid located genes; *nphT7*, *hbd*, *crt*, *yqhD* and *pduP* with the genes *acc* and *ter* integrated into the genome. With Δldh , Δpfl , $\Delta adhE$ and *pdh^{UP}*, plus deletion of acetic acid synthesis genes. In combination with the evolved *G. thermoglucosidasius* strain which is highly *n*-butanol tolerant and a sporulation mutant. Such modifications may improve the metabolic flux through the pathway by eliminating competing metabolism of the host organism and may result in increased production titres. It may also result in a more industrially relevant strain producing a single product, being tolerant to that product and also not able to sporulate during processing.

Currently protein production, growth and product formation are linked. Future efforts should attempt to uncouple biomass and product formation. If cells can be maintained in a quiescent state metabolic fluxes can be channelled to product formation.

7.4.1 Alternative potential pathways

There remains the possibility of exploring alternative production pathways. Branduardi et al. (2013) describe a novel pathway to produce butanol in *S. cerevisiae* using amino acid degradation, with glycine as a substrate. Utilising proteins derived from exhausted microbial biomass at the end of the fermentation process. The Ehrlich, or 2-keto-acid, pathway has also been explored for biofuel production by Lan and Liao (2013) who utilized the amino acid biosynthetic pathway of *E. coli* for alcohol synthesis. This pathway decarboxylates amino acid precursors, into aldehydes and reduces them to alcohols. Amino acid based alcohols include; *n*-propanol from isoleucine, isobutanol from valine and *n*-butanol from norvaline.

Ng et al. (2015) isolated a microorganism capable of producing butanol in high concentrations under aerobic conditions from soil samples in Singapore. The isolate was identified as a *Bacillus* species by 16S rDNA analysis. At 1 l scale fermentation with 0.2 l/min airflow, butanol titre reached 10.38 g/l in batch culture. Growth and butanol production of the isolated *Bacillus* sp. under both aerobic and anaerobic conditions, together with a different composition of by-products, suggested different metabolic networks from the native production in *Clostridium*. The specific pathway used by this *Bacillus* is yet to be reported. This suggests nature may already have a similar thermophilic solution present in the environment. If this is the case high throughput characterisation followed by manipulation to improve titre may be a more appropriate approach.

7.4.2 Use of weak promoters

Throughout this work, gene expression has relied upon the strong promoters' P_{ldh} and P_{GAPD} . Recently reports have suggested this may not be the best approach to achieve maximum expression levels. There may be the potential to improve production by use of weak promoters. Bond-Watts et al. (2011) assembled a synthetic pathway for *n*-butanol production in *E. coli*. They observed a 30 fold increase in product yield resulting from a decrease in promoter strength. They concluded that the major bottleneck in the pathway was the solubility and productivity of enzymes *in vivo*. Low expression could limit the protein overexpression burden and reduce any insoluble protein production.

For further control of expression levels it is possible to tune the expression level of several genes within operons using tuneable intergenic regions (TIGRs). Pflieger et al. (2006) generated and screened libraries of TIGRs containing control elements that include mRNA secondary structures, RNase cleavage sites and RBS sequestering sequences. Through an operon reporter system containing the genes encoding red and green fluorescent proteins, they

showed that TIGRs can vary the relative expression of two reporter genes over a 100-fold range. Using the TIGR approach, they balanced the expression of three genes in an operon that encodes a heterologous mevalonate biosynthetic pathway, resulting in a 7 fold increase in mevalonate production in *E. coli*. This technique could be applied here to optimize the expression of multiple genes in synthetic operons.

7.4.3 Balancing redox, cofactor generation and ATP synthesis

Maintaining the cellular redox balance is a basic requirement for living cells to sustain metabolism. Redox balance also plays an important role in the production of chemicals. Introduction of a heterologous pathway will alter the redox state of the cell and therefore the metabolic flux. For optimal biofuel production the host strain should maintain maximal carbon flux towards the target metabolite with no fluctuations in redox.

Redox balance can be maintained through a variety of techniques including; regulating gene expression levels, genome-scale engineering removing non-essential reactions, structural synthetic biotechnology such as DNA, RNA and protein scaffolds, systems metabolic engineering creating new metabolic pathways or cellular regulatory circuits, computational techniques, protein engineering to improve enzyme activity, change substrate specificity, modify cofactor specificity or construct multi-enzyme complexes. Cofactor regeneration can be used to alter the intracellular cofactor pool.

Ethanol fermentation in yeast and other ethanologenic hosts can occur at high yield due to precise redox balance with glycolysis in terms of NAD^+/NADH recycling in the presence of a non-oxidative pyruvate decarboxylase. In contrast, *E. coli* fermentation results in the production of mixed organic acids that favour direct reduction of pyruvate to lactate and the secretion of acetate as a product of glycolytic 'overflow' (Bond-Watts et al., 2011). As ethanol, *n*-butanol production from pyruvate can provide redox

balance with glycolysis to regenerate NAD⁺ for ATP synthesis. This offers the potential to realize yields from glucose similar to ethanol, if the host can be engineered to use *n*-butanol over native fermentation pathways.

Previous metabolic engineering studies have demonstrated that redox factor usage can significantly affect the yield of high-flux pathways because of differential regulation of NADH and NADPH pools. A change of enzyme cofactor from NADPH to NADH by an engineered enzyme enabled the anaerobic production of isobutanol at 100% theoretical yield in *E. coli* (Liao et al., 2016). The production of *n*-butanol and *n*-propanol also benefited from cofactor balancing in which an NADH-producing or NADPH-producing pathway is matched by a consumption pathway using the same redox cofactor.

7.5 Concluding remarks

In this work, molecular tools were developed to enable metabolic engineering of *Geobacillus*. A series of pathways for *n*-butanol biosynthesis were designed, constructed and evaluated in a *G. thermoglucosidasius* host. CoA dependent and CoA independent pathways were considered. Use of heterologous and native genes were considered. Individual enzymatic steps were investigated to characterise oxygen tolerant and thermostable enzymes. Modifications were made to the pathway design resulting in a functional pathway capable of *n*-butanol production. However the yields remain far from commercially relevant. Productivity and titres remain to be improved. Although efficient production of butanol as a major fermentation product has not been demonstrated here, the production of *n*-butanol demonstrates proof of concept and offers a platform to build upon, utilising the strain improvements previously discussed.

This work has demonstrated that the introduction of heterologous pathways into microbial host without an understanding of complex regulatory networks underlying biosynthesis pathways, will unlikely result in high-level production

of target chemicals, especially when working with a novel thermophilic host organism. However, the ever increasing development of novel synthetic biology tools, offers great potential for future work to elucidate complex regulatory networks, enhance gene expression, increase enzyme activities and substrate specificity, improve metabolic flux and boost product titre in heterologous microbial hosts. Ultimately, successful engineering strategies will be the key to generating efficient microbial based production of valuable fuels and chemicals.

As a thermophilic organism, *Geobacillus* was thought to offer advantages making it a promising potential industrial organism. Issues encountered during this work indicate this may not be the case. However, in order to develop efficient systems for commercial biofuel production, exploration of alternative platforms and strategies is essential.

BIBLIOGRAPHY

- Abdehagh, N., & Bagheri, M. (2014). Improved Acetone-Butanol-Ethanol (ABE) Solution Analysis Using HPLC: Chromatograph Spectrum Deconvolution Using Asymmetric Gaussian Fit. *American Journal of Analytical Chemistry*, 5(November), 1078–1089.
- Akhtar, M. K., Turner, N. J., & Jones, P. R. (2013). Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), 87–92.
- Aliyu, H., Lebre, P., Blom, J., Cowan, D., & De Maayer, P. (2016). Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Systematic and Applied Microbiology*, 39(8), 527–533.
- Anné, J., Vrancken, K., Van Mellaert, L., Van Impe, J., & Bernaerts, K. (2014). Protein secretion biotechnology in Gram-positive bacteria with special emphasis on *Streptomyces lividans*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8), 1750–1761.
- Araújo, K., Mahajan, D., Kerr, R., & Silva, M. da. (2017). Global Biofuels at the Crossroads: An Overview of Technical, Policy, and Investment Complexities in the Sustainability of Biofuel Development. *Agriculture*, 7(4), 32.
- Atsumi, S., Hanai, T. & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451, 86–89.
- Atsumi, S. & Liao, J. C. (2008). Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and butanol by *Escherichia coli*. *Appl. Environ. Microbiol.* 74, 7802–7808.
- Atsumi, S., Higashide, W. & Liao, J. C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* 27, 1177–1180.
- Atsumi, S., Wu, T. Y., Eckl, E. M., Hawkins, S. D., Buelter, T. & Liao, J.C. (2010). Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Appl. Microbiol. Biotechnol.* 85, 651–657.
- Bacon, L. F., Hamley-Bennett, C., Danson, M. J., & Leak, D. J. (2017). Development of an efficient technique for gene deletion and allelic exchange in *Geobacillus* spp. *Microbial Cell Factories*, 16(1), 58.
- Baez, A., Cho, K. M., & Liao, J. C. (2011). High-flux isobutanol production using engineered *Escherichia coli*: A bioreactor study with in situ product removal. *Applied Microbiology and Biotechnology*, 90(5), 1681–1690.

- Baral, N. R., Li, J., & Jha, A. K. (2014). Perspective and prospective of pretreatment of corn straw for butanol production. *Applied Biochemistry and Biotechnology*, 172(2), 840–853.
- Bartosiak-Jentys, J., Eley, K., & 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(16), 5945–5947.
- Bartosiak-Jentys, J., Hussein, A. H., Lewis, C. J., & Leak, D. J. (2013). Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology (United Kingdom)*, 159(PART7), 1267–1275.
- Basavaraj, G., Rao, P. P., Reddy, C. R., Ashok Kumar, A., Srinivasa Rao, P., & Reddy, B. V. S. (2012). A Review of the National Biofuel Policy in India: A critique of the Need to Promote Alternative Feedstocks. Working Paper Series No. 34, (34), 1–20.
- Berdichevets, I. N., Shimshilashvili, H. R., Gerasymenko, I. M., Sindarovska, Y. R., Sheludko, Y. V., & Goldenkova-Pavlova, I. V. (2010). Multiplex PCR assay for detection of recombinant genes encoding fatty acid desaturases fused with lichenase reporter protein in GM plants. *Analytical and Bioanalytical Chemistry*, 397(6), 2289–2293.
- Berezina, O. V., Zakharova, N. V., Brandt, A., Yarotsky, S. V., Schwarz, W. H. & Zverlov, V. V. (2010). Reconstructing the clostridial n-butanol metabolic pathway in *Lactobacillus brevis*. *Appl. Microbiol. Biotechnol.* 87, 635–646.
- Bergler, H., Fuchsbichler, S., Hogenauer, G., & Turnowsky, F. (1996). The Enoyl-[Acyl-Carrier-Protein] Reductase (FabI) of *Escherichia coli*, which Catalyzes a Key Regulatory Step in Fatty Acid Biosynthesis, Accepts NADH and NADPH as Cofactors and is Inhibited by Palmitoyl-CoA. *European Journal of Biochemistry*, 242(3), 689–694.
- Bhandiwad, A., Guseva, A., & Lynd, L. (2013). Metabolic Engineering of *Thermoanaerobacterium thermosaccharolyticum* for increased n-Butanol Production. *Advances in Microbiology*, 3(March), 46–51.
- Bhandiwad, A., Shaw, A. J., Guss, A., Guseva, A., Bahl, H., & Lynd, L. R. (2014). Metabolic engineering of *Thermoanaerobacterium saccharolyticum* for n-butanol production. *Metabolic Engineering*, 21, 17–25.
- Bisswanger, H. (2014). Enzyme assays. *Perspectives in Science*, 1(1–6), 41–55.
- Bond-Watts, B. B., Weeks, A. M., & Chang, M. C. Y. (2012). Biochemical and structural characterization of the trans-enoyl-coa reductase from *Treponema denticola*. *Biochemistry*, 51(34), 6827–6837.

- Bond-Watts, B. B., Bellerose, R. J., & Chang, M. C. Y. (2011). Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nature Chemical Biology*, 7(4), 222–227.
- Branduardi, P., Longo, V., Berterame, N. M., Rossi, G., & Porro, D. (2013). A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. *Biotechnology for Biofuels*, 6(1), 68.
- Brumm, P., Land, M. L., Hauser, L. J., Jeffries, C. D., Chang, Y.-J., & Mead, D. A. (2015). Complete genome sequences of *Geobacillus* sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, 10, 81.
- Bui, L. M., Lee, J. Y., Geraldi, A., Rahman, Z., Lee, J. H., & Kim, S. C. (2015). Improved n-butanol tolerance in *Escherichia coli* by controlling membrane related functions. *Journal of Biotechnology*, 204, 33–44.
- Cairns, L. S., Hogley, L., & Stanley-Wall, N. R. (2014). Biofilm formation by *Bacillus subtilis*: New insights into regulatory strategies and assembly mechanisms. *Molecular Microbiology*, 93(4), 587–598.
- Cava, F., Pedro, M. A. De, Blas-galindo, E., Waldo, G. S., Westblade, L. F., & Berenguer, J. (2008). Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environmental Microbiology*.
- Chen, C.-T., & Liao, J. C. (2016). Frontiers in microbial 1-butanol and isobutanol production. *FEMS Microbiology Letters Minireview – Biotechnology & Synthetic Biology*, 363.
- Chen, J., Zhang, Z., Zhang, C., & Yu, B. (2015). Genome sequence of *Geobacillus thermoglucosidasius* DSM2542, a platform hosts for biotechnological applications with industrial potential. *Journal of Biotechnology*, 216, 98–99.
- Cheong, S., Clomburg, J. M., & Gonzalez, R. (2016). Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nature Biotechnology*, (April), 1–8.
- Cobb, R. E., Sun, N., & Zhao, H. (2013). Directed evolution as a powerful synthetic biology tool. *Methods*, 60(1), 81–90.
- Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Atkinson, T. (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, 11(6), 398–408.

- Currin, A., Swainston, N., Day, P. J., & Kell, D. B. (2015). Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews*, 44(5), 1172–239.
- Daniell, J., Köpke, M., & Simpson, S. D. (2012). Commercial biomass syngas fermentation. *Energies* (Vol. 5).
- Darmani, A., Arvidsson, N., Hidalgo, A., & Albors, J. (2014). What drives the development of renewable energy technologies? Toward a typology for the systemic drivers. *Renewable and Sustainable Energy Reviews*, 38, 834–847.
- Davis, M. S., Solbiati, J., & Cronan, J. E. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *The Journal of Biological Chemistry*, 275(37), 28593–8.
- Dawson, L. F., Valiente, E., Faulds-Pain, A., Donahue, E. H., & Wren, B. W. (2012). Characterisation of *Clostridium difficile* Biofilm Formation, a Role for Spo0A. *PLoS ONE*, 7(12), e50527.
- Dellomonaco, C., Clomburg, J. M., Miller, E. N., & Gonzalez, R. (2011). Engineered reversal of the β -oxidation cycle for the synthesis of fuels and chemicals. *Nature*, 476(7360), 355–359.
- Don Paul, C., Traore, D. A. K., Byres, E., Rossjohn, J., Devenish, R. J., Kiss, C., ... 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(Pt 10), 1266–8.
- Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Mukhopadhyay, A. (2014). Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology*.
- Dwidar, M., Kim, S., Jeon, B. S., Um, Y., Mitchell, R. J., Sang, B.-I., Børresen, B. (2013). Co-culturing a novel *Bacillus* strain with *Clostridium tyrobutyricum* ATCC 25755 to produce butyric acid from sucrose. *Biotechnology for Biofuels*, 6(1), 35.
- Dworkin, J., & Losick, R. (2001). Differential Gene Expression Governed by Chromosomal Spatial Asymmetry. *Cell*, 107(3), 339–346.
- Favrot, L., Blanchard, J. S., & Vergnolle, O. (2016). Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry*, 55(7), 989–1002.

- Ferreira, L. S., & Trierweiler, J. O. (2009). Modeling and simulation of the polymeric nanocapsule formation process. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 7(PART 1), 405–410.
- Fujimori, S., Su, X., Liu, J., Hasegawa, T., Takahashi, K., Masui, T., & Takimi, M. (2016). Implication of Paris Agreement in the context of long-term climate mitigation goals Background. *SpringerPlus*, 5.
- Gadde, B., Menke, C., & Wassmann, R. (2009). Rice straw as a renewable energy source in India, Thailand, and the Philippines: Overall potential and limitations for energy contribution and greenhouse gas mitigation. *Biomass and Bioenergy*, 33(11), 1532–1546.
- García, V., Pääkilä, J., Ojamo, H., Muurinen, E., & Keiski, R. L. (2011). Challenges in biobutanol production: How to improve the efficiency? *Renewable and Sustainable Energy Reviews*, 15(2), 964–980.
- Gunatilake, H., Roland-Holst, D., & Sugiyarto, G. (2014). Energy security for India: Biofuels, energy efficiency and food productivity. *Energy Policy*, 65, 761–767.
- Hamon, M. A., & Lazazzera, B. A. (2001). The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*. *Molecular Microbiology*, 42(5), 1199–1209.
- Hawkins A. S., Han Y., Lian H., Loder A. J., Menon A. L., Iwuchukwu I. J., Keller M., Leuko T. T., Adams M. W. W. & Kelly R. M. (2011). Extremely Thermophilic Routes to Microbial Electrofuels. *ACS Catal* 1:1043-1050.
- Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, 78(1), 79–85.
- Hoffmeister, M., Piotrowski, M., Nowitzki, U., & Martin, W. (2005). Mitochondrial trans-2-enoyl-CoA reductase of wax ester fermentation from *Euglena gracilis* defines a new family of enzymes involved in lipid synthesis. *Journal of Biological Chemistry*, 280(6), 4329–4338.
- Hoshino, T., Ikeda, T., Tomizuka, N., & Furukawa, K. (1985). Nucleotide sequence of the tetracycline resistance gene of pTHT15, a thermophilic *Bacillus* plasmid: comparison with staphylococcal TcR controls. *Gene*, 37(1–3), 131–8.
- Huang, H., Liu, H., & Gan, Y.-R. (2010). Genetic modification of critical enzymes and involved genes in butanol biosynthesis from biomass. *Biotechnology Advances*, 28, 651–657.

- Hussein, A. H., Lisowska, B. K., & Leak, D. J. (2015). The Genus *Geobacillus* and Their Biotechnological Potential. In *Advances in applied microbiology* (Vol. 92, pp. 1–48).
- Jang, Y.S., Lee, J.Y., Lee, J., Park, J.H., Im, J.A., Eom, M.H., Lee, J., Lee, S.H., Song, H., Cho, J.H. & Lee, S.Y. (2012). Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *MBio* 3, e00314–12.
- Jawed, K., Mattam, A. J., Fatma, Z., Wajid, S., Abdin, M. Z., & Yazdani, S. S. (2016). Engineered production of short chain fatty acid in *Escherichia coli* using fatty acid synthesis pathway. *PLoS ONE*, 11(7).
- Jeuland, N., Montagne, X., & Gautrot, X. (2004). Potentiality of Ethanol as a Fuel for Dedicated Engine. *Oil & Gas Science and Technology*, 59(6), 559–570.
- Jilani, S. B., Venigalla, S. S. K., Mattam, A. J., Dev, C., & Yazdani, S. S. (2017). Improvement in ethanol productivity of engineered *E. coli* strain SSY13 in defined medium via adaptive evolution. *Journal of Industrial Microbiology & Biotechnology*, 44(9), 1375–1384.
- Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandea-Nelson, M. D., & Reilly, P. J. (2011). Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochemistry*, 12(1), 44.
- Kananavičiūtė, R., Butaitė, E., & Čitavičius, D. (2014). Characterization of two novel plasmids from *Geobacillus* sp. 610 and 1121 strains. *Plasmid*, 71, 23–31.
- Kananavičiūtė, R., & Čitavičius, D. (2015). Genetic engineering of *Geobacillus* spp. *Journal of Microbiological Methods*, 111, 31–39.
- Kataoka, N., Tajima, T., Kato, J., Rachadech, W., & Vangnai, A. S. (2011). Development of butanol-tolerant *Bacillus subtilis* strain GRSW2-B1 as a potential bioproduction host. *AMB Express*, 1(1), 10.
- Kataoka, N., Vangnai, A. S., Pongtharangkul, T., Tajima, T., Yakushi, T., Matsushita, K., & Kato, J. (2015). Construction of CoA-dependent 1-butanol synthetic pathway functions under aerobic conditions in *Escherichia coli*. *Journal of Biotechnology*, 204, 25–32.
- Kataoka, N., Vangnai, A. S., Pongtharangkul, T., Yakushi, T., & Matsushita, K. (2017). Butyrate production under aerobic growth conditions by engineered *Escherichia coli*. *Journal of Bioscience and Bioengineering*, 123(5), 562–568.
- Kearns, D. B. (2010). A field guide to bacterial swarming motility. *Nature Reviews. Microbiology*, 8(9), 634–44.

- Kehrenberg, C., Catry, B., Haesebrouck, F., de Kruif, A., & Schwarz, S. (2005). Novel spectinomycin/streptomycin resistance gene, *aadA14*, from *Pasteurella multocida*. *Antimicrobial Agents and Chemotherapy*, 49(7), 3046–9.
- Kim, J., & Kim, K.-J. (2014). Cloning, expression, purification, crystallization and X-ray crystallographic analysis of the (S)-3-hydroxybutyryl-CoA dehydrogenase PaaH1 from *Ralstonia eutropha* H16. *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 7), 955–8.
- Klümper, U., Riber, L., Dechesne, A., Sannazzarro, A., Hansen, L. H., Sørensen, S. J., & Smets, B. F. (2014). Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *The ISME Journal*, 9(4), 934–945.
- Kobayashi, K. (2007). *Bacillus subtilis* pellicle formation proceeds through genetically defined morphological changes. *Journal of Bacteriology*, 189(13), 4920–31.
- Kuisiense, N., Raugalas, J., & Chitavichius, D. (2009). Phylogenetic, inter, and intraspecific sequence analysis of *spo0A* gene of the genus *Geobacillus*. *Current Microbiology*, 58(6), 547–553.
- Kumar, V., Hart, A. J., Keerthiraju, E. R., Waldron, P. R., Tucker, G. A., & Greetham, D. (2015). Expression of mitochondrial cytochrome c oxidase chaperone gene (*COX20*) improves tolerance to weak acid and oxidative stress during yeast fermentation. *PLoS ONE*, 10(10), 1–16.
- Lai, M. C., & Lan, E. I. (2015). Advances in Metabolic Engineering of Cyanobacteria for Photosynthetic Biochemical Production. *Metabolites*, 5(4), 636–58.
- Lan, E. I., & Liao, J. C. (2012). ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16), 6018–23.
- Lan, E. I., Ro, S. Y., & Liao, J. C. (2013). Oxygen-tolerant coenzyme A-acylating aldehyde dehydrogenase facilitates efficient photosynthetic n-butanol biosynthesis in cyanobacteria. *Energy & Environmental Science*, 6(9), 2672.
- Lee, C. L., Siak, D., Ow, W., Kah, S., & Oh, W. (2006). Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *Journal of Microbiological Methods*.
- Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., & Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology*, 8(6), 536–546.

- Lehnherr, H., Maguin, E., Jafri, S., & Yarmolinsky, M. B. (1993). Plasmid Addiction Genes of Bacteriophage P1: *doc*, which Causes Cell Death on Curing of Prophage, and *phd*, which Prevents Host Death when Prophage is Retained. *Journal of Molecular Biology*, 233(3), 414–428.
- Lennen, R. M., & Pfleger, B. F. (2012). Engineering *Escherichia coli* to synthesize free fatty acids. *Trends in Biotechnology*, 30(12), 659–667.
- Li, H., Opgenorth, P. H., Wernick, D. G., Rogers, S., Wu, T. Y., Higashide, W., Malati, P., Huo, Y. X., Cho, K. M. & Liao, J.C. (2012). Integrated electromicrobial conversion of CO₂ to higher alcohols. *Science* 335, 1596–1596.
- Liao, J. C., Mi, L., Pontrelli, S., & Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Publishing Group*, 14.
- Lim, J. H., Seo, S. W., Kim, S. Y., & Jung, G. Y. (2013). Model-driven rebalancing of the intracellular redox state for optimization of a heterologous n-butanol pathway in *Escherichia coli*. *Metabolic Engineering*, 20, 56–62.
- Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., Liao, J. C. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metabolic Engineering*, 31, 44–52.
- Lin, P. P., Rabe, K. S., Takasumi, J. L., Kadisch, M., Arnold, F. H., & Liao, J. C. (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabolic Engineering*, 24, 1–8.
- Liu, M., Zhang, Y., Inouye, M., & Woychik, N. A. (2008). Bacterial addiction module toxin *Doc* inhibits translation elongation through its association with the 30S ribosomal subunit. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15), 5885–90.
- Loder, A. J., Zeldes, B. M., Dale Garrison, G., Lipscomb, G. L., Adams, M. W. W., & Kelly, R. M. (2015). Alcohol selectivity in a synthetic thermophilic n-butanol pathway is driven by biocatalytic and thermostability characteristics of constituent enzymes. *Applied and Environmental Microbiology*, 81(20), 7187–7200.
- Loftie-Eaton, W., Taylor, M., Horne, K., Tuffin, M. I., Burton, S. G., & Cowan, D. A. (2013). Balancing redox cofactor generation and ATP synthesis: Key microaerobic responses in thermophilic fermentations. *Biotechnology and Bioengineering*, 110(4), 1057–1065.
- Junca, H., Plumeier, I., Rgen Hecht, H.-J., & Pieper, D. H. (2004). Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiology*, 150, 4181–4187.

- Mann, M. S., & Lütke-Eversloh, T. (2013). Thiolase engineering for enhanced butanol production in *Clostridium acetobutylicum*. *Biotechnology and Bioengineering*, 110(3), 887–897.
- Mattam, A. J., & Yazdani, S. S. (2013). Engineering *E. coli* strain for conversion of short chain fatty acids to bioalcohols. *Biotechnology for Biofuels*, 6(1), 128.
- Mustard, A., & Aradhey, A. (2012). India Biofuels Annual 2012. USDA Foreign Agriculture Service, 15, 1–16.
- Nagórska, K., Ostrowski, A., Hinc, K., Holland, I. B., & Obuchowski, M. (2010). Importance of *eps* genes from *Bacillus subtilis* in biofilm formation and swarming. *Journal of Applied Genetics*, 51(3), 369–381.
- Nanda, S., Golemi-Kotra, D., Mcdermott, J. C., Dalai, A. K., Gökalp, I., & Kozinski, J. A. (2017). Fermentative production of butanol: Perspectives on synthetic biology. *New Biotechnology*.
- Ndaba, B., Chiyanzu, I., & Marx, S. (2015). N-Butanol derived from biochemical and chemical routes: A review. *Biotechnology Reports*, 8, 1–9.
- Ng, C. Y. C., Takahashi, K., & Liu, Z. (2016). Isolation, characterization, and optimization of an aerobic butanol-producing bacterium from Singapore. *Biotechnology and Applied Biochemistry*, 63(1), 86–91.
- Nielsen, D. R., Leonard, E., Yoon, S. H., Tseng, H. C., Yuan, C. & Prather, K. L. J. (2009). Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* 11, 262–273.
- Noguchi, S., Putri, S. P., Lan, E. I., Laviña, W. A., Dempo, Y., Bamba, T., ... Fukusaki, E. (2016). Quantitative target analysis and kinetic profiling of acyl-CoAs reveal the rate-limiting step in cyanobacterial 1-butanol production. *Metabolomics*, 12(2), 26.
- Okamura, E., Tomita, T., Sawa, R., Nishiyama, M., & Kuzuyama, T. (2010). Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), 11265–70.
- Ou, J., Ma, C., Xu, N., Du, Y., & Liu, X. (Margaret). (2015). High butanol production by regulating carbon, redox and energy in *Clostridia*. *Frontiers of Chemical Science and Engineering*, 9(3), 317–323.
- Park, S. H., Kim, S., & Hahn, J. S. (2014). Metabolic engineering of *Saccharomyces cerevisiae* for the production of isobutanol and 3-methyl-1-butanol. *Applied Microbiology and Biotechnology*, 98(21), 9139–9147.

- Parte, A. C. (2018). LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology*, 68; doi: 10.1099/ijsem.0.002786.
- Pásztor, A. (2015). *Advanced Biofuel Production: Engineering Metabolic Pathways for Butanol and Propane Biosynthesis*.
- Pásztor, A., Kallio, P., Malatinszky, D., Akhtar, M. K., & Jones, P. R. (2015). A synthetic O₂-tolerant butanol pathway exploiting native fatty acid biosynthesis in *Escherichia coli*. *Biotechnology and Bioengineering*, 112(1), 120–128.
- Peralta-Yahya, P. P., & Keasling, J. D. (2010). Advanced biofuel production in microbes. *Biotechnology Journal*, 5(2), 147–162.
- Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., & Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature*, 488(7411), 320–328.
- Pfaffl, M. W., Hamley-Bennett, C., Danson, M. J., Leak, D. J., Aldrich, H., Yeates, T., Yang, S. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9), 45e–45.
- Pfleger, B. F., Pitera, D. J., Smolke, C. D., & Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*, 24(8), 1027–1032.
- Pick, A., Rühmann, B., Schmid, J., & Sieber, V. (2013). Novel CAD-like enzymes from *Escherichia coli* K-12 as additional tools in chemical production. *Applied Microbiology and Biotechnology*, 97(13), 5815–5824.
- Ranganathan, S., & Maranas, C. D. (2010). Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions. *Biotechnology Journal*, 5(7), 716–725.
- Ravagnani, A., Jennert, K. C. B., Steiner, E., Grünberg, R., Jefferies, J. R., Wilkinson, S. R., ... Young, M. (2000). Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular Microbiology*, 37(5), 1172–1185.
- Reeve, B., Hargest, T., Gilbert, C., & Ellis, T. (2014). Predicting translation initiation rates for designing synthetic biology. *Frontiers in Bioengineering and Biotechnology*, 2, 1.
- Reeve, B., Martinez-Klimova, E., De Jonghe, J., Leak, D. J., & Ellis, T. (2016). The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology*.
- Reyes, L. H., Abdelaal, A. S., & Kao, K. C. (2013). Genetic determinants for n-butanol tolerance in evolved *Escherichia coli* mutants: Cross adaptation and

antagonistic pleiotropy between n-butanol and other stressors. *Applied and Environmental Microbiology*, 79(17), 5313–5320.

Rühl, J., Schmid, A., & Blank, L. M. (2009). Selected *Pseudomonas putida* strains able to grow in the presence of high butanol concentrations. *Applied and Environmental Microbiology*, 75(13), 4653–6.

Salvo, M. L. di, Budisa, N., & Contestabile, R. (2013). PLP-dependent Enzymes: a Powerful Tool for Metabolic Synthesis of Non-canonical Amino Acids. *Beilstein Bozen Symposium on Molecular Engineering and Control*.

Sarnklong, C. et al. (2010). Utilization of Rice Straw and Different Treatments to Improve Its Feed Value for Ruminants: A Review. *J. Anim. Sci*, 23(5), 680–692.

Schadeweg, V., & Boles, E. (2016). Increasing n-butanol production with *Saccharomyces cerevisiae* by optimizing acetyl-CoA synthesis, NADH levels and trans-2-enoyl-CoA reductase expression. *Biotechnology for Biofuels*, 9, 257.

Schmid, J., Sieber, V., & Rehm, B. (2015). Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Frontiers in Microbiology*, 6, 496.

Schwarz, K. M., Grosse-Honebrink, A., Derecka, K., Rotta, C., Zhang, Y., & Minton, N. P. (2017). Towards improved butanol production through targeted genetic modification of *Clostridium pasteurianum*. *Metabolic Engineering*.

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., & Liao, J. C. (2011). Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and Environmental Microbiology*, 77(9), 2905–2915.

Sheng, L., Kovács, K., Winzer, K., Zhang, Y., & 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(1), 5.

Sheng, L., Zhang, Y., & Minton, N. P. (2016). Complete Genome Sequence of *Geobacillus thermoglucosidasius* NCIMB 11955, the Progenitor of a Bioethanol Production Strain. *Genome Announcements*, 4(5).

Silvaggi, J. M., Popham, D. L., Driks, A., Eichenberger, P., & Losick, R. (2004). Unmasking novel sporulation genes in *Bacillus subtilis*. *Journal of Bacteriology*, 186(23), 8089–95.

Studholme, D. J. (2015). Some (bacilli) like it hot: genomics of *Geobacillus* species. *Microbial Biotechnology*, 8(1), 40–8.

Suzuki, H., Yoshida, K. Ichi, & Ohshima, T. (2013). Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus*

- kaustophilus HTA426. *Applied and Environmental Microbiology*, 79(17), 5151–5158.
- Tanaka, T., Shima, Y., Ogawa, N., Nagayama, K., Yoshida, T., & Ohmachi, T. (2010). Expression, identification and purification of Dictyostelium acetoacetyl-coa thiolase expressed in Escherichia coli. *International Journal of Biological Sciences*, 7(1), 9–17.
- Tang, Q., Yin, K., Qian, H., Zhao, Y., Wang, W., Chou, S.-H., He, J. (2016). Cyclic di-GMP contributes to adaptation and virulence of Bacillus thuringiensis through a riboswitch-regulated collagen adhesion protein. Nature Publishing Group.
- Tang, Y. J., Sapra, R., Joyner, D., Hazen, T. C., Myers, S., Reichmuth, D., Keasling, J. D. (2009). Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant Geobacillus strain. *Biotechnology and Bioengineering*, 102(5), 1377–1386.
- Taylor, M. P., Esteban, C. D., & Leak, D. J. (2008). Development of a versatile shuttle vector for gene expression in Geobacillus spp. *Plasmid*, 60(1), 45–52.
- Tominaga, Y., Ohshiro, T., & Suzuki, H. (2016). Conjugative plasmid transfer from Escherichia coli is a versatile approach for genetic transformation of thermophilic Bacillus and Geobacillus species. *Extremophiles*, 20(3), 375–381.
- Tran, T., Terwilliger, T. C., Waldo, G. S., & Pe, J. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1), 79–89.
- Tucci, S., & Martin, W. (2007). A novel prokaryotic trans-2-enoyl-CoA reductase from the spirochete Treponema denticola. *FEBS Letters*, 581(8), 1561–1566.
- Velappan, N., Sblattero, D., Chasteen, L., Pavlik, P., & Bradbury, A. R. M. (2007). Plasmid incompatibility: More compatible than previously thought? *Protein Engineering, Design and Selection*, 20(7), 309–313.
- Venkatasubramanian, P., Daniels, L., & Rosazza, J. P. N. (2007). Reduction of Carboxylic Acids by Nocardia Aldehyde Oxidoreductase Requires a Phosphopantetheinylated Enzyme. *Journal of Biological Chemistry*, 282(1), 478–485.
- Vick, J. E., Clomburg, J. M., Blankschien, M. D., Chou, A., Kim, S., & Gonzalez, R. (2015). Escherichia coli enoyl-acyl carrier protein reductase (FabI) supports efficient operation of a functional reversal of β -oxidation cycle. *Applied and Environmental Microbiology*, 81(4), 1406–16.
- Volker, M. and L.-M. (2001). Bacterial fermentation. *Encyclopedia of Life Science*, 1–7.

- Volodina, E., & Steinbüchel, A. (2014). (S)-3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase (FadB') from fatty acid degradation operon of *Ralstonia eutropha* H16. *AMB Express*, 4, 69.
- Wang, G., Wu, P., Liu, Y., Mi, S., Mai, S., Gu, C., Kotlar, H. K. (2015). Isolation and characterisation of non-anaerobic butanol-producing symbiotic system TSH06. *Bioenergy and Biofuels*.
- Wiesenthal, T., Leduc, G., Christidis, P., Schade, B., Pelkmans, L., Govaerts, L., & Georgopoulos, P. (2009). Biofuel support policies in Europe: Lessons learnt for the long way ahead. *Renewable and Sustainable Energy Reviews*, 13(4), 789–800.
- Wu, P., Wang, G., Wang, G., Børresen, B. T., Liu, H., & Zhang, J. (2016). Butanol production under microaerobic conditions with a symbiotic system of *Clostridium acetobutylicum* and *Bacillus cereus*. *Microbial Cell Factories*, 15(1), 8.
- Wu, Y.-C., & Liu, S.-T. (2010). A sequence that affects the copy number and stability of pSW200 and ColE1. *Journal of Bacteriology*, 192(14), 3654–60.
- Xiao, Z., Wang, X., Huang, Y., Huo, F., Zhu, X., Xi, L., & Lu, J. R. (2012). Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. *Biotechnology for Biofuels*, 5(1), 88.
- Xu, Y., & Ramanathan, V. (2017). Well below 2 °C: Mitigation strategies for avoiding dangerous to catastrophic climate changes. *Proceedings of the National Academy of Sciences of the United States of America*, 114(39), 10315–10323.
- Ye, W., Li, J., Han, R., Xu, G., Dong, J., & Ni, Y. (2017). Engineering coenzyme A-dependent pathway from *Clostridium saccharobutylicum* in *Escherichia coli* for butanol production. *Bioresource Technology*, 235, 140–148.
- Yoo, M., Croux, C., Meynial-Salles, I., & Soucaille, P. (2016). Elucidation of the roles of adhE1 and adhE2 in the primary metabolism of *Clostridium acetobutylicum* by combining in-frame gene deletion and a quantitative system-scale approach. *Biotechnology for Biofuels*, 9(1), 92.
- Yu, A.-Q., Pratomo Juwono, N. K., Leong, S. S. J., & Chang, M. W. (2014). Production of Fatty Acid-derived valuable chemicals in synthetic microbes. *Frontiers in Bioengineering and Biotechnology*, 2(December), 78.
- Yu, S., Zheng, B., Zhao, X., & Feng, Y. (2010). Gene cloning and characterization of a novel thermophilic esterase from *Fervidobacterium nodosum* Rt17-B1. *Acta Biochimica et Biophysica Sinica*, 42(4), 288–295.

- Zaki, A. M., Wimalasena, T. T., & Greetham, D. (2014). Phenotypic characterisation of *Saccharomyces* spp. for tolerance to 1-butanol. *Journal of Industrial Microbiology and Biotechnology*, 41(11), 1627–1636.
- Zeigler, D. R. (2014). The *Geobacillus* paradox: Why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology (United Kingdom)*, 160(PART 1), 1–11.
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Milo, R. (2013). Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Research*, 41(9), e98–e98.
- Zhang, F., Rodriguez, S., & Keasling, J. D. (2011). Metabolic engineering of microbial pathways for advanced biofuels production. *Current Opinion in Biotechnology*, 22(6), 775–783.
- Zhang, Y., Grosse-Honebrink, A., & Minton, N. P. (2015). A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLoS ONE*, 10(4), 1–21.
- Zhang, J., Yu, L., Lin, M., Yan, Q., & Yang, S.-T. (2017). n-Butanol production from sucrose and sugarcane juice by engineered *Clostridium tyrobutyricum* overexpressing sucrose catabolism genes and *adhE2*. *Bioresource Technology*, 233, 51–57.
- Zheng, Y. N., Li, L. Z., Xian, M., Ma, Y. J., Yang, J. M., Xu, X., & He, D. Z. (2009). Problems with the microbial production of butanol. *Journal of Industrial Microbiology and Biotechnology*, 36(9), 1127–1138.
- Zheng, J., Tashiro, Y., Wang, Q., & Sonomoto, K. (2015). Recent advances to improve fermentative butanol production: Genetic engineering and fermentation technology. *Journal of Bioscience and Bioengineering*, 119(1), 1–9.
- Zhou, J., Wu, K., & Rao, C. V. (2016). Evolutionary engineering of *Geobacillus thermoglucosidasius* for improved ethanol production. *Biotechnology and Bioengineering*, 113(10), 2156–2167.
- Zhu, H., Ren, X., Wang, J., Song, Z., Shi, M., Qiao, J., Zhang, W. (2013). Integrated OMICS guided engineering of biofuel butanol-tolerance in photosynthetic *Synechocystis* sp. PCC 6803. *Biotechnology for Biofuels*, 6(1), 106.

BIBLIOGRAPHY

APPENDIX

A1. Primers used in this study

OLIGO NAME	5'<-----SEQUENCE----->3'	LENGTH
Promoter gene amplification and cloning		
Pfdx_F	CTGACTGCAGgcgggccgctgtagtag	27
Pfdx_R	TCAGTCTAGACATcctccttaaaaattacacaactttatacga	43
PthI_F	CTGACTGCAGgcgggccgcttttaacaaaatatattg	37
PthI_R	TCAGTCTAGACATcctcctaaatTTgatacggggtaa	38
PfacOID_F	CTGACTGCAGaattcgcggccgcactagt	29
PfacOID_R	TCAGTCTAGACATcctcctaactagcttgacac	34
eCGP123 reporter gene amplification and cloning		
+eCGP123_F	catgattacgaattcgagctCCTTGCACCATGGATCCG	38
+eCGP123_R	ggatccccgggtaccgagctTTAGTGATGATGGTGGTGGTG	41
Sequencing primers for pMTL61112		
M13R	CAGGAAACAGCTATGACC	18
M13F	TGTA AACGACGGCCAGT	18
RepB_R1	CAGCAACTAAAATAAAAATGACGTTATTTTC	30
RepB_F1	GTATTGAAAACCTTAAAATTGGTTGCA	28
RepB_R2	CAGAAGTTCAAAGTAATCAACATTAGC	27
RepB_F2	GCTCAAGGATTTGCGCCGAAT	20
RepB_F3	CGGGCCAGTTTGTGAAGAT	20
RepB_F4	GCT GAA AGA GTA AAA GAT TGT GCT G	34
RepB_R4	CTTCATCATCGGTCATAAAATCCGT	25
Kan_F1	AGGAAGCAGAGTTCAGCCAT	20
ColE1_F1	GCTTGCAAACAAAAAACCCACCG	23
ColE1_F2	CCGCCTTTGAGTGAGCTGATA	21
ColE1_F4	ACC AAC TCT TTT TCC GAA GGT AAC	32
ColE1_R4	ATACGGTTATCCACAGAATCAGG	23
Ldh_F	gtctgtcatgaaatggacaacaata	26
Ldh_R	catataaatattttacgcgaaaggcgca	28
Spectinomycin resistance gene amplification and cloning		
aad9 F	TAGTGAGGAGGATATATTTGAATACATACGAACAAAT	37
aad9 R	ATTTTTTCAATTTTTTTATAATTTTTTTAAT	31
Toxin/antitoxin system amplification and cloning		
Tox_F	ggccggccATATTATGGAAAGGAGACGGAAAAGAA	35
Tox_R	gtttaaacTCACTCTTTTTGTTTTGGATGGG	32
qPCR primers for replicon copy number		
RepB_F	AACAAAATCGTGAAACAGGCG	21
RepB_R	CCTTTTGTGACTGAATGCCATG	22
Rep01_F	GGAACGATGGAACAAAAGGTG	21
Rep01_R	GGTGTATGTTTAAAAGTCCTTGTC	24
Rep02_R	GCGAGCTTTTGC GACTTTAG	20
Rep02_F	GGATTTTCGACATTGAACAGGC	21
Rep372_F	AACTACCTTCGCTCACATCG	20

Rep372_R	GTCAGCAAAGCAGGATCAG	20
RepC56_F	GTTCTCCATATCTCCACGGTAAC	23
RepC56_R	ACATTTAGGAGCGGACACAG	20
Geobacillus Chromosome _lepA1_F	CCCCTCTTTATTCTAGCGACG	21
Geobacillus Chromosome _lepA1_R	GTGGAAAACCTTGATAACGCCG	21
NEBuilder HiFi primers for construction of pMTL61112 +<i>thl,hbd,crt,adhE</i>		
FP1	CGCTCCATCAAGAAGAGCGA	20
RP1	GTTGAACACGTCCGTCAAGC	20
FP2	TGGTCTACGATGGCTTGACG	20
RP2	GAACACTTGAGCAATGCCCG	20
FP3	AATCTGTGTGATTGGCGCTG	20
RP3	CGGATGTGGGCGATATTTGC	20
FP4	GCAAATATCGCCCACATCCG	20
RP4	CCGTCGTAAAGCACAACGAG	20
FP5	GAAACGGACTTGATAACGGGA	21
RP5	TCGTTCTTTTTAGTTCTAAATCGT	25
NEBuilder HiFi primers for <i>Geobacillus</i> native gene constructs		
ldh_F	GAATTCGCGGCCGACTAGTGC	22
thl_F	cagctagtttatattgaaggaggcagagctcATGGGAAAAACG GTGATTTTAAG	54
thl_R	ccttcaatataaggtaccTTACGAATCAACCTGCACCA	38
hbd_F	ggttgattcgtaaggtacctatattgaaggaggATGATGGACGT GAAAACAATC	55
hbd_R	ctcctcaatataatctagaTACTCATACGTGTA AAAACCTC	43
crt_F	cacgtatgagtaattagattatattgaaggaggATGACTGAGTT TGCGCACAT	54
crt_R	caaatgcaggcttctattttttatggctagcttaattaacctaggTTAT CTTCCTTGAAAACGCG	65
Adhe_F	agtttatattgaaggaggcaATGGCTGTGGAGGAAAGAGTC	41
Adhe_R	aatataaactaaggcctTAAACTCCTTTAAACGCTTGG	39
Ter6192_F	aggagttaaaggcctTAGTTTATATTGAAGGAGGATGAAT GCAATG	47
Ter6192_R	gcaggcttctatttttatggctagcTCATCCTTCGCGAGGATC GA	46
Nat.AdhE_NV 3_F	ttaggatattaaGCGGGACGGGGAGCTGAG	31
Nat.AdhE_NV 3_R	tattttatggctagcttaattaacTAAACTCCTTTAAACGCTT GGCGATAAATATG	58
NV2_F_forNV 3	aagccgcttttcaaggaagataacGCGGGACGGGGAGCTG AG	43

NV2_R_forNV3	tatttttatggctagcttaattaacTCATCCTTCGCGAGGATCG	44
FabI gene amplification and cloning		
NV1_R_FabI_Ind	TGAGCTCTGCCTCCTTCA	18
FabI_F_Ind	attgaaggaggcagagctcaATGAATTTATCATTAAAAGGGC	42
FabI_R_Ind	gctagcttaattaacctaggTTAATATCCTAAAATATGATATCAG	46
NV1_F_FabI_Ind	CCTAGGTTAATTAAGCTAGCC	21
ldh_FabI_F	aagccgcgttttcaaggaagataacGCGGGACGGGGAGCTGAG	43
ldh_FabI_R	ctccccgtccgcTTAATATCCTAAAATATGATATCCAGAA TCGACATGG	50
Thioesterase gene amplification and cloning		
TesBF_FP	tacgCTGCAGTTATATTGAAGGAGGATGATCTACATGG CTTATCAATATCG	51
TesBF_RP	catCCCGGGTTAGTGATGGTGATGGTGATGTTCCG	34
TesBT_FP	tacgCTGCAGTTATATTGAAGGAGGATGTCGGAAGAA AACAAAATCGGCA	50
TesBT_RP	catCCCGGGTTAGTGATGGTGATGGTGATGTTCCGAAA ATCACTTTG	46
Tes_BF_Nde1_F	TACGCATATGATGATCTACATGGCTTATCAATATCGT	37
Tes_BF_Nhe1_R	catgctagcTTAGTGATGGTGATGGTGATGTTCCG	34
Tes_BT_Nde1_F	TACGCATATGATGTCGGAAGAAAACAAAATCGGCA	35
Tes_BT_Nhe1_R	catgctagcTTAGTGATGGTGATGGTGATGTTCCGAAAAT CACTTTG	46
TesF_F	attgaaggaggcagagctcaATGATCTACATGGCTTATCAAT ATC	45
TesF_R	gctagcttaattaacctaggTTAGTGATGGTGATGGTGATG	41
NV1_F_forTefsF	CCTAGGTTAATTAAGCTAGCCATAAAAATA	30
NV1_R_forTefsF	TGAGCTCTGCCTCCTTCA	18
Amplification of <i>Clostridium acetobutylicum</i> n-butanol genes		
thiL_1	ATCCCGGGG-AGGAGTAAAACATGAGAGA	29
thiL_2	ATCCCGGGCTCGAG-TTAGTCTCTTCAACTACGA	35
crtbcdetfhbd_3	ATCCCGGG-ATATTTTAGGAGGAT TAGTC	30
crtbcdetfhbd_4	ATCCCGGGAGATCT-CCATATTATAATCCCTCCTC	35

adhe1_5	ATCCCGGG-ATATCTATCTCCAAATCTGC	29
adhe1_6	ATCCCGGGCTCGAG-CCAATCATAATTGTCATCCC	35
adhe_7	CTCTCCCGGG-TATAAGGCATCAAAGTGTGT	31
adhe_8	CTCTCCCGGGCTCGAG-GTCTATGTGCTTCATGAAGC	37
Sequencing of <i>G. thermoglucosidasius</i> integration site		
PyrE_C1_F	CGAACAATTCTGTGACCCG	20
PyrE_C2_R	CCCATGCTGAAAATCCAGCTG	21
Primers in Megaplasmid 01		
F1	Tctggagaagagtccctccatcgctgcttggcg	32
R1	ggattggctcgagtttcgcgcggaatggcaaA	33
F2	AAT GGG GAA TAA AGT CCG GCT CC	30
R2	TC GGA AGC TAA AGG ATG ATC CTT TGG	34
F3	CGA TAA AAG GTG GAC TTA AGT CAA TTC CCC	39
R3	TCA GTG AAC TAT CCC TAC CTA CGC TTC	35

A2. Reporter gene sequences

pheB

ATGGCTATTATGCGGATCGGCAAGGCCGAAATAAGAGTCATGGATCTCGAAGAATC
TGTCAGTATTACACGAATGTGATTGGCCTGGAGGAAGTGGGAAGGAGTGAAGGA
AGAGTTTATTTAAAGGCATGGGATGAATTCGATCACCATAGCCTCATTCTTCAAGAA
GCCGATTCCCCTGGCCTTGATCACATTGCTTTTAAAGGTGGAACATGAAGACGATTTA
GCCAAGTACGAGAAGAAGATCGAGCAATTCGGGTGTACGTTAAAACGGATTTCCAA
AGGGACAAGGCTTGCAGAAGGAGAAGCAATTCGCTTCGAGCTTCCAACAGGGCAT
CAAGTGGAATTGTACCATGAAATCGTGCGTGTAGGCACGAAGACAGGAAATTTGAA
TCCAGCCCCATGGCCGGATGGAATGCGCGGGATTGCACCCGACCCGCTAGATCACT
TAGCGCTGACAGGAGAAGATATCAACACAGTGACAAGATTTTTTACAGAAGCCTTG
GATATGTTTCATTAGCGAAAAAATTATGACAGTAGATGGGGAAGAGATGGTAGGGA
GCTTTATATTTGCCAGAAACGGAAAAGCGCACGATGTTGCCTTTATTAAGGGCCA
GATAAGAAAATGCATCATGTGCGATTCTATGTGGACAATTGGTATGAAGTGTTAAA
GGCAGCGGATATTTTATCCAAAATAATGTCCAATTCGATGTGACACCGACCCGCCA
TGGGATTACGCGTGGACAAACCACCTACTTCTTTGATCCTTCAGGTAATCGCAATGA
AGCTTTTGCAAGCGTTACATTACGTATCCTGATTTTCTACCATAACATGGACAGA
AGACAAAATCGGTCAAGGAATCTTCTATCATAGAAGAGAATTGACGGAGTCATTCA
TCAAGGCGCTGACATAA

sfGFP

ATGAGCAAAGGCGAGGAACTGTTTACCGGTGTCGTCCCGATCCTGGTTCGAGCTGGA
TGGCGACGTTAACGGCCACAAGTTTTCTGTGCGCGGTGAAGGTGAGGGTGATGCG
ACGAACGGTAAACTGACCCTGAAATTCATTTGCACCACTGGTAAGCTGCCGGTGCC
GTGGCCGACCCTGGTTACCACGTTGGGCTACGGTGTTCAGTGTTCAGCCGTTATCC
GGACCACATGAAACGTACGACTTCTTTAAGAGCGCGATGCCTGAAGGCTACGTGC
AAGAGCGCACGATCAGCTTTAAAGACGACGGTACGTATAAGACGCGTGCAGAGGT

CAAGTTCGAGGGCGACACCCTGGTTAATCGCATCGAGCTGAAGGGTATCGACTTCA
 AAGAAGATGGCAATATTCTGGGTCATAAGCTGGAATACAACTTCAATTCCCATAATG
 TGTACATTACCGCGGATAAACAGAAGAACGGCATTAAAGCTAACTTCAAGATTCGT
 CACAATGTTGAGGACGGTAGCGTTCAACTGGCGGATCACTATCAACAGAACACCCC
 AATCGGTGATGGTCCGGTGCTGCTGCCGGATAATCACTACCTGAGCACTCAGTCGG
 TGTTGAGCAAAGACCCGAACGAAAAACGTGATCACATGGTCCTGCTGGAGTTTGT
 ACCGCAGCCGGCATCACGCATGGTATGGACGAATTGTATAAGGGTAGCCATCACCA
 CCATCACCATTAA

eCGP123

ATGGGCGCACACGCGTCCGTCATTAAACCTGAGATGAAGATTAACCTGCGCATGGA
 AGGTGCGGTCAACGGTCATAAGTTCGTGATCGAGGGTGAAGGTATCGGTAAGCCG
 TATGAAGGCACCCAGACGTTGGACCTGACCGTGAAAGAGGGCGCGCCGCTGCCGT
 TTTCTTATGATATTCTGACTCCGGCATTCCAATATGGTAACCGTGCCTTTACCAAGTA
 CCCGAAAGATATCCCGGACTACTTTAAACAGGCGTTTCCGGAAGGCTACAGCTGGG
 AGCGCTCGATGACCTACGAAGATCAAGGTATTTGCATCGCGACCAGCGACATCACG
 ATGGAAGGTGACTGCTTTTTCTACAAAATTCGTTTCGATGGTACGAATTTCCACCA
 AATGGCCCGTTATGCAAAAAGAAAACGCTGAAGTGGGAGCCGAGCACCGAGAAAA
 TGTACGTTTCGTGATGGCGTTCTGAAGGGTGACGTGAACATGGCGCTGTTGCTGGAG
 GGTGGCGGCCACTACCGTTGTGATTTCAAACACCTATAAAGCAAAGAAAGACGT
 CCGTCTGCCGGATGCCACGAAGTTGACCACCGCATTGAGATTCTGAGCCATGACA
 AGGACTATAACAAAGTGCGCCTGTACGAACATGCTGAGGCACGTTATAGCATGTTG
 CCGAGCCAGGCTAAGCCGATCCCGAATCCGCTGCTGGGTCTGGATAGCACGCACCA
 CCACCATCATCACTAA

licB

ATGAAAAACAGGGTAATTTTCATTATTAATGGCTTCCTTGCTTTTGGTTTTGTCGGTAA
 TTGTTGCTCCTTTTTACAAAGCGGAAGCCGCAACTGTGGTAAATACGCCTTTTGTG
 CAGTGTTCGAACTTTGACTCCAGTCAGTGGGAAAAAGCGGATTGGGCGAACGGT
 TCGGTGTTCAACTGTGTTTGAAGCCTTCACAGGTGACATTTTCGAACGGTAAAATG
 ATTTTGACCCTTGACAGGGAATATGGCGGTTTCATATCCGTATAAAAGCGGTGAATAT
 CGTACAAAATCATTTTTCGGATACGGTTATTATGAAGTAAGAATGAAAGCTGCCAAA
 AACGTAGGAATTGTTTCATCTTTCTTCACTTATACAGGACCTTCGGACAACAATCCAT
 GGGACGAAATCGATATCGAGTTTTTAGGAAAGGACACAACCTAAAGTTCAGTTCAAC
 TGGTACAAAAATGGAGTCGGTGGAAACGAGTATTTGCACAATCTTGGATTTCGATGC
 TTCCAGGATTTTCATACTTATGGATTTGAATGGAGGCCGGATTATATAGACTTCTAT
 GTTGACGGCAAAAAAGTTTATCGTGGAACCAGGAACATACCTGTTACTCCCGGCAA
 AATTATGATGAATTTGTGGCCAGGAATAGGAGTGGATGAATGGTTGGGACGTTAC
 GACGGAAGAACTCCTTTGCAGGCGGAGTACGAATATGTAAAATACTATCCTAACGG
 TGTTCCGCAAGATAATCCTACTCCTACTCCTACGATTGCTCCTTCTACTCCGACTAAC
 CCTAATTTACCTCTTAAGGGAGACGTAAACGGCGACGGTCATGTAACTCATCAGAC
 TATTCATTATTTAAAAGATATTTGCTCAGGGTTATTGATAGATTCCCTGTTGGAGATC
 CACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAG

A3. Promoter sequences

1) P_{ldh} *G. stearothermophilus* lactate dehydrogenase promoter

CGCGCCCCGAAAGGGGCGTTTTTTCTCAGGACATATGTCGCGCCCCGAAAGGGGG
CGTTTTTTTGCACCATGGATCCGCGGGACGGGGAGCTGAGTGCTCCCGTTGTTTGCC
GCGGCGTCTGTCATGAAATGGACAAACAATAGTCAAACAATCGCCACAATCGCGCA
TGCATTGCGGTGCGCCTTTCGCGTAAAATATTTATATGAAAGTGTTTCGCATTATATT
GAG

2) P_{idh} *G. thermoglucosidasius* isocitrate dehydrogenase promoter

CGATTTTTGCGTAAGCCGCATGTCTGGATGGCTTGACATATTTTGGAAACAATATG
ATAACAATCGCCTCATCCGTCCGCGTGCAGAATATACAGGTCCGGAGAAGCGGACG
TATGTTCCGATTGAACAACGAGGCTAAATTAGTTTATAAAAAGGTGAGAAGATAGTT
CTATTCTCACCTTTCACAACAAAAATATATTGGAGGTTGTT

3) P_{ala} *G. thermoglucosidasius* alanine dehydrogenase promoter

ACTGCAACGTCAATGAAACGGCGGAAGCGTTAAACATTCACCAAACACGCTCGCT
TATCGGCTGAAACGCATTGCCGAAATTGGCGAAATTGACTTGATGATATCAATCAA
AAGTAAACTATATATCGATATTAACCTGCCAAATACGAAGCGCTACATTGATTT
GTGGAAATCCACAAAATGATAGCGCTTCTTTTTTCTCCGAACAAAGAAAAAATCA
CGATAACTGTTTATA

4) *G. thermoglucosidasius* succinate dehydrogenase promoter (suc)

TTCTGATGATTTTTGGAGGCTGTTTTGTTTTAATTCAAATCTTTGAACATTTTGTGA
CTTTGTTTTCTTGACGCTTGTACAATGTGGGAGTACAATGAATGTGACACAATGATT
AATGGAAATGGAAGCAGTTACACGCTCTATTAATCATTAAATGTTCTATTTTTTGTG
GATGGGGAGCGCTATTTTCATCGTTGTTATTAT

5) P_{thl} *Clostridium acetobutylicum* thiolase promoter

GCGGCCGCTTTTTAACAAAATATATTGATAAAAATAATAATAGTGGGTATAATTAAG
TTGTTAGAGAAAACGTATAAATTAGGGATAAACTATGGAACCTATGAAATAGATTG
AAATGGTTTATCTGTTACCCCGTATCAAATTTAGGAGGTTAGTTCATATG

6) P_{fdx} *Clostridium sporogenes* ferredoxin promoter

GCGGCCGCGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAGAAGTAAGTGTT
ATATATGATGATTATTTTGTAGATGTAGATAGGATAATAGAATCCATAGAAAATATA
GGTTATACAGTTATATAAAAATTACTTTAAAATTAATAAAAACATGGTAAAATATA
AATCGTATAAAGTTGTGTAATTTTTAAGGAGGTGTGTTACATATG

7) P_{fac} A synthetic construct of *Clostridium pasteurianum*'s ferredoxin promoter with a lac operator

AATTCGCGGCCGCACTAGTGAGATAGTATATGATGCATATCTTTAAATATAGATAA
AGTTATAGAAGCAATAGAAGATTTAGGATTTACTGTAATATAAATTACACTTTTAAA
AAGTTTAAAACATGATACAATAAGTTATGGTTGGTGTGGAATTGTGAGCGCTCAC
AATTGGTGTCCAAGCTAGTTAGGAGGTTAGTTCAT

A4. TcdR/TcdB sequences

tcdR

ATGCAAAAGTCTTTTTATGAATTAATTGTTTTAGCAAGAAATAACTCAGTAGATGATT
 TGCAAGAAATTTTATTTATGTTTAAGCCATTAGTAAAAAACTTAGTAGAGTTTTACA
 TTATGAAGAGGGAGAAACAGATTTAATAATATTTTTTATTGAATTAATAAAAAATAT
 TAAATTAAGTAGCTTTTCAGAAAAAAGCGATGCTATTATAGTCAAATATATTCATAA
 ATCATTACTGAATAAGACTTTTTGAGTTGTCTAGAAGATATTCTAAAATGAAGTTTAAT
 TTTGTAGAATTTGATGAAAATATCTTAAATATGAAAAATAATTATCAAAGTAAGTCT
 GTTTTTGAGGAAGATATTTGTTTTTCGAATATATTTTGAAAGAATTATCTGGTATTC
 AAAGAAAAGTTATTTTTTATAAATATTTAAAAGGATATTCTGATAGAGAAATATCAG
 TGAAATTAATAATATCTAGACAAGCTGTTAATAAGGCTAAAAATAGAGCATTAAA
 AAAATAAAAAAAGACTATGAAAATTATTTAACTTG

TcdB promoter

TTAATGAATTTAAGAAATATTTACAATAGAAATCAAATTTTAGAATTAACTTTATTG
 TAAAATCAATAACTTAATCTAAGAATATCTTAATTTTTATTTTTATATAGAACAAAG
 TTTACATATTTATTTTCAGACAACGTCTTTATTCAATCGAAGAGCAAATTAATCAACTG
 AGTGTCTTCAATTTAAAATGTTAGGAAGTGAATGTATATGAAAACCTAAGTAGATAT
 TAGTATATTTTATAAATAGAAAGGAGGATATATAAAAGAGTTTTAGCATTTAGATGT
 AAAAATATTCAATAAAAAATATTATAGTAAAGGAGAAAAAT

A5. Sequences of Gram-positive replicons

UB110.1

CGAATTAATTCCTTAAGGAACGTACAGACGGCTTAAAAGCCTTTAAAAACGTTTTTTA
 AGGGGTTTGTAGACAAGGTAAAGGATAAAACAGCACAATTCCAAGAAAAACACGA
 TTTAGAACCTAAAAAGAACGAATTTGAACTAACTCATAACCGAGAGGTAAAAAAG
 AACGAAGTCGAGATCAGGGAATGAGTTTATAAAATAAAAAAAGCACCTGAAAAGG
 TGTCTTTTTTTGATGGTTTTGAACTTGTTCTTTCTTATCTTGATACATATAGAAATAAC
 GTCATTTTTATTTAGTTGCTGAAAGGTGCGTTGAAGTGTGGTATGTATGTGTTTTA
 AAGTATTGAAAACCTTAAAATTGGTTGCACAGAAAAACCCCATCTGTAAAGTTAT
 AAGTGACTAAACAATAACTAAATAGATGGGGTTTTCTTTAATATTATGTGTCCTA
 ATAGTAGCATTTATTCAGATGAAAAATCAAGGGTTTTAGTGGACAAGACAAAAAGT
 GGAAAAGTGAGACCATGGAGAGAAAAGAAAATCGCTAATGTTGATTACTTTGAACT
 TCTGCATATTCTTGAATTTAAAAGGCTGAAAGAGTAAAAGATTGTGCTGAAATATT
 AGAGTATAAACAAAATCGTGAAACAGGCGAAAGAAAGTTGTATCGAGTGTGGTTTT
 GTAAATCCAGGCTTTGTCCAATGTGCAACTGGAGGAGAGCAATGAAACATGGCATT
 CAGTCACAAAAGGTTGTTGCTGAAGTTATTAACAAAAGCCAACAGTTCGTTGGTTG
 TTTCTCACATTAACAGTTAAAATGTTTATGATGGCGAAGAATTAATAAGAGTTTG
 TCAGATATGGCTCAAGGATTCGCCGAATGATGCAATATAAAAAAATTAATAAAAAAT
 CTTGTTGGTTTTATGCGTGCAACGGAAGTGACAATAAATAATAAGATAATTCTTAT
 AATCAGCACATGCATGTATTGGTATGTGTGGAACCACTTATTTAAGAATACAGAA
 AACTACGTGAATCAAAAACAATGGATTCAATTTTGAAAAAGGCAATGAAATTAGA
 CTATGATCCAATGTAAAAGTTCAAATGATTGACCGAAAAATAAATATAAATCGGA

TATACAATCGGCAATTGACGAAACTGCAAATATCCTGTAAAGGATACGGATTTTAT
 GACCGATGATGAAGAAAAGAATTTGAAACGTTTGTCTGATTTGGAGGAAGGTTTAC
 ACCGTA AAAAGGTTAATCTCCTATGGTGGTTTGTAAAAGAAATACATAAAAAATTAA
 ACCTTGATGACACAGAAGAAGGCGATTTGATTCATACAGATGATGACGAAAAAGCC
 GATGAAGATGGATTTTCTATTATTGCAATGTGGAATTGGGAACGGAAAAATTATTTT
 ATTAAGAGTAG

NCI001

TCCCTTGA ACTCCTCAATCAAGTCATACCATTGGCATATGGTTTTAGTGTATCAGTGGC
 TGGTCCACTACACCACAAATGTAACATAAAAACGGTTACTTTGTAAATATCTATTTTG
 AATTATCTGAAAAATATAAAATGGCAGTCGCACA ACTTAAACCATTAGGTGAAGCCC
 AAATTATTGTGGACATTTCCCTTGATGCGACCGCTTCATTCTCCCTCCATTGCCTAA
 GGCATGCCCTCAGCTCCACTTTGCTTCAAAGATTTGGCTTAGAACTTGTCCAGCA
 GTTCAGCTAATCCCTCCTCTTATCGATAATCTGTGCAAAAAGTTCATTGTTTAGTAT
 GATCTATTGTGCGCTTCCGTGTTTATTTCCATTGTTTCACCTGCTAAAGAGGAAGC
 AACATCTCTTTTTCTATTATATGTAATTA ACTGTCTCTAACGTTTTCGCTCTGTTTGT
 TCTTGTTTAAACA ACTTGTATATCTGTTTATAACTTGTATACTTGTACGGG
 CTTCTGGATCCCTGGGCGGCAAGGGATTGCGGATTTAAACACAACAAAATTGGG
 GTGAATGCACAACAATTTGGGGCGAACGCATAACAATTTGGGGTCAATGCACAA
 CAAAATTGGGGCGAACACAACAATTTGGGGTCAACGCACAACAATTTGGGGGTG
 AAAACAACAATTTGGGGTCTGAAAAACAACATTTTGGGGTGTCCGCAAACGA
 AATTGGGGCCG CAGCGTTCCTTATGAATAAATGGAATGGCTCAAGCCTTATCGGCA
 AAGGAAATGGCATGCACAATCAACAGTTTTGCGGGTGATTGATTTCCCTTCGATTC
 GTTTCATGTTAAATGCAAACA AAAATAAAAAAATGTTTGCTTTTATTTTCTAGTTAT
 TTTATAATAGATTTACCACTATTTTTGCCATATGGGGGATAGAGAGGTAATCAACA
 TGACGATGAAACAAAACCTCTGTTTTAAACCAATTAACTTAGTGTCTAAATCAAA
 TTCGCTCGTTGAAGCGCGTTATAAGTTGGGAACGATGGAACAAAAGGTGATTGCCG
 CCATTGCATCCAATATTTGCCCCGATGACAAGGACTTTCAAACATACACCTTTCCAT
 AAAAGAATTTAAAGAACTCATCGTTTGAAGTCGAAAAATATTTACCATGAGATCG
 ACAAGTTATGACCAAGCTTATGCAGCCGTTCAATTTCATTAAATGCGGAGGGGAAG
 CCGACAAAATGCTTGGCTGTCCAAAGCCACTTATAACGTTGGCGAAGGGACGGT
 CACGGTTCGCTTTGATCCCGATTTGAAACCTTTCTTTTGTTCCTAATGAGAAGTTTA
 CGAGATATAAATTAGGGAATATCATTCAATTTGCGAAGAAGCTACTCGATCAGAATAT
 ACGAATTGTTAAAGAGTTACGAAGGATTGGCCGAGCGAACATTACCCTAGAGGAG
 CTGCGAAGCAAATTGGGAGTAGAAAACAATATCCGCATTGGATCAACTCCGGCA
 AAGGGTGTGGACCCTGCAAAGGAAGAACTAGAGGAAAAGACGGATATTATGTTT
 ACATACGAGCCCATCAAAAAGGCCGTTTCGATCGAAAAAGTCAAGTTTATCATCAA
 GCCGAATCCGAAAAATAAACGTTGATGACGAATATTGTGGAACAATTAGAAAAGG
 TGTTAGAAAACGAAGATGTGCTGGAAATTCAAGAAAAGGAAAAGAACTAGGATT
 GGCATTGCCTAAAAAATTAGTTCAAGAATGGCTGAAATACGAAAAGAAAACGTAA
 TGAAGCTAATGGAGTCCATCCAATATCGCGGCGATATAGAAAACAAGATTGGCTAT
 ATCTCTCCGTTTTGCGAAAAAAGCGGAAAGCGAAGAGAAAATGGCGGATGTGG

ACCAAGTGCAAAAAAGATTTCGCGAAGTCATTCATTATTTTACACCTCGTCAAGGAG
 CGAGAAAAGTAGATTTGTTGCCTGAGTGGCTGGTGAAAAGCACTGCGGTGGATAT
 ATTTGCAAAAGACATGACGCCGAGGAAGCAGAACAGCTTTGGGAAGAAAAAGA
 GATGAAATTATGGAGGAAATCAATAACCGAAGGCGCAAAGTGACCACGTGAACGC
 CCTTTTTAGAGTAATGGTTTGGTATTAGAGCGCACATGTTGGCGTTCTGCTACATTC
 TAAAAGAATCGTTGCTGATCTAGCAATGGAGCTAGGTGGCAAAGAGGCGGCATGC
 CGTGTGGGCCTTTCACAGCGAGAGGCTTGGGGATTCCGCTATGGAGGAATGATATA
 AAAATGCGGAAAGAGGCTGTCCCAAAGGTTGCTAAAACGACCTTTTGCGGACAGC
 CTTCTTATTCATAATAAAGGAAGATAATATAAAAATACAAACAACAATCTTTTCTTATA
 TAAAACAGTGCATTTCTATGAAAATGTGTCCTTGGAGCGGCCATTTTCATGCCGGGT
 GGAGAGCAAAGCGCGCTCATGGAACCTTTTTAATCTAATTGGTCCAACCTCCATTGTA
 TATAGCTGAGAGCAGACCAAGTACCAACAGCTCGAACTTTTGTATCGTGGAATACA
 GTAGCTTTTTGCCATCTTATACCGATATTTAATTCAAATTACCTTCGACATTTCTAAT
 ATAATCAAAGTTATTATATTCATACCAATCCTCTACTGATTTACCAGTATAACTCCATC
 CTGAAGAATAGATCTTATAACCTTCCTTTGGGCTAATAGAGAGAATTTTACTCCATC
 GTAATACCATAGATTTTCCGTCCAAAACCTCGAACAAGTCAGTACCATAGATATTCTC
 ACCGATTAGTGTTTTATGAAACCAATTATAATCGGTTATAGTAGCTTCAGCTTTGGTC
 GGAAGAATGA

NCI002

ACAGGCTGACCAAGGTTTTGACCTTTAGGTCAGCCTGTTTTATATTCACCATATTTTC
 CATTTCAGTGTTTGGATTAACCTCCGAATCGCTGCTGATTCGGTAATATTTTTGTCT
 GACAATATTGTTTCAACAACCTGATAACTCTCATCGTCCAACCGGACGGTTACACGAT
 GCGATAGCGTATTTTCTCTTTAGGGCGTCCCTGTTTTAGCTTTTTCCGTTACATACGTT
 AAATCATCATATACCACGACATCTAAATATCCAAAGCGATCATAAATGTTTTGATTTT
 TCCGCACGAATTCTTCCCACATTTGAGGAATCTCACGAATCAGGCGCAAATATTCTT
 GTTGTAGCTCCCATTTTTCTTTTCTGTAGACGGTCGGCGTTGGGCAAACCTGGGATA
 AGATCAATACCTGAAAATCATGTTGTTTATAGCATCGTCTTCTCTCATAGAAATCAC
 CTTTTTGTCTGACGTAATTTATTTCTATATTATCACAAAAAATAGAATGCGGGAC
 ATCACCTACTACGACATATAAACGAGGACTACGCTCTTTTATTTAACACCTAAATTA
 ATCTTGATCCCTCCACTATATTTAGTTGGGACGCTGTCTTTGGTTTCGGTCGTTGTT
 CTTCTTTCGGCTTCGCTGGCTTCAACGACTGCATAAATTGCCTGAACGATTGCGCTT
 GTATCACCCGAAACGGGTAGCTGTCGTCATAGCGTACCGTTGTCGGATATAATTA
 GCGTCAAACCTTTGGTTCGATTGGGCTCACTTGTGCTGGCACTTTCCAGTTCCTTCGTGA
 AATCGTACCAAATTAAGCCGGTCATTTTATAATCGGTCTCGGCTGGCTTGCAGC
 GAGCTTTTGCGACTTTAGTCTTCGTTTTCACACGTTTCATTAAGTTGTTAAGAGCCTT
 ATCGATGCTAGTCTCATCGGTCATTCATCTCTATATTTTTTCAACCGTTCCTCAAGTTC
 CCGTCGTGCCTGTTCAATGTCGAAATCCTCATCATCCTTCGGTTGGTTATAGTCCATC
 TTCAACCAATCCGGTACTAATTCTTTACGTGTCGGTTTCTGAATTTTTTGCCTTCGCT
 AGAATTAGAGTCTACCATTTTCTTACGGTTTCAAATGTCTTTTCTATGTAGGCCATG
 TATTTGAATCGTTTTCTTTTAGTTTCTCCAACAGTTTGTTTTTCAAGTCATCATCATC
 AATCTTGTTGGCTTGTAGAACTCATGGATCAAATCGGATTGCTGTTGGTTAAAGCC
 TTCAATCAGTGGATGATGATGATTATTATTTGAATCATTATTTATATCATTGAAATCA

TTATTTATATCATTGAATCATTATTAGGTTTCTCTGCGAAACCACCCCGGTTTCTAT
 AGGAAACCACCCCTGGTTTCCTACAGAAACCACCCCTGGTTTCATATTGAAACCATC
 CGAACTTCGTTCTGTTTTCTCCGCCTTTTTGGCAAAGTTCTGCTGCTTGTTCAT
 ATTCAGTATCGTAATCACGAATCTTTCAAGTGGTTGATATTTCTTTGAATATCGTTT
 TGCGGATACTCGTAACTACAATATTTACAGGTTTGTCCCTGGATATTCGTCATCTT
 CATATGTGACAAGATCAATCAAACCATAATTCCATAAAGGCCGGATGATCTCATTGT
 AGAATTTTTCTTTCCAACCCCTAAACGTTTCATGACTTTGGAAAGGCTTGAAGGAAT
 CACATCATTATATTTATCTTTTCGCTGATCTGAACGATCACACCAACTGTAGAATTTT
 AGCCAGGCTGCGAAGGCTTCAGGGCCAAGTTTATCAATCCAATCGTCAGCAACAAC
 AAAATGAAGAATTGGAAGCTGCAATTCGTTTCTGTTTTCTTTGCTCTAGTAAACTCT
 ACTTTTGCCATATATATTCTCCTCTGTTTCCCCTGAAATTCGGGACAAACAAAAGAG
 CGTGCCCGCGTACGTTTATACGTAGGCAAACACGCTCTCATCGGAGGAAAAACGAA
 AATTTCCAGGCGTTTCGCTCTTGCCTTAAACACATTTATGCATTAATAAATGAAGACAA
 GGCATAAATATGTTAACACAAGAGCTTTTCGAGCAGGTGTTTCCGTAGGGTGAGTA
 CGGCAGACGGTTGTAGCGGACCCCAATCCGACTACAACGCGCTTCGAAAGCTCTT
 CTATTTTGTCCGTTCTATTATGTTGTCTCTATTGTACTACTAATCTTCGAGTCGCGAC
 AAGGTCTATTTCCAAAATTCTCGACAGCCATCCCGTTTTTTTTGTCCCCTTGTCCCTC
 CGTTGTGCTGCTGTCTCCAACGCGGCCTGCTTCTGTTCTTGAATGGCGCGAATAGA
 TTCCATGAGTTTTCGGTCGCGTTCTTCCAGTCGTTGGTTGATGTAGCGTTGCTGCTCG
 TTGAGCTTTTCAATCAATTGACGATTCACCATTTCCAAGCGCTCGATTGCTCGATTA
 GTTGATCTAGCTTGTGTCATAGCGCGTCATATCGCTAGGTACGTCTAGCTCTTGAG
 CGATTTCTCCGTTTCTTGACACGATTAACGACTTCGTTTACCGCGTTTTCAAGCGA
 CATTCTTCGTCTTGTGTGAGTCGTTTTAAATGTCGTAGGGCAACGATATCTTTTTCG
 TAATAGAGACGCTGTCCTTTGTCGTTCTTCTTAAACGGGTACCCGGCGTCTTCGAGC
 ATGGCGCTCCATTTCCGTAAAGTTGACGCGGCAATCTCCAAATTATCCGACACGCTT
 TTCGAATAAAACGCGATTTTGTCCGACATAGCGCTATCAGCTACCTTCCAAACAGTT
 TTTCTTTATTATACAACGTTTTAGACTCGCTATGTAGCGATTTTTGTGCGATCCAGTG
 CGCTATGAGATAGTAAATATTGAAAATGCATTTTGAAAATGGCGCGAATAGATTCC
 ATGAGTTTTCGGTCGCGTTCTTCCAGTCGTTGGTTGATGTAGCGTTGCTGCTCGTTG
 AGCTTTTCAATCAATTGACGATTCACCATTTCCAAGCGCTCGATTTGCTCGATTAGTT
 GATCTAGCTTGTGTCATAGCGCGTCATATTGCTGGGTACGTCTAGCCCTTGAGCGA
 TTTCTCCGTTTCTTGACACGATTAACGACTTCGTTCCGACACTTTCTAGCGCCATC
 CCTTCGTTTTGTGTGAGTCACTTCAAATGGCGAAGGGTGACGTTGTCCTTTTCGTAG
 TAGAGATGTTGTCCCTTGTGTTTTCTGAATGGATACCCGGTGTCTTCGAGCATG
 GCGCTCCATTTCCGCAAAGTTGACGCGGCAATCTCCAAATTATCCGACATGCCTTTC
 GAATAAAACGCGATTTTGTCCGACATAGCGCTATCAGCTACCTTCTCATAGTTTTTC
 CTTATTATACTTGGGTGCGCTATATAGCGATTTTTGTGTGATCTCGTACGCTATTAAA
 CAGGTAATATTGTATATGCATTTGGAAACTGATATACTTAGATTAACCTTTGGACTT
 AAAGTTAGTCAAATTCGAGGCGAAGGGTTGGGTATATTGGT

BST1

GCGCGCCTCGCGCGTTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCC
 CGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCA

GGGCGCGTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATCA
GAGCAGATTGTAAGTGCACCATATGTTCCCTAAGGAACGTACAGACGGCTT
AAAAGCCTTTAAAAACGTTTTTAAAGGGTTTGTAGACAAGGTAAAGGATAAAACAG
CACAATCCAAGAAAAACACGATTTAGAACCTAAAAAGAACGAATTTGAACTAACTC
ATAACCGAGAGGTAAAAAAGAACGAAGTCGATTGAGCACTACAGGAACAAAA
AGAGCGATGATTGAAGGAACGCAAAAAACAAGAGAAGGTCACATTGAAATAGCAG
CGACAACGGAGGACAAGACAACAGAGGACAAGAGGACAAAAAACAGGGATGGC
TGTCGAGAATTTTTGGAAAATAGACCTTGTGCGGACTCGAAGATTAGTAGTACAATA
GAGACAACATAACAGAACGGGACAAAATAGAAGAGCTTCCGAAGCGTGTGTAAAGT
GGTGTCCCAACACCGTTCGCCGTAACGCGCTACGGCAAACACCTGCTCGAAAGCT
CTTGTGTTAACATATTTATGCCTTGTCTTCATTTAATGCATAGATATGTTGAATGCA
AGAGCGAAACGCTTGAAATTTTCGTTTTCTCCGATGAGAGCGTGTTCCTGCG
TATTAATGTACGGCAACGCGCTCTTTGTTTGTCCGAAAAATTCGGGAGAAACGG
GAGGAGAACGGAAATGTACATAACTAGCGAACAATTCACGCAACTGAAAGAA
ATCGAGAGAGTCACAGATTTGTTTTATAAACGCCGTAAGATTAATAATGAGCTGAT
GGAGCGTATCGCCGACGCATTAGGCCGCAAGTTCGCAAAGCGCAAAAAACAACG
CGCGTCATCGACGAAATTGTTGGCTCGCAACAGACCGCGGTTTTTCGTTCCAGGG
CGCGAAACGCTTGCCAAAAAGTTAGGCGTGTCTTGCCGACTGTTGATCGTGCTACT
CGCATGCTCAAAGACTCGGGCGAAGTCGTCGTTTGCTATCGTGAGAATCCGAATAG
TAACGGACCTAAAACGCCTGTATTCTTTCCGGAGCCATGCAAATTTGAGCGTAT
TGCAGCCGTTTTAACTTGCGTGACAACGAAGCTGATAAAGTAGAAAACGCCGAAA
AGCCTACGGAATCAAGCGATTTAGCGCGTAAAACGGACGCTACCATTAGTTCACCA
GTTGAAAAACATGACGATATTAATAAATAATAGACAATACACATCCGTTGGATAAA
ATTGTAATAACGTCACTATTAAGTGAACGAAGTGCAGAAGCGTCATCCTTACGG
CATTGATATCTATCCGCATACATAAAACAAACACTCGACGACCTAATGCAACGTGC
AAAAGCGAAGGTCAAAGCCGAAAAGCTCGCGCCAACAACGCAAACAAGCCGAC
TATCAACCGACCGCGACCGGCTTAATTTGGTACGACTTCACGGCGGACACAAGCGC
CAGCCAGCCGACAAGCCAGCGCAAGCCAGCGTTTGACAGCGAAGACATGTTTGCGC
CGTTATTGCAGGCTGTTGCACGAATTAAGGGCGTTGAGGTGTGACACACTCGTTTTA
AGGCGTCTAACAGGCGCGTAGAGACGCTTTAAGTCGGGGGAATATTAATAATCC
ACGCCTCCTAAAACGCGATATACGCCGATTATGACGGTCGTTTGCTTACGTAACGT
GCAACTACCTTCGCTCACATCGTCCAAAATCGTCATAGTCAACGTCTTTCTTCGTGA
GCTGGCGCAAAGCCTTAATGATTTTCGATGCCGTTTTTATTGTCGGCGAAAATTTAT
CTGACTGGCAAAGGCGGCTGATCGTGCTTTTGCTGACGCCTGCTTTGTTGCGACGT
CTTGCTGCGTGATGCCGTGTTGTCTAAAATCGCGCAAATTTGATCGAGATCAGG
GAATGAGTTTATAAATAAAAAAAGCACCTGAAAAGGTGCTTTTTTTGATGGTTTT
GAACTTGTCTTTCTTATCTTGATACATATAGAAATAACGTCATTTTTATTTAGTTGC
TGAAAGGTGCGTTGAAGTGTGGTATGTATGTGTTTTAAAGTATTGAAAACCCTTAA
AATTGGTTGCACAGAAAAACCCCATCTGTAAAGTTATAAGTACTAAACAAATAAC
TAAATAGATGGGGTTTTCTTTAATATTATGTGTCCTAATAGTAGCATTTATTCAGAT
GAAAATCAAGGGTTTTAGTGGACAAGACAAAAGTGGAAGGTGAGACCATGGA
CTCTAGTGGATCGACTAGAGTCCGACGACTGCTCCGGCCGTCCAGAATGGAAAAG
AAGAGGGGCGGATGTCTCTCCGTTGGCGTGGCGACGATTAATATTTCAAGCACG

GGAAATCGTCGTACACCGAAGTGCTGGAATGGGCCGCTGCGCAAAAAATCGACCG
 AAAAAATAATATAAATCGGATATACAATCGGCAATTGACGAAACTGCAAAATATCCT
 GTAAAGGATACGGATTTTATGACCGATGATGAAGAAAAGAATTTGAAACGTTTGTC
 TGATTTGGAGGAAGGTTTACACCGTAAAAGGTTAATCTCCTATGGTGGTTTGTTAA
 AGAAATACATAAAAAATTAACCTTGATGACACAGAAGAAGGCGATTTGATTCATA
 CAGATGATGACGAAAAAGCCGATGAAGATGGATTTTCTATTATTGCAATGTGGAAT
 TGGGAACGGAAAAATTATTTTATTAAGAGTAGTTCAACAAACGGGCCAGTTTGT
 GAAGATTAGATGCTATAATTGTTATTAAGGATTGAAGGATGCTTAGGAAGACGA
 GTTATTAATAGCTGAATAAGAACGGTGCTCTCCAATATTCTTATTAGAAAAGCAA
 ATCTAAAATTATCTGAAAAGGGAGGCCGCC

GEOTH02

CCATCACCCACCACACTGCGATGTTCTCGTGTCTTTTCTTCTCCATCCACTTCGTT
 TAGCTTCTTTAATCACAAATCCTCATGACACACACTCACACCCAGACACCCACCATA
 ATATCCTAATCCCTCACACAAATCCCCACACCTACAGGATCCAAAGGATCCAATCA
 ACTAACCGAAAAAAGTCATATCTCTCACAACTGTAACCGAACCTCTCTCATCAG
 ACATCCCCCTCTCCGATAAATCAATAACTCCATCCATAACCACATAATCCATAACC
 ACAAAGATTATTAGCAACACAGGCCAACATTCTTACCCTAGAATACAAAAAACA
 TCCCCAGTAATATAAAGAAACACCCTAAGAAATACTTAGAGCGTTAAGAAAGCCA
 ACCCCTTATCCAATTATAAAAACGGTTTAAAGGGACAGACCATTAACTTTTTTATAA
 TACTCCAGCAACAGCACATGTCCATTATAAAGAACGGTTTAAAGAGACAAACCATT
 TCACGGGAATATATGTTCTTATAGTTCCTTTTTAATTGCTGAACACGTCTGATTGAA
 ACACCAAGAATAGCGGCTAGTTCCTTATTTGTTGCCTTTGGGTTCTTCTCCAGCAACT
 CTTGTAATTTAAATAATTTGTCATCCGTCTTATCATGCTGCTGTTAATATATTCGTCG
 CGTTGCAACATACCTTTTTCACGACGTTTTTTCGTTTTACGCTCTGTGTCAGTTTTTG
 TTTCTCATTATCATCGATAAGAGTGGTCATCTGTTGCATTTCTCCGGCGTAATCTCC
 AATTCCTCGATGATCGTCTCGTTTTTCATCGGCCGCTTAATCCGTCCCTTGCTGT
 ACCACATTCTAAAATCCCGCTTCTGGAACCTCGTCAAAGAACTCCATGGCATCGTCGT
 AGGCATTGCCAGCGGTCCGTTTGACTTCACTTGCTTTCTGTGGTTCTTGAAACGAT
 CATTGAGCTGCAGAGCAAAATCAATCGTGGCTTGCTTGTTTTTCAGGATTAACGCCA
 CTGTATAAGCGTATATGTAGGTTAGAACATTTCTGTTTTCTATATCTCCATTTGGAAG
 GGTAACGAGTAATTCTAAATCGTTTTTACGCGCCGATTTAGGCTGTACAGCGTTTTT
 AATCCCCTTTAGGTGTAATATACTAACCGTTCCTTTTCGGCGTTTTACGGGCTTTC
 TACGGCGTTCTAGCGGTGTACAATAACTATATAACTCTTCCAACCTGACTCTAATGT
 TCTCCATATCTCCACGGTAACCTGTTTTCCATTTCCGCCATTAATGCTGTAAGGCAAG
 CGGAACACACGAGTCACATCAGTAGCGGCTGTGTCCGCTCCTAAATGTTTTAATTG
 GCAATATATTGAGTGGTAATGTATTGCGTGAGAAAAGCCATTGCCGGCGCCGCTCC
 GCCGCTAATGGAGTAAACCAGCTGCAGACCGCGCCCCGAGAAAATCACTAAATTGG
 GATTAGGGATTTTCCCTTAATGATCAGCTGCTTAATCTCTTCTAAAGCCTTTGAAAT
 GGAAACATTGACTTTGTAGCAATCAATATCAACACCGATGTTACGGATCTGTTTTAA
 TGCTTTGGTGGTACGTGAACCATATTCAAAGCGTTCAGGCTTAATAACACATCAGT
 TTTTCGTTTTTTCGCCTCTTGAGTAGCGCATTTAAATCGTTAAGCCATAGAAGCGG

TGTTTTTTGTTGGCCAACTGCAGCGTACAAACAAAACCCGTTTTCTTCGATTCCGGACA
 AATAGCAATCGTGCCATTCTTCAATGTACGTTTCTTGTACTTGCAACAAAAAGCCAC
 GTCGAAAACCTCTTCAACGTGGCTGCATCTCCTTTTAGCAACAAACCCCAAAAAT
 GTATGGAAAACTGGACACATCGTTATTGAAATATGTCCTCACTTAGTTTATTATGTT
 GATAGGTATGTAGTGGTGGGGTGGTTGCTTCCCAGTGTTGCCAGCACTGTGTGGGG
 ATGCAGCCACTTTTTATTTTTCCGTTCTCTCTATCTGTCTGTTATTCTATCAAACCTTG
 ACACAAATCTACAACCTTAGGGTTGAAAACGAAAAGAAGGGGAACGCCACTTTGCGT
 TATCCCCTCTTGAATGTTATAATTAATTAGCCTGTTTTTTACCAATAAGCCATCGTG
 TTGCTGCACGGTGGC

A6. Sequences of marker genes

Kanamycin resistance (*kan*)

GTGAATGGACCAATAATAATGACTAGAGAAGAAAGAATGAAGATTGTTTCATGAAAT
 TAAGGAACGAATATTGGATAAATATGGGGATGATGTTAAGGCTATTGGTGTATG
 GCTCTCTTGGTCGTCAGACTGATGGGCCCTATTCGGATATTGAGATGATGTGTGCA
 TGCAACAGAGGAAGCAGAGTTCAGCCATGAATGGACAACCGGTGAGTGGAAGGT
 GGAAGTGAATTTTTATAGCGAAGAGATTCTACTAGATTATGCATCTCAGGTGGAATC
 AGATTGGCCGCTTACACATGGTCAATTTTTCTATTTTGCCGATTTATGATTCAGGT
 GGATACTTAGAGAAAGTGTATCAAACCTGCTAAATCGGTAGAAGCCCAAAAGTTCCA
 CGATGCGATTTGTGCCCTTATCGTAGAAGAGCTGTTTGAATATGCAGGCAATGGC
 GTAATATTCGTGTGCAAGGACCGACAACATTTCTACCATCCTTGACTGTACAGGTAG
 CAATGGCAGGTGCCATGTTGATTGGTCTGCATCATCGCATCTGTTATACGACGAGCG
 CTTCCGGTCTTAAGCAAGCAGTAAAGCAATCAGATCTTCCCTCAGGTTATGACCATCT
 GTGCCAGTTCGTAATGTCTGGTCAACTTTCCGACTCTGAGAACTTCTGGAATCGCT
 AGAGAATTTCTGGAATGGGATTGAGGAGTGGACAGAACGACACGGATATATAGTG
 GATGTGTCAAACGCATACCATTTTGA

Spectinomycin resistance (*aad9*)

TTGAATACATACGAACAAATTAATAAAGTGAAAAAATACTTCGGAAACATTTAAAA
 AATAACCTTATTGGTACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCA
 AATAGTGATCTTGACTTTTTAGTCGTCGTATCTGAACCATTGACAGATCAAAGTAAA
 GAAATACTTATACAAAAAATTAGACCTATTTCAAAGAAAATAGGAGATAAAAGCAA
 CTTACGATATATTGAATTAACAATTATTATTACAGCAAGAAATGGTACCGTGGAATCA
 TCCTCCCAAACAAGAATTTATTTATGGAGAATGGTTACAAGAGCTTTATGAACAAGG
 ATACATTCCTCAGAAGGAATTAATTCAGATTTAACCATAATGCTTTACCAAGCAAA
 ACGAAAAAATAAAGAATATACGGAAATTATGACTTAGAGGAATTAATACTGATA
 TTCCATTTTCTGATGTGAGAAGAGCCATTATGGATTCGTCAGAGGAATTAATAGATA
 ATTATCAGGATGATGAAACCACTCTATATTAACCTTTATGCCGTATGATTTAACTAT
 GGACACGGGTAAAATCATACCAAAAAGATATTGCGGGAAATGCAGTGGCTGAATCTT
 CTCCATTAGAACATAGGGAGAGAATTTTGTAGCAGTTCGTAGTTATCTTGGAGAG
 AATATTGAATGGACTAATGAAAATGTAAATTTAACTATAAACTATTTAAATAACAGA
 TTAAAAAATTATAA

Gentamicin resistance (*aac*)

ATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTT
 AGGTGGCTCAAGTATGGGCATCATTGCACATGTAGGCTCGGCCCTGACCAAGTCA
 AATCCATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCT
 ACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACCTGCTCCGTAGTAAGACAT
 TCATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACG
 TTCTGCCAGGTTTGAGCAGCCGCTAGTGAGATCTATATCTATGATCTCGCAGTCT
 CCGGCGAGCACCGGAGGCAGGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCAT
 GAGGCCAACGCGCTTGGTGCTTATGTGATCTACGTGCAAGCAGATTACGGTGACGA
 TCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAAGTGATGCACTTTG
 ATATCGACCCAAGTACCGCCACCTAA

A7. Sequences of the Toxin and Antitoxin genes**Prevent host death protein (*phd*) antitoxin**

ATGGGCGTGAAATGTTGGGGAACAAATATGAAAAATGTACACACGGCGGATTA
 ACCAAGTTGGGAATAGTTTGTCCGTGAATATCCCGAAGGATTTAGCGACGATGTTA
 AACATAAATAAAGGGGACGAAATGGAAATATATTACGATAAAGAACGGGGGGAAA
 TCGTGATGAAACGCGCAAACCGGATTCCAAAAGGGGTCCGTCCCGAAGTAGTGATG
 GCGATGAACCGTGCGATTGCCAAATATGATGAAGCGCTGCGGAACCTGAAAGACA
 GATAA

Death on curing protein (*doc*) toxin

ATGGTTTATTATTTAACAGCCGAAGAAATCATCTTCTTGCAATTATACGATTATGGAAA
 TGTACGGTGATGCCGAACAAGCGGGGATCAAATTCCCGATAAATTTGCCTGGATG
 TTGGAAAGGCCGAAAACGAAATTGTTCCGGGGAGGAGCAGTTTTCTTCGATCATTGA
 AAAAGCGTGCTGCTATTACCATTCCATCGCGACGGGGCATATTTTTACAATGGCAA
 TAAACGGACGGCTTTAGCCGTATTCGTCACTTTCTTGATTTGAACGGGTATGAATT
 AACGATGACCAATAAAGAGGGCAGAAGATTTACGGTGTACCTTGCCGAAGACGCCA
 AGTTCCGGGGAAATGATTGCATTCAGCACCTTGTCCGGAATTAAGATTACATCC
 GCCCATCCAAAACAAAAGAGTGA

A8. Codon optimisation table

<i>Geobacillus thermoglucosidasius</i> [gbbct]: 27 CDS's (10634 codons)											
fields: [triplet] [frequency: per thousand] ([number])											
UUU	24.5(261)	UCU	7.4(79)	UAU	23.7(252)	UGU	2.8(30)
UUC	9.2(98)	UCC	8.4(89)	UAC	11.3(120)	UGC	3.1(33)
UUA	27.7(295)	UCA	7.1(75)	UAA	1.3(14)	UGA	0.9(10)
UUG	21.9(233)	UCG	13.2(140)	UAG	0.3(3)	UGG	7.9(84)
CUU	20.1(214)	CCU	7.6(81)	CAU	20.2(215)	CGU	7.5(80)
CUC	7.3(78)	CCC	0.7(7)	CAC	4.8(51)	CGC	18.9(201)
CUA	5.5(58)	CCA	7.7(82)	CAA	27.6(294)	CGA	3.8(40)
CUG	14.3(152)	CCG	15.9(169)	CAG	11.1(118)	CGG	7.8(83)
AUU	46.6(496)	ACU	6.4(68)	AAU	21.8(232)	AGU	4.1(44)
AUC	21.6(230)	ACC	8.1(86)	AAC	15.9(169)	AGC	12.9(137)
AUA	8.6(91)	ACA	18.1(193)	AAA	60.5(643)	AGA	7.9(84)
AUG	25.0(266)	ACG	17.5(186)	AAG	12.3(131)	AGG	1.3(14)
GUU	17.4(185)	GCU	15.4(164)	GAU	40.4(430)	GGU	12.5(133)
GUC	15.2(162)	GCC	11.7(124)	GAC	19.1(203)	GGC	22.7(241)
GUA	14.7(156)	GCA	14.7(156)	GAA	76.7(816)	GGA	19.2(204)
GUG	17.5(186)	GCG	30.9(329)	GAG	19.2(204)	GGG	12.4(132)

Coding GC 43.21% 1st letter GC 54.05% 2nd letter GC 33.65% 3rd letter GC 41.93%

Figure A8.1. Codon usage table for *G. thermoglucosidasius*. (Source; Codon usage database. Data source; NCBI-GenBank Flat File Release 160.0 [June 15 2007]).

A9. *n*-butanol gene sequences

A9.1 CoA dependent synthetic genes

thl

ATGAAAGAAGTGGTGATTGCGTCGGGAGTTCGCACAGCGGTTGGAAAATTTGGAG
GCACATTGTAAACGTCCCGGCTGTGGATTTAGGCGCTGTCAATAACAAACGCTCG
ATTAACGCGCTAATGTCAAACCGGAAGACGTATCGGAAGTCATCATGGGCAATGT
GTTACAAGCGGGCTTAGGCCAAAATCCTGCTCGCCAAGCTGAAATCAAAGCTGGAA
TCCCTGTGGAAGTGCCTGCGATGACGGTTAATATGGTGTGTGGCAGCGGCTTACGC
GCGTTACGTTAGCTGCGCAAGCGGTCATGTTAGGCGATGCGGATATCGTGGTTGC
TGGCGGCATGGAAAACATGTGCGCGCTCCGTACATTTTGAATGATGCGCGCTTTG
GCTATCGCATGAACAATGGCCAATTAGTCGACGAAATGGTCTACGATGGCTTGACG
GACGTGTTCAACCAGTATCACATGGGCATTACGGCTGAAAACCTGGCGGAAAAATA
CGGCATCAGCCGCGAAGAACAAGACGAATTTGCGTATCGCAGCCAGAAATTGGCGT
CGGAAGCTATCAGCTCGGGCCGCTTTGAAGATGAAATTGTCGCGTATTGTGCCG
CAGAAAAAAGGCGAACCGATCGAGTTCAAAGTGGACGAACATGTGCGCCCGAATA
CGACGATTGAAGCGTTGGCGAAATTAACCGGCGTTCCAAAAGATGGCACAGTC
ACGGCGGGAATGCGAGCGGAATCAACGACGCTGCTGCGGCGGTGGTGGTCATGA

GCAAAGAAAAAGCGTGCGAATTAGGCATCAAAACAATTGCTACAATCAAATCGTTC
 GGCTACGCTGGCGTGGATCCGTCGATTACAGGAATCGGCCAGTGTACGCGACGC
 GCAAAGCGTTAGAAAAAGCTAACTTAACGGTTGACGACTTGGACTTGATCGAAGCT
 AATGAAGCTTTTGCGGCGCAGAGCTTAGCGGTCGCGAAAGAATTGAAATCAACAT
 GGATCGCGTCAACGTGAACGGAGGCGCGATTGCGATCGGCCATCCTATCGGCGCTT
 CGGGCTGCCGCATCTTGGTCACGTTATTGTATGAAATGCAGAAACGCAACAGCCAT
 ACGGGCTTGGCTACATTATGCATTGGAGGAGGAATGGGAATCGCTATGGTCGTCGA
 ACGCGCTAGTGATTATAAAGATGACGATGACAAATAATAA

hbd

ATGCAAAAAATCTGTGTGATTGGCGCTGGCACAATGGGCTCGGGCATTGCTCAAGT
 GTTCGCTCAGAACGGATTTGAAGTCATCTTGC GCGACATCGACATGAAATTCGTCGA
 AAAAGGCTTCGGCACGATCGAAAAAATCTACAAAGAAATGTTGACGAAAGGCAAA
 TTACAACAAATGCGCAAACGCATTTTATCGCGCATTTCGCGGCACAACGAACCTTAGAA
 GACGCTAAAGAAGCTGACTTCGTGGTAGAAGCGGGCATTGAAAACATGGACTTGA
 AAAACAGATCTTCAAAGAATTGGACGAAATCTGCAAAATGGAAACGATTTTAGCG
 AGCAATACGTCGTCGTTATCGATTACAGAAATTGCGAGCGCGACGAAACGCCCGGA
 AAAAGTCATTGGCATGCATTTTTTAAATCCGGTGCCTGTGATGAAATTAGTTGAAGT
 GATTAAAGGCTTAAAACGTCGGAACAGACGTTTAAATGTGGTCCGCGAATTGGCTT
 TGAAAGTGGATAAAACACCGATTGAAGTAAAAGAAGCGCCGGGCTTTGTGGTGAA
 CCGCATCTTAATCCCGATGATCAACGAAGCGATCGGCATTTTGGCGGTGGTGTTAG
 CGACAGACAAAAGCATCGACGAAGCGATGAAATTAGGCGCGAATCATCCTATCGGG
 CCGTTAGCGTTGAGCTCGTTAATCGGCAACGATGTGCTGTTGGCGATCATGAACGT
 GTTGATGAAGAATATGGCGACAGCAAATATCGCCACATCCGTTGTTGAAAAAAG
 TGGTGC GCGGGCTTGTAGGCCGAAAACGGGCAAAGGCTTCTCGAATACAAA
 ATCAACTTGTGCGCCGCCGCATTTTCGGCTAGTGATTATAAAGATGACGATGACAAA
 TAATAA

crt

ATGGACTTTAACAATGTCTTGTGAACAAAGATGACGGCATCGCGTTAATCATTATT
 AACCGCCCGAAAGCGTTGAATGCTTTGAACTACGAAACGTTGAAAGAATTAGACAG
 CGTCTTGGACATCGTAGAAAATGACAAAGAAATCAAAGTGTTGATTATTACGGGCT
 CGGGAGAAAAAACATTTGTGGCGGGAGCGGATATCGCGGAAATGTGGAACATGAC
 ACCGTTGGAAGCGAAAAAATTTAGCTTGTACGGCCAGAAAGTTTTCCGCAAAATTG
 AAATGTTGTCGAAACCGGTGATCGCTGCTGTGAACGGCTTTGCGTTAGGCGGAGGC
 TGCGAATTATCGATGGCGTGCGATATTCGCATCGCGAGCAAAAATGCGAAATTCGG
 ACAACCGGAAGTGGGCTTAGGCATTATTCGGGCTTCTCGGGCACACAGCGCTTAC
 CGCGCTTGATTGGCACGAGCAAAGCTAAAGAATTGATTTTACGGGGCGACATGATC
 AATAGCGACGAAGCTTACAAAATCGGCTTAATCTCGAAAGTCGTGGAATTGTGCGGA
 TTTAATCGAAGAAGCGAAAAAATTGGCTAAAAAATGATGAGCAAATCGCAGATTG
 CTATTTGTTGGCGAAAGAAGCTATCAACAAAGGAATGGAAACGGACTTGGATACG
 GGAAATACGATTGAAGCTGAAAAATCTCGTTGTGCTTTACGACGGACGATCAAAA
 AGAAGGCATGATTGCGTTTAGCGAAAAACGCGCTCCAAAATTTGGCAAAGCTAGTG
 ATTATAAAGATGACGATGACAAATAATAA

Ter₇₃₃₄

CTAGTTTATATTGAAGGAGGATGAATGCAATGCAAATCAAACCAATGATCCGCAAC
AACATCTGCTTGAACCTCGCATCCTAAAGGATGTGCTAAAATGGTGCAGAACCAGATT
GAATATGTGCGCCAGCACATGGCTGGCTTGTGCAACACAAAAGGAGCGCCAAAATT
AGCTTTGGTGATTGGATGCAGCACAGGCTATGGCTTAGCGAGCCGCATTGGCGCGG
CTTTTGGATATGGAGCGGCTACGGTTGGAATCTCGTTTAAAAAGCTGGAAGCTCG
ACGAAACCGGGCACACCGGGCTGGTATGACAACTTAGCGTTTACACAGAAGCTGC
TAAAGCGGGCTTAATTAGCAAACGTTGGATGGCGACGCGTTTTCGAATGAAATGA
AAGATCAGGCGATTGCTGCTATCCGCGAAGTAGCGGCGAAAGCTGATATTCCGGCG
CAAGTGGACTTAGTGATCTATTCGTTGGCTTCGCCTGTCCGCACAGATCCTCGCGAC
GGCGTGATGTATCGCTCGGTGATTAACCGATTGGCCAGGCTTACAGCGGAAAAAC
AATGGACATGTTGACGGGCAAAATCAGCATCGCGTCGGCTGAACCGGCTACAGAA
GAAGAAATTGCGCATACAGTCAAAGTTATGGGCGGAGAAGATTGGGAATTGTGGA
TTGACGCTTTGCAATCGGCTGGCGTGTTAGCGCCGCAAGTTCGCACGTTGGCGTAC
ACGTATATCGGCCCTGAATTATCGTGGCCGATCTACAAAAATGGAACAATCGGAAA
AGCGAAAGAACATTTGAAAAAACGAGCCATTTGTTACATGAACGCTTGAGCAAAA
CAGGCGGAGCTGCGTGGGTGTCGGTGAATAAAGCGTTAGTTACACGCGGAGCGC
GGTGATCCCGATCATCCCGTTGTACATCTCGGCTTTGTTTAAAGTAATGAAAGAAAA
AGGATTGCATGAAGGCTGCATTGAACAGTTGGTACGCTTGTATAAAGACCGCTTAT
ATACGGCGGAAGCGGCGTTGAATCCGAACAAAGTTCGACGGACGATGCTCAACG
CATTTCGATTGATGACTGGGAAATGCGCGATGATGTGCAGCGCGAAACAAAAGCG
CGCATGGAAGCGATCACGGAAGAAAATTTGTTACGTTAGCTGACGTCGAAGGCTT
CAAACATGACTTCTAGAAGCTCATGGATTGATGTGGCTGGAGTTGATTATGAAG
CGGATGTGGATCCGTCGACAATT

Ter₆₁₉₂

GAATTCGCGGCCGCACTAGTTTATATTGAAGGAGGATGAATGCAATGGTCATTA
CCTATGATTCGCAACAATATCTGCATGAATGCTCATCCGGAAGGATGCCGCTTATTC
GTGGATCGCCAAATCGACTATGTCCGCCAACGCGGACCGTTAAGCAGCGGACCTAA
AAATGTGTTAGTGGTTGGATCGTCGGGCGGATATGGCTTAGCGACACGCATTGCTG
CGGGCTTCGGCGCTGGCGCTGCTACGATCGGAGTGATTTTTGAAAAAGAACCGTCG
GAAAAACGCACGGGCACACCGGGCTGGTATGCGGCTCGCCATTTCTAAAACGCGC
GGAAGCGGCTGGCTTAAAGTGATGACATCAACGGCGACGCGTTCTCGTTTGAAG
TTAAAGAACAGGTGCTGGACTTAATCAAAAAAGAATATGGAACGATTGACTTGTT
GTCTACAGCTTAGCGAGCGGCGTTTCGCACAGATCCGGTGTCTGGGAGTGACATATCG
CTCGGCGTTAAAACCGATTGGCCGCACGTATACATCGAAAGCTTTAGATCCGTTGAA
ACGCGAAGTTATTCAGGCGACAATCGAACCGGCGAGCGAAGAAGAAATCGAAGCT
ACGGTGAAAGTGATGGGCGGAGAAGATTGGCAGTTATGGGTAGAAGCTTTGAAAG
AAGGAGGCGTCTTAGCGGAAGGCGCTGTGACGGTTGCGTATTCGTATATTGGACCG
GAATTGACACGCCCATCTATCGCGAAGGCACAATTGGCAAAGCTAAAGAACATTT
AGAAGCGACGGCGAAAACGTTGGATGAAGCGTTAGCGTCGGTGGGAGGAAAAGC
GTATGTGTCGGTCAATAAAGCTGTGGTACACGCGGAGCGCGGTGATTCCGGCTG

TCCCGTTGTA CTTGAGCATCTTGTTTACAGTCATGAAAGCGAAAGGCATCCATGAAG
 GCTGCATTGAACAGATGTATCGCATGTTGGCTGAACGCTTATATGCGGGAGGCCCT
 GTGCCTGTGGATGACGAAGGCCGCATTTCGCATGGACGATTGGGAAATGCGCGAAG
 ACGTGCAACGCGAAGTGGCGGAAGTGTGGGAACGCATTACAACAGAAACATTAGA
 CCGCGAAGCGGATGTGGACGGCTACTTAACAGACTTTTTGAACTTGCATGGCTTTG
 GCTTCAAAGAAATTGACTACGAAAAAGAAGTCGATCCTCGCGAAGGA

adhE

CTAGTTTATATTGAAGGAGGATGAATGCAATGGCGACAGCTAAAATGGAAACGGAT
 GTACAGAAACAGATCGATTTGTTGGTCTCGAAAGCTCAGGAAGCTCAGAAAAAATT
 CATGTCGTACACACAGGAACAGATTGATGCGATTGTTAAAGCTATGGCGTTGGCGG
 GAGTTGATAAACATGTCGAATTAGCTAAATTGGCTTACGAAGAAACAAAAATGGGA
 GTATATGAAGACAAAATCACGAAAACTGTTTGTACAGAATATGTCTACCATGAT
 ATAAAAACGAAAAACAGTGGGCATTATCAATGAAAATATCGAAGAAAATTACAT
 GGAAGTTGCTGAACCGATTGGAGTTATCGCGGGCGTGACGCCGGTGACAAATCCA
 ACGTCGACAACGATGTTCAAATGTTTGATTAGCATCAAAACACGCAACCCAATTATC
 TTCTCGTTTCATCCGAAAGCGATTAATGTAGCATCGCGGCTGCGAAAGTGATGTAT
 GAAGCGGCGTAAAAGCTGGCGCTCCTGAAGGCTGCATCGGATGGATCGAAACAC
 CATCGATTGAAGCTACACAGTTGTTAATGACGCATCCTGGCGTCTCGTTAATCTTAG
 CGACAGGCGGAGCTGGCATGGTGAATCGGCGTATAGCTCGGGCAAACCAGCGTT
 AGGAGTGGGACCTGGCAACGTCCCGTGTACATTGAAAAATCGGCTAATATCAAAC
 GCGCTGTTTCGGACTTGATCTTATCGAAAACGTTTCGACAATGGAGTAATCTGCGCGA
 GCGAACAAGCGGTCATTATCGATGAAGAAATTGCTGATGAAGTTAAAAAATTGATG
 AAAGAATACGGCTGTTACTTCTTGAATAAAGATGAAATCAAAAAATTGGAAAAATTT
 GCGATCGATGAACAGTCGTGTGCGATGAGCCCGGCGGTGGTGGGCCAACCGGCGA
 CGAAAATTGCGGAAATGGCTGACTTTAAAGTTCCGGATGGCACAAAAATCTTAGTC
 GCGGAATACGAAGGAATCGGCCCGAAATATCCGTTGAGCCGCGAAAAATTATCGCC
 AATCTTGGCTTGCTATACAGTTAAAGATTACAACGAAGGCATTAAAAAATGTGAAG
 AAATGACAGAATTTGGCGGCTTAGGCCATTCGGCGGTTATCCATAGCGAAAACCAG
 GACGTCATCAACGAATTTGCTCGCCGCGTGCGCACAGGACGCTTGATTGTGAACTC
 GCCGTCGTCGAAGGCGCTATTGGAGACATCTATAATACAAACACACCTAGCTTGAC
 ATTAGGCTGCGGCTCGATGGGCCGCAACAGCACGACAGACAACGTATCGGTTAATA
 ACTTATTGAACATTAACGCGTCGTCATCCGCAAAGACCGCATGAAATGGTTAAAA
 TCCCTCCGAAAATCTATTTTGAATCGGGATCGTTACAGTATTTATGCAAAGTAAAAA
 AAAAAAAGCGTTTATTGTAACAGACCCGTTTATGGTCAAATTAGGATTTGTAGATA
 AAGTTACATACCAGTTGGACAAAGCTAACATTGACTATGAAATTTTACGCGAAGTAG
 AACCAGATCCGTCGGTCGATACGGTCATGAACGGCGTAAAAATCATGAACTCGTAT
 AATCCAGATTTGATCATCGCTGTGGGAGGAGGAAGCGCGATTGACGCGGCTAAAG
 GAATGTGGTTGTTCTACGAATATCCTGATACGGAGTTCGAAACGTTACGCTTAAAT
 TCGCGGACATTCGCAAACGCGCTTCAAATTTCTGAATTGGGCAAAAAAGCGTTGT
 TCATTGCTATTCGACAACATCGGGCACAGGCTCGGAAGTGACAGCGTTTCGCGGTA
 ATCACAGATAAAAAACGCAACATCAAATACCCATTAGCTGATTACGAATTAACACCT
 GATATCGCTATCATCGACCCTGACTTGACAAAAACGGTACCTCCTTCGGTTACAGCG

GACACAGGCATGGACGTCTTAACGCATGCTATCGAAGCTTATGTGTCGGTGATGGC
 GTCGGATTACACAGATGCTTTAGCGGAAAAAGCTATCAAATTGGTGTTCGAATACTT
 ACCAAAAGCTTACAAAAACGGAAACGATGAAGTGGCGCGCGAAAAAATGCATAAT
 GCGTCGTGCATGGCGGGCATGGCGTTTACAAATGCTTTCTTAGGCATCAACCATAG
 CATGGCTCATATCTTGGGAGGCAAATTCATATCCCGCATGGCCGCGCTAACGCGAT
 CTTGTTACCGTACGTGATTAATAACAATGCGGAAAAACCGACGAAATTTGTGGCGT
 ATCCACAATACGAATATCCTAAAGCGGCGGAACGCTATGCTGAAATTGCGCGCTTCT
 TAGGCTTACCTGCGAGCACAGTGGAAAGAAGGCGTGGAAGCTTAATCGAAGCGAT
 CAAAACTTAATGAAAGAATTAACATCCCTTTGACGTTGAAAGACGCTGGAATCA
 ACAAAGAACAGTTCGAAAAAGAAATCGACGAAATGAGCGACATTGCGTTCAATGAC
 CAATGTACGGGAAGCAATCCGCGCATGCCGTTAACGAAAGAAATCGCGGAAATCTA
 CCGCAAAGCGTACGGAGCG

nphT7

ATGACAGACGTCCGTTTTTCGCATCATTGGAACGGGCGCTTATGTGCCGGAGAGGAT
 AGTTTCTAACGACGAGGTTGGCGCACCAGCCGGGGTTCGATGACGACTGGATCACAC
 GTAAACTGGAATCCGTCAACGACGATGGGCTGCGGATGATCAAGCTACTTCCGAT
 TTGGCTACCGCAGCTGGTCGCGCTGCTTTGAAGGCAGCAGGAATAACACCAGAACA
 ACTCACAGTAATTGCTGTGCAACTTCAACGCCTGATAGACCGCAGCCACCTACAGC
 AGCGTACGTACAACACCATTAGGTGCCACTGGAACAGCGGCCTTTGATGTGAACG
 CCGTATGCAGCGGAACCGTCTTCGCTCTAAGTTCCGTGGCGGGGACGCTAGTGTAT
 CGCGGTGGGTATGCTTTAGTTATCGGGGCGGACCTCTATTGCGGCATTTTAAATCCA
 GCGGACAGAAAAACGTTGTTTTGTTTGGTGACGGTGCAGGGGCGATGGTTTTAG
 GTCCTACGAGCACAGGGACAGGTCCGATTGTCCGGCGGGTGGCCCTTACACCTTT
 GGAGGCTTGACCGATTTAATTCGCGTTCGGGCGGGAGGATCAAGACAGCCGCTGG
 ATACGGATGGGTTGGATGCCGGCTTACAGTACTTCGCAATGGATGGTAGAGAAGT
 GCGCCGCTTTGTAACAGAACATTTACCTCAATTAATTAAGGCTTCTGTCATGAAGC
 GGGAGTAGATGCGGCGGATATTTCTATTTTCGTCGCGCATCAAGCCAATGGAGTGA
 TGCTTGATGAAGTATTTGGAGAATTGCATCTGCCGCGGGCAACGATGCATCGGACA
 GTGGAAACGTATGGCAATACGGGCGCGGCGTTCGATTCCGATTACGATGGATGCGG
 CGGTCCGCGCGGGCAGCTTTGCCCCGGGCGAACTTGTTCTTCTGGCGGGCTTTGGC
 GGCGGCATGGCGGCGTCTGTTTGCCTTATTGAATGG

acc

ATGCCACCGTTTTAGCAGAGTGTTAGTCGCGAATCGGGGAGAAATTGCGGTCAGAGT
 CATGAAAGCGATTAAAGAAATGGGCATGACGGCGATTGCGGTCTATAGCGAAGCG
 GATAAATATGCGGTACATGTTAAATATGCGGATGAAGCGTATTATATTGGACCGAG
 CCCTGCGTTAGAATCGTATCTTAATATTCCTCATATTATTGATGCGGCGGAAAAAGC
 CCATGCGGATGCGGTGCATCCAGGCTATGGCTTTCTTTCCGAAAATGCGGATTTTGT
 AGAAGCGGTTCGAAAAAGCGGGCATGACGTATATTGGACCGTCCGCGGAAGTAATG
 CGCAAAATTAAGATAAATTAGATGGCAAACGCATTGCGCAACTTTCGGGAGTGCC
 GATTGCGCCAGGCTCGGATGGCCCTGTGAATCCATTGATGAAGCGCTTAACTTG
 CGGAAAAAATTGGCTATCCAATTATGGTCAAAGCGGCGTCCGGTGGGGGCGGGGT

AGGCATTACGAAAATTGATACACCAGATCAACTTATTGATGCGTGGGAACGTAACA
 AGCGGTTGGCGACACAAGCGTTTGGACGCTCGGATTTGTATATTGAAAAAGCGGCG
 GTGAATCCTAGACATATTGAATTTCAACTTATTGGAGATAAATATGGCAATTATGTA
 GTCGCGTGGGAACGTGAATGTACGATTCAACGCCGCAATCAAAAACCTTATTGAAGA
 AGCGCCGAGCCCTGCGATTACGATGGAAGAACGCTCCCGCATGTTTGAACCTATAT
 ACAAATATGGCAAACCTATTAATTATTTTACGTTGGGCACGTTTGAACAGTGTTC
 CGATGCGACACGGGAATTTTATTTTCTTGAACCTTAACAAACGCTTGAAGTCGAACA
 TCCTGTGACAGAATTAATTTTTCGCATTGATCTTGTCAAACCTCAAATTCGCCTTGCG
 GCGGGCGAACATCTTCCGTTACACAAGAAGAACTTAATAAGCGCGCGAGAGGCG
 CGGCGATTGAATTTTCGATTAATGCGGAAGATCCAATTAACAATTTTTCCGGCAGCT
 CCGGCTTTACTTATTATCGTGAACCTACAGGACCCGGTGTACGCATGGATTCCGG
 GAGTAACAGAAGGCAGCTGGGTTCCACCGTTTTATGATAGCCTTGTCTCCAACTTA
 TTGTCTATGGAGAAGATAGACAATATGCGATTCAAACAGCGATGAGAGCGTTAGAT
 GATTACAAGATTGGAGGTGTCAAACGACGATTCCTTTGTATAAGCTAATTATGAGA
 GATCCTGATTTTCAAGAAGGCCGCTTAGCACGGCGTATATTAGCCAAAAAATTGAT
 AGCATGGTCAAAAAACTTAAAGCGGAAGAAGAAATGATGGCGTCCGTCGCGGCGG
 TGTTACAATCCCGTGGCCTTCTTCGTAAGAAAGCGTCCGCGCCACAAGAACAAGCG
 AAACCAGGCTCGGGCTGGAAATCGTATGGCATTATGATGCAATCGACACCTCGTGT
 CATGTGGGGCTGA

CAATCAAAAAGGGGGCATAGCGCCATGAACTTTATAGAGTCCATGCGGATACCGG
 CGATACGTTTATTGTAGCGCATGATCAAAAAGAAAATAAGGATCGTCTTAAAACAG
 AAAATAACGAGTTTGAATTGAATATGTGGGCCAAGGCACACGGGAGGGAGAAAT
 TATTCTTAAAATTAATGGCGAAATGCATAGAGTGTTTATTGATAATGGCTGGATTAT
 TCTTGATAATGCGCGTATTTTTCGTGCGGAACGGGTGACGGAACCTTCTACGCAAG
 AAGGCCAAACGTTGGATGAAATGATTAAGGCAAAGAAGGCGAAGTACTTAGCCC
 TCTTCAAGGCAGAGTGGTGCAAGTCAGAGTTAAAGAAGGGGATGCGGTCAACAAG
 GGACAACCGTTACTTAGCATTGAAGCGATGAAATCCGAAACGATTGTGAGCGCGCC
 GATTTCCGGCTTAGTCGAAAAAGTGCTTGTCAAAGCGGGACAAGGCGTCAAAAAG
 GGCGATATTTTGGTAGTAATTAATAA

TGATGATTAAGGGAGTTGTTTGAATATGACAGCGACATTTGAAAAACCGGATATG
 TCCAAACTTGTGGAAGAACTTAGAGCGCTTAAAGCGAAAGCGTATATGGGTGGCGG
 GGAAGAACGGGTTCAAGCGCAACATGCGAAAGGCAAACCTTACAGCGAGAGAACGC
 CTTAATCTTCTTTTATGATGAAGGCACGTTTAAACGAGGTCATGACGTTTTCGACAACG
 AAAGCGACAGAATTTGGCCTTGATAAATCCAAAGTCTATGGAGATGGAGTAGTAAC
 GGGCTGGGGCCAAGTTGAAGGACGCACGGTGTGTTGCGTTTTCGCAAGATTTTACGA
 GCATTGGAGGCACGCTAGGCGAAACACATGCGTCCAAAATTGCGAAAGTCTATGAA
 CTTGCGCTTAAAGTGGGAGCGCCAGTAGTCGGCATTAAATGATTCGGGAGGCGCGC
 GGATTCAAGAAGGCGCGGTCGCGTTAGAAGGCTATGGCACAGTCTTTAAAGCGAAT
 GTCATGGCGTCCGGAGTTGTACCACAAATTACGATTATGGCGGGACCGGCGGCGG
 GAGGCGCGGTCTATAGCCCTGCGCTTACGGATTTTATTATTATGATTAAGGCGATG
 CGTATTATATGTTTGTGACAGGACCTGAAATTACGAAAGTTGTGTTGGGCGAAGAT
 GTGAGCTTTCAAGACCTTGGAGGCGCGGTAATTCATGCGACGAAATCGGGAGTTGT

GCATTTTATTGCGGAAAACGAGCAAGATAGCATTAAATATTACGAAAACGCCTTCTTTC
 CTATCTTCCTAGCAACAACATGGAAGAACCACCGTTTATGGATACAGGGCGATCCAGC
 GGATCGTGAGATGAAAGATGTAGAATCCGTGGTGCCGACAGATACAGTCAAACCTT
 TTGATATGAGAGAAGTGATCTATCGTACAGTCGATAATGGCGAGTTCATGGAAGTG
 CAAAAACATTGGGCGCAAATATGGTGGTCGGCTTTGGCCGTGTCGCGGGCAATGT
 AGTGGGCATTGTAGCGAATAACTCCGCGCATCTTGGCGCGGCGATTGATATTGATG
 CGTCCGATAAAGCGGCGCGTTCATTTCGGTTTTGTGATGCGTTTAATATTCCGTTGA
 TTTTCGTTAGTCGATACACCAGGCTATATGCCTGGCACGGATCAAGAATACAAAGGC
 ATTATTCGACACGGAGCGAAAATGTTGTATGCGTTTTCGGAAGCGACAGTTCCTAA
 AGTAACAGTAGTGGTCAGACGCAGCTATGGAGGCGCGCATATTGCGATGTCCATTA
 AATCGTTGGGTGCGGATCTTATCTATGCGTGGCCATCCGCGGAAATTGCGGTGACG
 GGCCCTGAAGGCGCGGTTTCGCATTTTGTATCGCCGGGAAATTCAAAATAGCAAATC
 CCCAGATGATCTTATTAAGAACGTATTGCGGAATACAAAAAACTTTTTGCGAATCC
 GTATTGGGCGGCGGAAAAAGGACTTATTGATGATGTCATTGAACCTAAAGATACTC
 GAAAAGTAATTGCGTCCGCGCTTAAAATGCTTAAAAATAAGCGTGAATTTGCTATC
 CTAAAAAACATGGCAATATTCCACTTAA

AAAATATATTGGAGGTTGTTTCTAGATTGCAAATTGTCAAACCTTGATCGTGTGACGT
 CCACACAAGATTTTTCGGAAGCGGTCTCCGAAATGATTGATGGAGATTTTGTGGTC
 GTCGCGAAAGAACAACGAAAGCGAGGGGGCGCTATGGCAGAGAATGGTATAGC
 CCTCGTGGCGGACTTTGGGTGACGTATGTGGTCAAAAATTTAATGTGAAGAAAT
 TGGCATTAGCACGCTTAAAGTGGCGTTGGCGATTTCGCAGCATTCTTCCAAAATGGT
 AGATGCGAAAATTCGCTGGCCTAATGATATTGTTATTAATCATAAAAAAGTATCGGG
 CGTGTTACTTGAAGCGATTACAGTAGGGGAACGCTCGACAGGCTTTATTGGCTTGG
 GAGTCAATACGAATGTGGTCTCGTTTCTCAAGGCATTTCCGCGACGAGCCTTAAAC
 TTGAATTAGGACGGGATGTCGATAATGATAGCTTGCTTATGGAAATTCCTTGATAAAA
 TTGCGGAATACTTAGGCAAAGCGCAAGATGATGTCATGCGTGAACCTAACGAGAAT
 CTTAGCATTATGGATAAAGAAGTCAATCTTAAAGGCCGGGATTGGGAAAAACGGTG
 CAAAGCGTTGTTTGTAGATAAATTTGGACGGTTAGTTACGGAATGCGGCATTTTTGA
 AGTAGAAGAAGTGTTACGCCTTGAAACAACACATCATCATCATCATTA

yqhD

ATGAACAACCTTAATCTGCACACCCCAACCCGCATTCTGTTTGGTAAAGGCGCAATC
 GCTGGTTTACGCGAACAATTCCTCACGATGCTCGCGTATTGATTACCTACGGCGGC
 GGCAGCGTGAAAAAACCAGCGTTCGATCAAGTTCTGGATGCCCTGAAAGGCAT
 GGACGTGCTGGAATTTGGCGGTATTGAGCCAAACCCGGCTTATGAAACGCTGATGA
 ACGCCGTGAAACTGGTTCGCGAACAGAAAGTGACTTTCTGCTGGCGGTTGGCGGC
 GGTTCTGTACTGGACGGCACCAAATTTATCGCCGACGGCTAACTATCCGAAAA
 TATCGATCCGTGGCACATTCTGCAAACGGGCGGTAAAGAGATTAAGCGCCATCC
 CGATGGGCTGTGTGCTGACGCTGCCAGCAACCGTTTCCAGAAATCCAACGCAGGCGC
 GGTGATCTCCCGTAAAACCACAGGCGACAAGCAGGCGTTCCATTCTGCCATGTTCA
 GCCGGTATTTGCCGTGCTCGATCCGTTTATACCTACACCCTGCCGCCGCGTCAGGT
 GGCTAACGGCGTAGTGGACGCCTTTGTACACACCGTGGAACAGTATGTTACCAAAC

CGGTTGATGCCAAAATTCAGGACCGTTTCGCAGAAGGCATTTTGCTGACGCTAATC
 GAAGATGGTCCGAAAGCCCTGAAAGAGCCAGAAAACACTACGATGTGCGCGCCAACG
 TCATGTGGGCGGCGACTCAGGCGCTGAACGGTTTGATTGGCGCTGGCGTACCGCA
 GACTGGGCAACGCATATGCTGGGCCACGAACTGACTGCGATGCACGGTCTGGAT
 CACGCGCAAACACTGGCTATCGTCCTGCCTGCACTGTGGAATGAAAAACGCGATAAC
 CAAGCGCGCTAAGCTGCTGCAATATGCTGAACGCGTCTGGAACATCACTGAAGGTT
 CCGATGATGAGCGTATTGACGCCGCGATTGCCGCAACCCGCAATTTCTTTGAGCAAT
 TAGGCGTGCCGACCCACCTCTCCGACTACGGTCTGGACGGCAGCTCCATCCCGGCTT
 TGCTGAAAAAACTGGAAGAGCACGGCATGACCCAACCTGGGCGAAAATCATGACATT
 ACGTTGGATGTCAGCCGCCGTATATACGAAGCCGCCCGCTAA

pduP

ATGAATACTTCTGAACTCGAAACCCTGATTGCGACCATTCTTAGCGAGCAATTAACC
 ACGCCGGCGCAAACGCCGGTCCAGCCTCAGGGCAAAGGGATTTTCCAGTCCGTGAG
 CGAGGCCATCGACGCCGCGCACCAGGCGTTCTACGTTATCAGCAGTGCCCGCTAAA
 AACCCGCGAGCGCCATTATCAGCGCGATGCGTCAGGAGCTGACGCCGCTGCTGGCGC
 CCCTGGCGGAAGAGAGCGCCAATGAAACGGGGATGGGCAACAAGAAGATAAATT
 TCTCAAAAACAAGGCTGCGCTGGACAACACGCCGGGCGTAGAAGATCTCACCACCA
 CCGCGCTGACCGGCGACGGCGGCATGGTGCTGTTTGAATACTCACCGTTTGGCGTT
 ATCGGTTGCGTCCGCAAGCACCAACCCGACGGAAACCATCATCAACAACAGTAT
 CAGCATGCTGGCGGGCGGCAACAGTATCTACTTTAGCCCGCATCCGGGAGCGAAAA
 AGGTCTCTCTGAAGCTGATTAGCCTGATTGAAGAGATTGCCTTCCGCTGCTGCGGCA
 TCCGCAATCTGGTGGTGACCGTGGCGGAACCCACCTTCGAAGCGACCCAGCAGATG
 ATGGCCACCCGCGAATCGCAGTACTGGCCATTACCGGCGGCCCGGGCATTGTGGC
 AATGGGCATGAAGAGCGGTAAGAAGGTGATTGGCGCTGGCGCGGGTAACCCGCC
 TGCATCGTTGATGAAACGGCGGACCTGGTGAAAGCGGCGGAAGATATCATCAACG
 GCGCGTCAATTCGATTACAACCTGCCCTGCATTGCCGAGAAGAGCCTGATCGTAGTG
 GAGAGTGTGCCGAACGTCTGGTGCAGCAAATGCAAACCTTCGGCGCGCTGCTGTT
 AAGCCCTGCCGATACCGACAAACTCCGCGCCGTCTGCCTGCCTGAAGGCCAGGCCGA
 ATAAAAAACTGGTCGGCAAGAGCCCATCGGCCATGCTGGAAGCCGCCGGGATCGC
 TGTCCCTGCAAAGCGCCGCTGCTGATTGCGCTGGTTAACGCTGACGATCCGT
 GGGTACCAGCGAACAGTTGATGCCGATGCTGCCAGTGGTAAAAGTCAGCGATTTT
 GATAGCGCGCTGGCGCTGGCCCTGAAGGTTGAAGAGGGGCTGCATCATACCGCCA
 TTATGCACTCGCAGAACGTGTCACGCCTGAACCTCGCGGCCCGCACGCTGCAAACCT
 CGATATTCGTCAAAAACGGCCCCTTTATGCCGGGATCGGCGTCCGGCGGCGAAGGC
 TTTACCACCTTCACTATCGCCACACCAACCGGTGAAGGGACCACGTCAGCGCGTACT
 TTTGCCCGTTCCCGGCGCTGCGTACTGACCAACGGCTTTTCTATTGCTAA

A9.2 CoA dependent *Geobacillus* native genes**Native *thl***

ATGGGAAAAACGGTGATTTTAAGTGGGGTTCGCACGCCGTTTGGAAAATTCGGCG
 GTGCGTTAAAGCCGCTTACGGCGGCGCAGCTCGGCGGCATTGCTGTGAAAGAAGC
 GCTTCATCGCGCCAGAGTCAGCGGCGAGCAGGTCGATCAAGTCATTTTAGGAACGG
 TGCTGCAAGGAGGGCAAGGCCAGCTTCCATCCCGCCAAGCGATGCGCTATGCTGGC
 ATCCGTGGGAAGTGCAGACAGAAACGGTTAATAAAGTGTGCGCTTCGGGAATGC
 GGAGCGTCACACTTGGCGATCAAATCATCCGCCTGGGAGATGCCGAAGTCATCGTT
 GCGGGCGGAATGGAATCGATGAGCAATGCGCCATACATTTTGCCAAATGCGCGTTG
 GGGATTGCGAATGGGCGATGCGGCTGTCAAAGATTTAATGGTATATGATGGTTTAA
 CATGCAGTTTTACCGGAGTTCATATGGGGTCTACGGCGGCGAAACGGCGAAAGA
 ATTGGAGATTTCCCGCGAAGAACAGGATGAGTGGGCATACCGCAGCCATCAGCGG
 GCGATTGCGGCGATGGAAGCCGATTATTGGCGGAGGAGATCGTACCAGTGGAAG
 TTCCGCAGCACAAAGGGAAGCCGCTGTTGGTGGAGCATGACGAAGCACCGAGAAA
 AGATACGTCTCTTGA AAAACTGGCGAAGCTGCCTCCTGTATTTGACCCGGAAGGGA
 CGATAACAGCGGAAACGCTCCTGGAGTCAATGATGGCGCTGCCGCGCTTGCTG
 ATGAGTGAAGAAAGAGCGGCGCGGAAGGAAGAAAACCGCTCGCGACGATTTTGG
 CCCATGCATCGATTGCCGTTGCGGCAAAAGACTTTCCGAAAACGCCGGGGTTTGTG
 ATCAACGAGCTGCTTCGAAAACGGGGAAAACAGTCGATGACATTGATTTATTTGA
 AATTAACGAGGCGTTTGGCGCTGTGCGCTCGCCAGCATAAAAATCGCCGGCATCG
 ATCCAGAAAAAGTGAATGTCAATGGCGGCGCGGTCGCGTTAGGCCATCCGATCGGC
 GCGAGCGGCGCGCATTATCATTACATTAATTCACGAATTAAGCGCCGCGGCGG
 CGGCATCGGGATCGCGGCGATCTGCAGCGGCGGCGGCAAGGCGATGCCGTTATG
 GTGCAGGTTGATTCGTAA

Native *hbd*

ATGATGGACGTGAAAACAATCATGGTTGTGCGGCCCGGACAGATGGGGTCCGGCA
 TCGCCCAAGTATGCGCGGTGTGCGGATATGATGTTTATCTTCATGACATCAGCCAGA
 CGCAAATCGATAAAGGGCTTGCCCATATTGAAATTATTGTCCCGTCAAGTAGAAAA
 GGGAAAAATGACCGACGCCGAAAAGCGGCGGCGTTGTGCGCCTCATTCTTCGA
 CCAATTTGAAAACGCTGCGAAAGTGGATTTAGTGATTGAAGCGTTGTGAAAAC
 ATGGATGTGAAAACAAAATTGTTTGCCGAACTTGATCAAATCGCTCCGCCGCACACG
 ATTTTGGCGACAAACACGTCTTCGCTGCCAATTACGGAAATCGCAGCGGCGACGAA
 ACGGCCGAAAAAGTGATCGGGATGCATTTTATGAACCCGTTCCGGTCATGAAGC
 TTGTCGAAATTATCCGCGGTTGGCGACCGCCGACGAAGTATATGAAACGATTGAA
 GCCTCACTCGCAAGTTAACAAGTTCTGTTGAAGTCAATGACTTTCCTGGCTTTAT
 CTCCAACCGCTTGTGATGCCGATGATCAACGAAGCGATTTACGCGTTGTATGAAG
 GGGTGGCGACGAAAGAAGCGATTGACGAAGTTATGAAGCTCGGCATGAACCATCC
 GATGGGCCGTTAACATTAGCAGATTTTATCGGGCTAGACACGTGCTTATACATTAT
 GGAAACACTTCATGAAGGATTTGGCGATGATAAATACCGTCCATGCCATTGTTGC
 GGAAATATGTGAAAGCCGGCTGGCTCGGCCGAAAACAGGCAGAGGTTTTTACAC
 GTATGAGTAA

Native *crt*

ATGACTGAGTTTGCACACATTCGCACACATGTTGATGAGAATGTTGCTGTGATTGAA
 CTGCATCGCCCAGACGTGTTAAACGCCTTGAATAGGAAAATGGTAACGGAAATCGT
 CAGCGCGATGGAATTTTACGACCGCGATGAACATGTGCGCGTCATCGTCTTGACAG
 GAAGCGGCCGTGCCTTTGCAGCAGGAGCGGATATCGACGAAATGGCAAACGATGA
 TTCCATTACGCTGGAATTGCTTAATCAATTTGCCGAATGGGATCGCCTTGCACTTATT
 AAAAAGCCTACGATTGCCGCCGTACACGGTTTTGCATTAGGCGGAGGATTGCAACT
 CGCGCTTAGCTGTGACCTGTTGTTTGCGGCCGATACAGCGGAATTCGGATTTCCGG
 AAGTGAATATCGGTGTGATGCCAGGTGCGGGCGGAACGCAGCGCCTTACTAAATTA
 ATCGGCAAAACAAAAGCGCTCGAGTGGCTTTGGACAGGCGAGCGTATGTCGGCAA
 AAGAAGCGCATGAATTAGGGATTGTCAACCGTGTGTCATCGCGCCAGAGTTATTGGAT
 GAGGAAACAATGCGTTTTGCGAAAAAGCTGGCAAAACAGCCTCCTTTATCCGTCCG
 CCTAATTAAGAAGCCGTGAATAAGGCAGTAGACTATTCGCTATATGAAGGCATGC
 AGTTTGAGAGAAAAAATTTTTATCTCTTTTTTCGTCTGAGGATCAAAAAGAGGGAA
 TGGCGGCATTTATCGAAAAACGAAAGCCGCGTTTTCAAGGAAGATAA

Native *adhE*

ATGGCTGTGGAGGAAAGAGTCGTCGATAAAAAAATCGAAGCAGCAAAAATGATTG
 ATGAGCTTGTGCTAATGCACAGAAAGCGTTGGAACAATTTGCGCCTTACGATCAA
 GAAACGATCGATCATATCGTCAAAGAAATGGCGTTAGCCGGGCTCGACAAGCATAT
 GGCATTAGCCAAGCTTGCAGTAGAAGAAACAAAACGCGGTGTATATGAAGATAAA
 ATCATAAAAAACCTTTTTGCGACAGAATATATATACCACAATATTAAGTATGATAAA
 ACAGTCGGGATTATTCATGAAAATCCGCATGAAGAAATTATCGAAATTGCTGAGCCT
 GTTGGTGTATTGCTGGGATTACGCCAGTGACAAACCCGACATCGACAACGATGTTT
 AAAGCGTTAATCTCGATAAAAAACGCAACCCGATTATTTTCGCTTCCATCCATCGG
 CGCAACGATGCAGCAGCGAAGCGGCAAGAGTGCTGCGCGATGCGGCGGTCCGGG
 CAGGGGCTCCAGAACATTGCATTCAATGGATTGAAACTCCTTCGCTTGATGCAACCA
 ATCAGCTTATGCACCATCTGGCGTTTCTCATTGTTGGCAACTGGTGGCGCCGGCA
 TGGTCAAAGCAGCGTACAGCTCTGGAAAACAGCTTTGGGCGTCCGACCTGGCAAT
 GTGCCTTGCTATATTGAAAAACGGCAAAACATAAAAACGGGCGGTAAATGACTTAAT
 TTTATCGAAAACGTTTGATAACGGCATGATTTGCGCTTCTGAACAAGCAGTCATTAT
 TGATAAAGAAATTTATGAACAAGTAAAGAAAGAAATGATAGAAAACCATTTGTTATT
 TCTTAAATGAAGAAGAAAAGAAAAAAGTAGAAAAACTCGTTATCAATGAAAATACA
 TGCGCCGTCAACCCGGATATCGTCGAAAGCCAGCTTATGAAATTGCGAAAATGGC
 CGGCATCGCTGTGCCGGAAGACACAAAATTCTTGTGCTGAGTTAAAAGGGGTCCG
 GGCCAAAATATCCGTTGTCTCGGGAAAAATTAAGCCCTGTCCTTGCTTGCTATAAAG
 TTAACAGCACGGAAGAAGGATTTAAGCGCTGTGAAGAAATGCTGGAATTTGGCGG
 CTTGGGACATTCGGCTGTCATCCATTCCGATAATCAAACGTGGTTACCGAATTTGG
 CAAACGGATGAAAGCGGGACGGATTATCGTTAATGCGCCATCTTCGCAAGGAGCAA
 TCGGCGATATTTACAATGCGTACATTCCGTCATTAACGCTGGGATGCGGCACATTTG
 GCGGAAACTCTGTTTCGACAAACGTGAGTGGGATTGCTTATCAATATAAAAAGAA
 TGCAAAAAGGACGGTAAATATGCAATGGTTTAAAGTGCCGCCGAAAATTTATTTT
 GAAAAAATGCTGTACAATACTTAGCGAAAATGCCGGATATTTCCAGAGCTTTTATC

GTCACCGACCCGGGAATGGTCAAGCTCGGATATGTCGATAAAGTGCTGTACTT
GCGCAGACGCCGGATTATGTGCATAGTGAAATTTCTCCGAAGTAGAGCCAGATC
CTTCAATTGAGACGGTAATGAAAGGTGTCGATATGATGAGAAGTTTCGAGCCGGAT
GTGATTATCGCGCTTGAGGCGGCTCGCCAATGGATGCGGCAAAGCGATGTGGC
TCTTTTACGAGCATCCGACAGCGGATTTCAACGCATTAACAAAAATTTTAGATA
TTCGAAAACGCGTTTATAAATATCCAAAAGTGGGCCAAAAGCGAAATTTGTCGCCA
TTCCGACGACATCAGGAACAGGATCGGAAGTAACGTCCTTTGCCGTCATTACCGATA
AAAAACGAATATAAATATCCGTTGGCAGATTATGAATTGACACCGGACGTCGCG
ATTGTGGATCCGCAATTTGTCATGACCGTGCCAAAACATGTCACCGCCGATACGGG
AATGGATGTATTGACACATGCGATCGAAGCGTATGTCTCCAATATGGCCAATGATTA
TACCGATGGTCTTGCCATGAAAGCAATCCAAGTTCGATTTGAATATTTGCCGCGGGC
ATATCAAACGGAGCGGATGAGCTTGCCCGGAGAAAATGCATAACGCCTCTACGA
TTGCGGGAATGGCATTGCCAACGCGTTTTTAGGCATTAACCATAGTTTGGCTCATA
AACTTGGCGCGGAATTCCATATTCCGCATGGGCGCGCAATACCATTTTGATGCCGC
ATGTCATTGCTATAACGCAGCGAAACCGAAAAATTTACCGCATTTCGAAATACG
AATATTTCAAAGCGGACCAGCGCTATGCAGAAATTGCGAGAATGCTCGGCTTGCCG
GCCCCACAACGGAAGAAGGGTTCGAAAGCCTCGTTCAGGCGATCATTAAAGCTGG
CAAAACAGTTGGATATGCCGCTGAGCATTGAAGCATGCGGCGTCAGCAAACAAGA
ATTTGAAAGCAAAGTTGAAAATTAGCCGAATTGGCTTTCGAAGACCAATGTACTAC
TGCTAACCCGAAACTCCCCTTGTTAGCGATTTAGTTCATATTTATCGCCAAGCGTTT
AAAGGAGTTTAA

fabI

ATGAATTTATCATTAAAAGGGCGCACATATGTTGTGATGGGAGTGGCGAATAAGCG
CAGCATCGCTTGGGGAATCGCGAGATCTTTGCATGCGGCGGGGGCGCGTCTTATTT
TTACATACGCCGGTGAGCGTTTTGAAAAAGAGGTTTCGTTCCCTTGTAGATACGCTGG
AAGATGAAAGCTCCCTGCTTCTCCTTGCGATGTGACGAAAGATGAGGAAATTGTC
CAGTGTTCAGCAAATTAAGGAGCAAGTCGGCGTCATTCATGGGGTTGCCATTG
CATTGCGTTTGCGAACAGAGAAGATTTAAGCGGAGAATATATGAAAGTAAGCCGCG
AAGGATTTTTGCTAGCCCACAACATCAGCGCCTATTCCTTAACCGCGTTGCTAGAG
AAGCGAAAGAGCTGATGACGGAAGGCGGAAGCATCGTAACATTAACATATCTTGG
CGGCGAGCGCGTCGTGCAAAATTATAACATTATGGGAGTGGCGAAAGCGTCGTTG
GATGCCAGCGTTAAATATTTAGCAAACGATTTAGGAAAGTATGGCATCCGCGTCAA
CGCGATTTGCGCCGACCGATTGCGACGCTGTCGGCAAAGGAATTAGCGATTTTA
ACGCGATTTTAAAACAAATTGAAGAACGCGCTCCGCTCCGGCGCACGACGACGCAA
GAAGAAGTCGGCGATACGGCGCTGTTTTATTTAGCGATTTGTCGCGTGGCATTAC
GGGCGAAATTATCCATGTCGATTCTGGATATCATATTTTAGGATATTA

A9.3 CoA independent genes

tesBT thioesterase

(*E. coli* optimised, Sequence Length:762, GC%:48.41, His-tagged)

ATGTCGGAAGAAAACAAAATCGGCACCTACCAATTCGTTGCTGAACCGTTCACGTT
 GACTTTAATGGCCGCTGACGATGGGTGTGCTGGGCAACCATCTGCTGAATTGCCG
 AGGTTTTACGCTTCTGATCGTGGCTTCGGTATCGCAACCCTGAACGAAGACAATTA
 TACGTGGGTGCTGTCCCGCTGGCGATTGAACTGGATGAAATGCCGTATCAGTACG
 AAAAATTTTCAGTTCAAACCTGGGTTGAAAACGTCTATCGTCTGTTTACGGACCGCA
 ATTTCCGGTCATCGATAAAGACGGCAAAAAGATTGGTTACGCGCGTTCCGGTGTGG
 GCCATGATCAACCTGAATACCCGTAAGCCGGCAGATCTGCTGGCACTGCATGGTGG
 TAGTATCGTCGATTATATTTGCGACGAACCGTGTCCGATCGAAAAACCGAGCCGTAT
 TAAGGTGACGTCTAACCAGCCGGTTGCAACCCTGACGGCTAAATACTCTGATATTGA
 CATCAACGGCCATGTTAACAGTATCCGTTACATCGAACACATTCTGGACCTGTTTCC
 GATCGAACTGTACCAAACCAAGCGCATTTCGTCGCTTTGAAATGGCGTATGTTGCCG
 AATCATACTTCGGCGATGAACTGTGCTTTTTCTGCGACGAAGTCAGCGAAAACGAA
 TTCACGTCGAAGTGAAAAGAATGGTAGTGAAGTGGTTTGTGCTCCAAAGTGATT
 TTCGAA**CATCACCATCACCATCAC**TAA

TSCA operon

(*Geobacillus* optimised, His-tagged)

CTTAAAGGAGGTTTTATATTATGTGCGGAAGAAAACAAAATTGGCACGTATCAATTTG
 TGGCGGAACCGTTTCATGTGGATTTAATGGACGCTTAACAATGGGCGTGTTAGGC
 AATCATTTATTGAATTGTGCGGGATTTTCATGCGTCGGATCGCGGATTTGGCATTGCG
 ACATTAATGAAGATAACTACACATGGGTGTTATCGCGCTTAGCGATTGAATTAGAT
 GAAATGCCGTACCAATATGAAAATTCTCGGTGCAAACGTGGGTGCAAAAATGTGTA
 TCGCTTATTTACGGATCGCAACTTTGCGGTGATTGATAAAGATGGCAAAAAAATCG
 GCTATGCTCGCTCGGTGTGGGCTATGATTAATTTGAATACACGCAAACCGGCTGATT
 TGTTAGCGTTACATGGCGGATCGATTGTGGATTACATTTGTGATGAACCGTGTCCGA
 TTGAAAACCGTCGCGCATTAAAGTGACATCGAATCAACCGGTGGCGACGTTAACA
 GCGAAATATTCGGATATTGATATTAACGGCCATGTGAACTCGATTGCTACATTGAA
 CATATCTTGACTTGTTCCCGATCGAACTGTATCAAACAAAACGCATTTCGAGATTT
 GAAATGGCGTATGTTGCGGAATCGTATTTTGGAGATGAACTGTCGTTTTTTTTGCGAC
 GAAGTGTGCGAAAATGAATTTTCATGTGCAAGTGAAAAAAAACGGCTCGGAAGTGG
 TGTGCCGCTCGAAAGTGATTTTTGAATAAGATATCGAAATAAAGGAGGTGAGTATA
 TGAAAATCTATGGCATTATATGGATCGCCCTCTGTCGCAAGAAGAAAACGAACGCT
 TTATGTCGTTTATTTCTCCGAAAAACGCGAAAAATGCCGTCGTTTTATCATAAAG
 AAGATGCGCATAGAACGTTGTTGGGAGATGTGTTAGTGCCTCGGTGATTTCTCGC
 CAATATCAATTAGATAAATCGGACATTCGCTTTTCGACGCAAGAATATGGCAAACCG
 TGCATTCCGGATTTACCGGATGCGCATTTTAACATTTTCGCATAGCGGACGCTGGGTG
 ATTTGCGCGTTTGATTTCGCAACCGATTGGCATTGACATTGAAAAACAAAACCGATT
 TCGCTGAAAATCGCGAAACGCTTTTTTAGCAAAACGGAATACTCGGACTTACTGGC
 GAAAGATAAAGACGAACAAACGGACTACTTTTACCATCTGTGGTCGATGAAAGAAT
 CGTTCATTAAACAAGAAGGCAAAGGCTTATCGTTACCGTTAGATTCGTTTTCGGTGC

GCTTACATCAAGATGGCCAAGTGAGCATTGAATTACCGGATAGCCATTCGCCGTGC
 TACATTAACCGTATGAAGTGGATCCAGGCTATAAAATGGCGGTGTGCGCTGCTCA
 TCCAGATTTTCCGGAAGATATTACAATGGTGTGCGTATGAAGAAGTGTGTAAGAATT
 CTTTTAAAGGAGGTATAAGCTATGTCGCCGATTACACGCGAAGAACGCTTAGAAAG
 ACGCATTCAAGATTTGTATGCGAACGATCCGCAATTTGCGGCTGCGAAACCGGCTA
 CAGCGATTACAGCGGCTATTGAACGCCAGGATTACCGTTACCGCAAATTATTGAAA
 CAGTGATGACAGGCTATGCGGATCGCCAGCGTTAGCGCAACGCAGCGTGGAATTT
 GTGACAGATGCTGGCACAGGCCATAACAATTACGCTTATTACCGCATTTTGAAACG
 ATTTGATGCGCAATTATGGGATCGCATTTCGGCGTTAGCGGATGTGTTATCGACA
 GAACAAACAGTGAACCGAGGCGATCGCGTGTGCTTATTAGGCTTTAATTCGGTGGA
 TTATGCGACGATTGATATGACATTAGCTCGCCTTGGAGCGGTGGCGGTGCCGTTAC
 AACATCGGCTGCGATTACACAATTACAACCGATCGTGGCGGAAACACAACCGACA
 ATGATTGCGGCTTCGGTTGATGCGTTGGCGGATGCGACAGAATTAGCGTTATCGGG
 ACAAACAGCGACACGCGTTTTAGTGTGTTGATCATCATAGACAAGTGGACGCGCATA
 GAGCGGCTGTGGAATCGGCTCGCGAACGCCTTGGCGGATCTGCGGTGGTGAAAC
 ATTAGCGGAAGCGATTGCTCGCGGAGATGTGCCACGCGGAGCGTTCGGCTGGATCT
 GCGCCAGGCACAGATGTGTCGGATGATAGCTTAGCGTTGTTGATTTATACATCGGG
 CTCGACAGGCGCTCCGAAAGGCGCTATGTATCCACGCAGAAATGTGGCGACATTTT
 GGCGCAAACGCACATGGTTTGAAGGCGGATATGAACCGTCGATTACGTTGAACTTT
 ATGCCGATGTGCGATGTGATGGGACGCCAAATTTTGTATGGCACATTATGCAATGG
 CGAACAGCGTATTTTGTGGCGAAATCCGATTTAAGCACGCTGTTTGAAGATTTAGC
 GTTAGTTCGCCAACCGAATTAACATTTGTGCCTCGCGTGTGGGATATGGTGTTGA
 CGAATTTCAAAGCGAAGTGGATAGACGCTTAGTGGATGGCGCTGATCGCGTGGCG
 TTAGAAGCGCAAGTGAAGCGGAAATTCGCAATGATGTTTTAGGCGGACGCTATAC
 GTCGGCGTTAACAGGCTCGGCTCCGATTTCCGATGAAATGAAAGCGTGGGTTGAA
 GAATTGCTGGATATGCATTTAGTGAAGGCTATGGCTCGACGGAAGCGGGAATGA
 TTTTGATTGATGGTGCGATTTCGACAGCCGGCTGTGTTAGATTATAAACTTGTGGATG
 TGCCGATCTGGGCTATTTTTGACAGATCGCCACATCCGAGAGGCGAATTGTTAG
 TGAAAACGATTTCGTTATTTCCGGGATATTATCAACGCGCTGAAGTACAGCGGAT
 GTTTTTGATGCGGATGGCTTTTATCGCACAGGCGATATTATGGCTGAAGTTGGACCG
 GAACAATTTGTGATTTAGATCGCCGTAACAACGTGCTGAAATTGTCGCAAGGCGA
 ATTTGTCACAGTGTGCAAAATTAGAAGCGGTGTTTGGCGATTTCGCCGTTAGTGAGAC
 AAATTTACATTTATGGCAATTCGGCGAGAGCGTATTTATTGGCGGTTATTGTGCCGA
 CACAAGAAGCGTTAGACGCGGTTCCGGTGAAGAATTAAGGCTCGCTTAGGCGAT
 TCGTTACAAGAAGTGGCGAAAGCGGCTGGCTTACAATCGTATGAAATTCCTCGCGA
 CTTTATCATTGAAACAACACCGTGGACATTGGAAAATGGCTTATTAACAGGCATTTCG
 CAAATTAGCACGCCACAGCTTAAAAACATTACGGCGAATTACTGGAACAAATCTA
 TACGGATTTAGCGCATGGCCAAGCGGATGAATTACGCTCGTTACGCCAAAGCGGAG
 CGGATGCTCCGGTGTAGTGACAGTGTGTCGCGCTGCGGCAGCGTTATTAGGCGGA
 TCGGCGTCGGATGTGCAACCAGATGCTCATTTACAGATCTTGGCGGAGATTTCGTTA
 AGCGCGTTATCGTTTACGAATCTGCTGCATGAAATTTTCGACATTGAAGTGCCGGTG
 GCGGTGATTGTGTCGCCAGCGAATGATTTACAAGCGCTGGCGGATTATGTGGAAGC
 TCGAGAAAACCGGGATCGAGCAGACCGACATTTGCGTCCGGTGCATGGTGCGTGC

AATGGACAAGTGACGGAAGTGCATGCTGGCGATTTATCGTTGGATAAATTCATTGA
TGCGGCTACACTTTCGGAAGCGCCACGCTTACCGGCAGCGAATACACAAGTGCGCA
CAGTGTATTGACAGGTGCGACAGGCTTTTTAGGACGCTATTTAGCGCTTGAATGGT
TAGAACGCATGGATTTAGTTGACGGCAAACGATTTGCTTAGTGAGAGCTAAAAGC
GATACGGAAGCTCGCGCACGCTTAGATAAAAACGTTTGATAGCGGAGATCCGGAATT
ATTGGCGCATTATAGAGCGTTAGCTGGCGATCATTTAGAAGTGTTAGCAGGCGATA
AAGGCGAAGCGGATTTAGGCTTAGACCGCCAAACATGGCAACGCTTAGCGGATAC
AGTGGATTTAATTGTGGATCCGGCTGCGTTAGTGAATCATGTGTTACCGTATTGCGA
ATTGTTTGGACCGAATGCGCTTGGCACAGCGGAATTACTTCGCTTAGCGCTTACGTC
GAAAATCAAACCGTATTCTATACATCGACAATTGGCGTGGCGGATCAAATTCGCGC
ATCGGCGTTTACGGAAGATGCGGATATTCGCGTGATTTGCGGCGACAAGAGCGGTG
GATGATTCCTATGCGAATGGATATAGCAATTCGAAATGGGCTGGCGAAGTGTTATT
ACGCGAAGCGCATGATTTATGCGGCTTACCGGTTGCGGTGTTTCGCTGCGATATGA
TTTTGGCGGATACAACGTGGGCTGGCCAATTAATGTTCCGGATATGTTTACACGCA
TGATCTTAAGCTTAGCGGCTACGGGCATTGCGCCTGGCTCGTTTTATGAATTAGCGG
CTGATGGCGCTCGCCAACGCGCTCATTATGATGGATTACCGGTGGAATTTATTGCTG
AAGCGATTTGACGTTGGGAGCGCAATCGCAAGATGGCTTTCATACATATCATGTC
ATGAACCCGTATGATGATGGCATTGGCTTAGATGAATTTGTGGATTGGTTAAATGA
ATCGGGCTGTCCGATTCAACGCATTGCTGATTATGGCGATTGGTTGCAACGGTTTGA
AACAGCGTTACGTGCGTTGCCGGATCGCCAAAGACATTCGTCGTTATTACCGTTGTT
ACATAATTATCGCCAACCGGAACGCCCTGTTGCGGTTGATTGCGCCTACAGATCG
CTTTCGCGCAGCGGTGCAAGAAGCGAAAATTGGTCCGGATAAAGATATTCGCGCATG
TTGGCGCACCGATTATTGTGAAATATGTTAGCGATTTACGCTGTTGGGCTTGTGT
AACATATGTCTAGATAGATAAAGGAGGTAGAACTATGCATCATCACCATCACCCTC
GATGATTAATCGTATGCGGCTAAAGAAGCTGGCGGAGAATTAGAAGTTTATGAAT
ATGATCCAGGCGAACTTCGCCACAAGATGTGGAAGTGCAAGTGGATTATTGCGGC
ATTTGCCATTCCGATTTATCGATGATTGACAACGAATGGGGCTTTTCGCAATATCCG
TTAGTTGCGGGACATGAAGTGATTGGACGCGTGGTTGCGTTAGGCTCGGCAGCGC
AAGATAAAGGCTTACAAGTGGGCCAACGCGTCGGCATTGGATGGACAGCTCGCAG
CTGTGGCCATTGTGATGCGTGATTTTCGGGCAATCAAATTAAGTGCGAACAAGGTG
CGGTGCCGACAATTATGAATCGCGGAGGCTTTGCGGAAAACTTCGCGCTGATTGG
CAATGGGTTATTCCGTTACCGGAAAACATCGATATTGAATCTGCGGGACCGTTACTT
TGCGGAGGCATTACAGTTTTTAAACCGCTGTTAATGCATCATATCACAGCGACATCT
CGCGTGGGCGTCATTGGAATTGGAGGCTTAGGACATATTGCGATTAACTGTTACA
TGCTATGGGCTGTGAAGTTACAGCGTTTTTCGTCGAATCCTGCGAAAGAACAAGAAG
TTTTAGCTATGGGAGCTGACAAAGTGGTTAATTCGAGAGATCCGCAAGCTTTAAAA
GCATTAGCGGGACAATTTGATCTGATCATCAATACAGTGAACGTGTCGTTAGATTG
GCAACCGTATTTTGAAGCGCTGACGTATGGCGGAAATTTTCATACAGTCGGAGCGG
TTTTAACACCGTTATCGGTTCCGGCTTTTACATTAATTGCTGGGGATCGCTCGGTTAG
CGGCTCGGCGACAGGCACACCGTATGAATTGCGCAAATTAATGCGCTTTGCGGCTC
GCAGCAAAGTGGCTCCGACAACAGAATTATTTCCGATGAGCAAATTAACGACGCG
ATTCAACATGTTTCGCGACGGCAAAGCACGCTATCGCGTTGTGTTAAAAGCGGACTTT
TAA

Acyl-ACP thioesterase (TesBT)**CTTAAAGGAGGTTTTATATT**

MSEENKIGTYQFVAEPFHVDFNGRLTMGVLGNHLLNCAGFHASDRGFGIATLNEDNYT
 WVLSRLAIELDEMPYQYEKFSVQTWVENVYRLFTDRNFAVIDKDGKKIGYARSVWAMI
 NLNTRKPADLLALHGGSSIVDYICDEPCPIEKPSRIKVTSNQPVALTLAKYSDIDINGHVNSI
 RYIEHILDLFPIELYQTKRIRRFEMAYVAESYFGDELSFFCDEVSENEFHVEVKKNGSEVVC
 RSKVIFE*Stop
 247aa, 741bp

Phosphopantetheinyl transferase (Sfp)**GAAATAAAGGAGGTGAGTAT**

MKIYGIYMDRPLSQEENERFMSFISPEKREKCRRFYHKEDAHRTLLGDVLVRSVISRQYQ
 LDKSDIRFSTQEYKPCIPDLPAHFNISHSGRWVICAFDSQPIGIDIEKTKPISLEIAKRFF
 SKTEYSDLLAKDKDEQTDYFYHLWSMKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELP
 DSHSPCYIKTYEVDPGYKMAVCAAHPDFPEDITMVSYEELL*Stop
 224aa, 672bp

Carboxylic acid reductase (Car)**TTTTAAAGGAGGTATAAGCT**

MSPITREERLERRIQDLYANDPQFAAAKPATAITAAIERPGLPLPQIETVMTGYADRPAL
 AQRVSEFVTDAGTGHTTLRLLPHFETISYGELWDRISALADVLSTEQTVKPGDRVCLLGF
 NSVDYATIDMTLARLGAVAVPLQTSAAITQLQPIVAETQPTMIAASVDALADATELALS
 GQTATRVLVFDHHRQVDAHRAAVESARERLAGSAVVETLAEAIARGDVPRGASAGSAP
 GTDVSDDSLALLIYTSGSTGAPKAMYPRRNVATFWRKRTWFEGGYEPSITLNFMPMS
 HVMGRQILYGTLCNGGTAYFVAKSDLSTLFEDLALVRPTELTFVPRVWDMVFDEFQSE
 VDRRLVDGADRVALEAQVKAIEIRNDVLGGRYTSALTGSAPISDEMKAWVEELDMHLV
 EGYGSTEAGMILIDGAIRRPVLDYKLVDPDLGYFLTDRPHPRGELLVKTDSLFPGYQ
 RAEVTADVFDADGFYRTGDIMAIEVGPEQFVYLDRRNNVLKLSQGEFVTVSKLEAVFGD
 SPLVRQIYIYGN SARAYLLAVIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQSYEIPR
 DFIIETTPWTLNGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSRQSGADAP
 VLVTVCRAAAALLGGSASDVQPDAHFTDLGGDSLALSFTNLLHEIFDIEVPVGVIVSPA
 NDLQALADYVEAARKPGSSRPTFASVHGASNGQVTEVHAGDLSLDKFIDAATLAEAPRL
 PAANTQVRTVLLTGATGFLGRYLALEWLERMDLVDGKLI CLVRAKSDTEARARLDKTFD
 SGDPPELLAHYRALAGDHLEVLGDKGEADLGLDRQTWQRLADTVDLIVDPAALVNHVL
 PYSQLFGPNALGTAELLRLALTSKIKPYSYSTIGVADQIPPSAFTEDADIRVISATRAVDD
 SYANGYSNSKWAGEVLLREAHDLCLPVAVFRCMILADTTWAGQLNVPDMFTRMIL
 SLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAI STLGAQSQDGFHTYHVMNPYD
 DGIGLDEFVDWLNESGCPRIADYGDWLQRFETALRALPDRQRHSSLLPLLHNYRQPE
 RPVRGSIAPTDRFRAAVQEAKIGPDKDIPHVGAPIIVKYVSDLRLLGLL*Stop
 1174aa, 3522bp

Aldehyde reductase (Ahr)

TAGATAAAGGAGGTAGAACTATG-**CATCATCACCATCACCAC**-
 SMIKSYAAKEAGGELEVYEDPGELRPQDVEVQVDYCGICHSDLSMIDNEWGFSQYPL
 VAGHEVIGRVVALGSAAQDKGLQVGQRVIGWTARSCGHCDACISGNQINCEQGAV
 PTIMNRGGFAEKLRADWQWVIPLPENIDIESAGPLLCGGITVFKPLLMHHITATSRVGI
 GIGGLGHIAIKLLHAMGCEVTAFSSNPAKEQEVLAMGADKVVNSRDPQALKALAGQFD
 LIINTVNVSLDWQPYFEALTYGGNFHTVGAVLTPLSVPAFTLIAGDRSVSGSATGTPYEL
 RKLMRFAARSKVAPTTTELPMSKINDAIQHVRDYGKARYRVVLKADF*Stop
 339aa, 1017bp

TSCA operon RBS'

1. Sporogenes RBS 8: **CTTAAAGGAGGTTTTATATT**_ATG (TES)
2. Sporogenes RBS 5: **TTTTAAAGGAGGTATAAGCT**_ATG (CAR)
3. Sporogenes RBS 9: **TAGATAAAGGAGGTAGAACT**_ATG (Ahr)
4. Sporogenes RBS 6: **GAAATAAAGGAGGTGAGTAT**_ATG (Sfp)